ФГБОУ ВО «Московский государственный
университет имени М. В. Ломоносова»

На правах рукописи

Мусатов Даниил Владимирович

Комбинаторные методы в теории колмогоровской сложности с ограничением на ресурсы

01.01.06 — математическая логика, алгебра и теория чисел

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Москва — 2014

Работа выполнена на кафедре математической логики и теории алгоритмов Механико-математического факультета ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова».

Научный руководитель:	**Верещагин Николай Константинович**, доктор физико-математических наук, профессор.
Официальные оппоненты:	**Аблаев Фарид Мансурович**, доктор физико-математических наук, профессор (ФГАОУ ВО «Казанский (Приволжский) федеральный университет», Институт вычислительной математики и информационных технологий, кафедра теоретической кибернетики, зав. кафедрой); **Подольский Владимир Владимирович**, кандидат физико-математических наук (ФГБУН Математический институт им. В.А. Стеклова РАН, отдел математической логики, старший научный сотрудник).
Ведущая организация:	ФГБУН Санкт-Петербургское отделение Математического института им. В.А. Стеклова РАН.

Защита диссертации состоится 13 февраля 2015 г. в 16 ч. 45 мин. на заседании диссертационного совета Д 501.001.84, созданного на базе ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова», по адресу: Российская Федерация, 119991, Москва, ГСП-1, Ленинские горы, д. 1, ФГБОУ ВО МГУ им. М.В. Ломоносова, Механико-математический факультет, ауд. 14-08.

С диссертацией можно ознакомиться в Фундаментальной библиотеке ФГБОУ ВО МГУ им. М.В. Ломоносова по адресу: Москва, Ломоносовский проспект, д. 27, сектор А, а также в сети Интернет по адресу `http://mech.math.msu.su/~snark/index.cgi`.

Автореферат разослан 13 января 2015 г.

Учёный секретарь
диссертационного совета
Д 501.001.84, созданного на базе
ФГБОУ ВО МГУ им. М.В. Ломоносова,
доктор физико-математических наук,
профессор

Иванов Александр Олегович

Общая характеристика работы

Актуальность темы

Понятие колмогоровской сложности появилось в 1960-х годах в работах Колмогорова[1], Соломонова[2] и Чейтина[3,4]. Колмогоровскую сложность можно определять для любых конечных объектов, но достаточно рассматривать двоичные слова, т.е. конечные последовательности из нулей и единиц. Неформально говоря, сложность слова есть сложность его алгоритмического описания, т.е. длина кратчайшей программы, которая печатает это слово на пустом входе. Также рассматривают условную сложность одного слова относительно другого, т.е. длину кратчайшей программы, печатающей первое слово после получения на вход второго. Разумеется, эти определения зависят от того, какой язык программирования используется для описания, но согласно теореме Колмогорова–Соломонова существует оптимальный язык программирования, при котором сложности всех слов минимальны с точностью до аддитивной константы. Это позволяет не ссылаться на конкретный язык программирования и при этом формулировать различные утверждения о связи сложностей различных слов. Эти утверждения формулируются с точностью до аддитивной константы, зависящей только от выбора оптимального языка программирования и не зависящей от конкретных слов. В формулировках эта константа традиционно обозначается через $O(1)$. (Некоторые утверждения формулируются с точностью до другого аддитивного слагаемого: $O(\log n)$, $O(\log^3 n)$ и т.п.) К сожалению, до сих пор не сложилось единого обозначения для сложности. В диссертации сложность слова x обозначается через $\mathrm{C}(x)$, а условная сложность слова x относительно y — через $\mathrm{C}(x|y)$.

Также рассматривают колмогоровскую сложность с ограничением на ресурсы.[5] В этом случае программа, генерирующая слово, должна быть

[1] А.Н. Колмогоров. Три подхода к определению понятия «Количество информации» // Проблемы передачи информации. — 1965. — Т.1. №1. — С. 3–11.

[2] R.J. Solomonoff. A Formal Theory of Inductive Inference, Part 1, Part 2 // Information and Control (now Infromation and Computation). — 1964. — V. 7. — P. 1–22, 224–254.

[3] G.J. Chaitin. On the length of programs for computing binary sequences // Journal of the ACM — 1966. — V. 13. №4. — P. 547–569.

[4] G.J. Chaitin. On the length of programs for computing binary sequences: statistical considerations // Journal of the ACM. — 1969. — V. 16. №1. — P. 145–159.

[5] M. Li, P.M.B. Vitányi. An Introduction to Kolmogorov Complexity and its Applications. — 3rd

не только короткой, но и использующей ограниченные вычислительные ресурсы, а именно время работы и память. Разумеется, при наложении этих дополнительных ограничений сложность не уменьшается. Важно, что в отличие от сложности без ограничений сложность с ограничениями является вычислимой функцией.

Для доказательства того, что сложность некоторого слова мала, часто используют следующий принцип, восходящий к Сипсеру:[6] элементы любого небольшого перечислимого множества просты. Более точно: если перечислимое множество S содержит не больше K слов длины n, то все эти слова имеют сложность не больше $\log K + \log n + O(1)$. Действительно, каждое из них можно задать перечисляющим алгоритмом, длиной n и номером в перечислении слов данной длины. Это наблюдение позволяет доказывать различные утверждения о колмогоровской сложности через комбинаторные конструкции: сначала строится некоторое перечислимое множество, затем комбинаторными методами доказывается верхняя оценка на его размер, и, наконец, делается вывод о малой сложности любого входящего в него слова. Более того, если перечисляющий алгоритм вычислительно эффективен, то сложность с ограничением на ресурсы также мала. В диссертации изучаются два примера такого подхода.

Во-первых, изучается теорема Мучника об условном кодировании.[7] Она гласит, что для любых слов a и b, таких что $\mathrm{C}(a|b) < k$, найдётся слово p длины k, для которого одновременно верны соотношения $\mathrm{C}(a|b,p) = O(\log n)$ и $\mathrm{C}(p|a) = O(\log n)$, где $n = \mathrm{C}(a)$. Иными словами, среди всех описаний a при известном b найдётся p, простое относительно a. Принцип доказательства такой: рассматривается некоторое семейство хеш-функций, отображающих слова длины n в слова длины k. Слово p строится как образ слова a, поэтому для его описания требуется лишь задать номер хеш-функции. Для того, чтобы a можно было восстановить по b и p, нужно чтобы p могло быть выбрано как хеш-значение у не слишком большого числа слов, имеющих сложность не больше k при условии b. Этого можно добиться, наложив специальные

Edition. — Springer, 2008. — XXIV, 792 p.

[6] M. Sipser. A Complexity Theoretic Approach to Randomness // Proceedings of the 15th Annual ACM Symposium on Theory of Computing. — 1983. — P. 330–335.

[7] An.A. Muchnik. Conditional Complexity and Codes // Theoretical Computer Science. — 2002. — V. 271. №1–2. — P. 97–109.

комбинаторные требования на семейство функций. Сам Ан.А.Мучник использовал свойство экспандера, А.Шень[8] использовал свойство существования онлайн-паросочетаний, а в диссертации показано, что можно рассмотреть свойство экстрактора.

Понятие экстрактора было введено в начале 1990-х годов в работе Н.Нисана и Д.Цукермана[9] как технический инструмент. Неформально говоря, экстрактор — это функция, которая получает на вход две «не очень хорошие» случайные величины и превращает их в «почти случайные». В последующие годы было разработано множество конструкций и приложений экстракторов[10,11,12], к которым диссертация добавляет ещё одно.

Ан.А.Мучник также доказал вариацию теоремы для двух условий. Она гласит, что для любых слов a, b и c, таких что $\mathrm{C}(a|b) < k$ и $\mathrm{C}(a|c) < l$, где $l \leqslant k$, найдётся слово p длины k, такое что для него верны соотношения $\mathrm{C}(a|b,p) = O(\log n)$ и $\mathrm{C}(p|a) = O(\log n)$, а для его начала q длины l верно $\mathrm{C}(a|c,q) = O(\log n)$. Можно обобщить эту теорему на любое константное и даже полиномиальное число условий. Иными словами, программы, превращающие разные слова в a, не только просты относительно a, но и являются префиксами одной и той же длинной программы с точностью до небольших добавок логарифмической длины. Принцип доказательства этой теоремы такой же, но комбинаторное условие на семейство функций будет несколько другим.

Вторым примером параллелизма между сложностными и комбинаторными утверждениями являются теоремы о существовании колмогоровских экстракторов (обычных и усиленных). Колмогоровским экстрактором называется функция двух аргументов, значение которой имеет меньший *дефект случайности*, т.е. разность между длиной и сложностью, чем каждый из её аргументов. Если «условный дефект случай-

[8] A. Shen. Combinatorial Proof of Muchnik's Theorem // M. Hutter, W. Merkle, P. Vitanyi, eds. Kolmogorov complexity and applications. — 2006. — Dagstuhl Seminar Proceedings 06051. — http://drops.dagstuhl.de/opus/volltexte/2006/625.

[9] N. Nisan, D. Zuckerman. Randomness is Linear in Space // Journal of Computer and System Sciences. — 1996. — V. 52. №1. — P. 43–52.

[10] R. Shaltiel. Recent Developments in Explicit Constructions of Extractors // Current and Trends in Theoretical Computer Science: The Challenge of the New Century. Volume 1: Algorithms and Complexity. — World Scientific, 2002. — P. 189–228.

[11] R. Shaltiel. An Introduction to Randomness Extractors // Proceedings of the International Conference on Automata, Languages and Programming. — 2011. — LNCS V. 6756. №2. — P. 21–41.

[12] S. Vadhan. Pseudorandomness. — Draft survey/monograph. — 2012. — http://people.seas.harvard.edu/ salil/pseudorandomness/

ности», т.е. разность между длиной и условной сложностью значения функции при условии одного из аргументов, также уменьшается, то такая функция называется усиленным колмогоровским экстрактором. Колмогоровские экстракторы были определены Дж. Хичкоком и соавторами[13,14] и затем изучались М. Зимандом[15,16,17]. Зиманд определил комбинаторные понятия пёстрых и равномерно пёстрых таблиц, доказал их существование и показал, что они являются колмогоровскими экстракторами (обычными и усиленными).

Основной темой диссертации является распространение известных результатов о колмогоровской сложности на сложность с ограничением на ресурсы. Достижений в этом направлении не так много. Можно отметить результат Т. Ли и А.Е. Ромащенко[18] о симметрии информации, а также работу Г. Бюрмана, Т. Ли и Д. ван Мелкебеека о сжатии языков.[19] Для сложности с ограничением на ресурсы становится важна модель вычислений: использование недетерминизма и/или случайности может существенно уменьшить сложность. В частности, изучают CAM-сложность для недетерминированных вероятностных вычислений. Эта модель была введена Л. Бабаем[20] и получила метафорическое название игр Артура–Мерлина. Метафора объясняется так: вычисления проводит Артур, беседующий с магом Мерлином. Артур может бросать монетку (т.е. получать случайные биты) и совершать полиномиальные вычисления. Мерлин может вычислять любые функции (даже не вычислимые алгоритмически) и мгновенно узнаёт случайные биты Артура. Взаимо-

[13] J.M. Hitchcock, A. Pavan, N.V. Vinodchandran. Kolmogorov Complexity in Randomness Extraction // ACM Transactions on Computation Theory. — 2011. — V. 3. №1. — P. 1:1–1:15.

[14] L. Fortnow, J. Hitchcock, A. Pavan, N.V. Vinodchandran, F. Wang. Extracting Kolmogorov Complexity with Applications to Dimension Zero-One Laws // Information and Computation. — 2011. — V. 209. №4. — P. 627–636.

[15] M. Zimand. Two Sources are Better than One for Increasing the Kolmogorov Complexity of Infinite Sequences // Proceedings of the 3rd Computer Science Symposium in Russia. — 2008. — LNCS, V. 5010. — P. 326–338.

[16] M. Zimand. Extracting the Kolmogorov Complexity of Strings and Sequences from Sources with Limited Independence // Proceedings the 26th Symposium on Theoretical Aspects of Computer Science. — 2009. — P. 697–708.

[17] M. Zimand. Impossibility of Independence Amplification in Kolmogorov Complexity Theory // Proceedings of the 35th International Symposium on Mathematical Foundations of Computer Science. — 2010. — LNCS, V. 6281. — P. 701–712.

[18] T. Lee, A. Romashchenko. Resource-Bounded Symmetry of Information Revisited // Theoretical Computer Science. — 2005. — V. 345. № 2–3. — P. 386–405.

[19] H. Buhrman, T. Lee, D. van Melkebeek. Language Compression and Pseudorandom Generators // Computational Complexity. — 2005. — V. 14. — P. 247–274.

[20] L. Babai. Trading Group Theory for Randomness // Proceedings of the 17th annual ACM symposium on Theory of computing. — 1985. — P. 421–429.

действие устроено так: сначала Артур кидает монетку необходимое число раз, затем Мерлин узнаёт результаты и посылает некоторое сообщение, затем Артур проводит полиномиальные вычисления и выдаёт ответ. Ответом может быть либо двоичное слово, либо символ ошибки $\perp$. Слово x называется результатом работы программы, если с вероятностью хотя бы $\frac{2}{3}$ одновременно выполнены два условия: во-первых, Артур выдаст x при некоторых сообщениях Мерлина; во-вторых, при любых других сообщениях Мерлина Артур выдаст либо то же x, либо $\perp$.

Важным техническим инструментом, используемым в диссертации, является генератор псевдо-случайных чисел Нисана–Вигдерсона.[21,22] Этот генератор «обманывает» все схемы из функциональных элементов полиномиального размера и константной глубины, т.е. такие схемы выдают асимптотически одинаковый результат на случайном слове и на случайном значении генератора. В диссертации генератор используется «наивным» образом, сродни работе А.Е. Ромащенко:[23] вместо случайного комбинаторного объекта рассматривается псевдослучайный, что существенно уменьшает область перебора и позволяет доказать теоремы для ограниченных ресурсов.

Цели и задачи работы

Основной задачей работы является изучение связей между комбинаторными конструкциями и утверждениями о колмогоровской сложности и использование этих связей для распространения известных теорем о колмогоровской сложности на сложность с ограничением на вычислительные ресурсы.

Основные результаты

Работа содержит четыре основных результата. Результаты являются новыми, получены автором самостоятельно и состоят в следующем.

Во-первых, установлена связь теоремы Мучника об условном кодировании и теории экстракторов. С использованием леммы Бюрмана–

[21] N. Nisan. Pseudorandom Bits for Constant Depth Circuits // Combinatorica. — 1991. — V. 11. №1. P. 63–70.

[22] N. Nisan, A. Wigderson. Hardness vs. Randomness. // Journal of Computer and System Sciences. — 1994. — V. 49. — P. 149–167.

[23] A.E. Romashchenko. Pseudo-Random Graphs and Bit Probe Schemes with One-Sided Error // Theory of Computing Systems. — 2014. — V. 55. №2. — P. 313–329.

Фортноу–Лаплант[24] доказано, что комбинаторное свойство экстрактора позволяет доказать теорему Мучника об условном кодировании. Также показано, что аналогичной техникой можно получить теорему Мучника для нескольких условий (вплоть до полиномиального числа). Для этого используются префиксные экстракторы.

Во-вторых, доказаны аналоги теоремы Мучника для сложности с ограничением на память. Применяются две техники: использование явных экстракторов и использование псевдослучайных генераторов. Комбинацией двух подходов получен аналог теоремы для логарифмической точности при любом ограничении на память. Также доказаны аналоги теоремы Мучника для нескольких источников. Применение техники генераторов псевдослучайных чисел позволило доказать теорему не только для полиномиального, но и для экспоненциального числа условий при полиномиальном ограничении на память.

В-третьих, доказан аналог теоремы Мучника для CAM-сложности с ограничением на время. Здесь кодирование осуществляется обычным полиномиальным алгоритмом, а декодирование происходит при помощи алгоритма из класса AM.

В-четвёртых, определены и построены обычные и усиленные колмогоровские экстракторы с ограничением на память. Конструкции Зиманда, доказывающие существование таких экстракторов с оптимальными параметрами, упрощены и переложены для новых определений. Использована разработанная при доказательстве аналогов теоремы Мучника техника псевдослучайных генераторов.

Основные методы исследования

В работе использованы комбинаторные методы доказательства утверждений о колмогоровской сложности. Использованы такие комбинаторные конструкции, как экстракторы (в частности, экстрактор Тревисана) и генераторы псевдослучайных чисел.

[24] H. Buhrman, L. Fortnow, S. Laplante. Resource bounded Kolmogorov complexity revisited // SIAM Journal on Computing. — 2002. — V. 31. №3. — P. 887–905.

Теоретическая и практическая ценность

Работа имеет теоретический характер. Полученные результаты представляют интерес для специалистов по теории колмогоровской сложности. Методы, разработанные автором в диссертационной работе, были использованы[25,26] и могут быть использованы в дальнейшем для доказательства других результатов о колмогоровской сложности с ограничением на вычислительные ресурсы.

Апробация работы

Основные результаты, полученные в работе, были изложены на международных конференциях "Computer Science in Russia" в Новосибирске (18–23 августа 2009 г.), в Санкт-Петербурге (14–18 июня 2011 г.) и в Нижнем Новгороде (3–7 июля 2012 г.)

Кроме того, результаты были доложены на следующих международных и всероссийских научных конференциях и семинарах:

- 8-ая международная конференция по вычислимости, сложности и случайности (CCR), Москва, 21–23 сентября 2013 г.
- 56-я научная конференция МФТИ, секция дискретной математики, Долгопрудный, 25–30 ноября 2013 г.
- 55-я научная конференция МФТИ, секция дискретной математики, Долгопрудный, 19–25 ноября 2012 г.
- Колмогоровский семинар по сложности вычислений и сложности определений, механико-математический факультет МГУ им. М.В.Ломоносова, 2006–2012 гг.
- Кафедральный семинар кафедры дискретной математики факультета инноваций и высоких технологий МФТИ(ГУ), 2009–2013 гг.
- Семинар по дискретной математике Петербургского отделения математического института им. В.А.Стеклова РАН, 2012 г.
- Семинар Давида Гамарника в Массачусетском Технологическом институте (США), 2012 г.

[25]M. Zimand. Symmetry of Information and Bounds on Nonuniform Randomness Extraction via Kolmogorov Extractors // Proceedings of the 26th IEEE Conference in Computational Complexity. — 2011. — P. 148–156.

[26]M. Zimand. On the Optimal Compression of Sets in PSPACE // Proceedings of the 18th International Symposium on Fundamentals of Computation Theory. — 2011. — P. 65–77.

Публикации

Основные результаты диссертации были опубликованы в сборниках трудов международных конференций “Computer Science in Russia” [2; 4; 6] (серия Lecture Notes in Computer Science, входит в систему Scopus) и специальных выпусках журнала Theory of Computing Systems [1; 3; 5] (входит в систему Web of Science). В совместных статьях [5] и [6] соискателю принадлежат метод доказательства теоремы Мучника при помощи экстракторов и теорема для CAM-сложности. Упрощённое доказательство одной из теорем было опубликовано в материалах 55-ой научной конференции МФТИ [7].

Структура диссертации

Диссертация состоит из введения, обзора основных понятий, четырёх глав с изложением основных результатов работы, заключения и списка литературы из 65 наименований. Диссертация содержит 18 иллюстраций и 6 алгоритмов, записанных на псевдокоде. Общий объём диссертации составляет 128 страниц, включая 118 страниц основного текста.

Основное содержание работы

Глава 1 является введением диссертации. Она содержит описание актуальности темы, цели работы, список основных результатов, сведении об апробации работы и список используемых обозначений.

В **главе 2** приведён обзор основных понятий и результатов, используемых в диссертации. Освещены такие области, как колмогоровская сложность, экстракторы, колмогоровские экстракторы, схемы из функциональных элементов, генераторы псевдослучайных чисел. Также приведены формулировки важных технических результатов: вычислительной XOR-леммы и неравенства Чернова–Хёффдинга. В автореферате приведены только определения, используемые в формулировках соответствующих теорем.

В **главе 3** излагается новое доказательство теоремы Мучника об условном кодировании при помощи экстракторов.

Определение 1. Пусть U — универсальная вычислимая функция двух аргументов. *Условной колмогоровской сложностью* слова x относи-

тельно слова y называется длина кратчайшего слова p, такого что $U(p, y) = x$. Это число будем обозначать $\mathrm{C}(x|y)$. *Безусловной колмогоровской сложностью* слова x называется $\mathrm{C}(x) = \mathrm{C}(x|\varepsilon)$, где ε — пустое слово.

Неформально теорема Мучника гласит следующее: среди всех программ, переводящих b в a и имеющих близкую к минимальной длину, найдётся программа p, простая относительно a. Формально теорема формулируется так:

Теорема 2. *Пусть даны слова из нулей и единиц a и b и числа n и k, такие что $\mathrm{C}(a) < n$ и $\mathrm{C}(a|b) < k$. Тогда существует такое слово p, что:*

- $\mathrm{C}(a|p, b) \leqslant O(\log n)$;
- $|p| \leqslant k + O(\log n)$;
- $\mathrm{C}(p|a) \leqslant O(\log n)$,

где константы в $O(\cdot)$-обозначениях не зависят от a, b, n, k, но могут зависеть от выбранной универсальной функции в определении C.

Идея доказательства состоит в следующем: значение p выбирается среди значений хеш-функций из некоторого семейства. Чтобы была мала сложность $\mathrm{C}(p|a)$, нужно чтобы семейство было небольшим. Чтобы была мала сложность $\mathrm{C}(a|p, b)$, нужно чтобы для хеш-функции было немного коллизий в множестве $S_b = \{x \mid \mathrm{C}(x|b) < k\}$ при любом b. В диссертации показано, что в качестве такого семейства функций можно взять любой экстрактор. (Если рассмотреть функцию двух аргументов как семейство функций первого аргумента, проиндексированное значениями второго аргумента).

Определение 3. Функция $\mathrm{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ называется (k, ε)-*экстрактором*, если для любых независимых случайных величин ξ на $\{0,1\}^n$ и η на $\{0,1\}^d$, таких что $\Pr[\xi = x] \leqslant 2^{-k}$ при любом фиксированном x, а η распределена равномерно, и любого $S \subset \{0,1\}^m$ выполнено $\left|\Pr[\mathrm{Ext}(\xi, \eta)] - \frac{|S|}{2^m}\right| < \varepsilon$.

Ан.А. Мучник также доказал расширение теоремы об условном кодировании на случай нескольких условий:

Теорема 4. *Пусть даны двоичные слова a, b и c и числа n, k и l, для которых выполнено* $\mathrm{C}(a) < n$, $\mathrm{C}(a|b) < k$ *и* $\mathrm{C}(a|c) < l$. *Тогда существуют слова p длины $k + O(\log n)$ и q длины $l + O(\log n$, одно из которых является началом другого, для которых величины* $\mathrm{C}(a|p, b)$, $\mathrm{C}(a|q, c)$, $\mathrm{C}(p|a)$ *и* $\mathrm{C}(q|a)$ *имеют порядок* $O(\log n)$.

В диссертации показано, что этот результат также можно доказать при помощи экстракторов. Для этого вводится новое понятие префиксного экстрактора и вероятностным методом[27] доказывается существование таких объектов.

Определение 5. Будем говорить, что (k, ε)-экстрактор $\mathrm{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, где $m \geqslant k$, является *префиксным*, если для любого $i \leqslant k$ его префикс длины $m - i$ является $(k - i, \varepsilon)$-экстрактором. Под префиксом понимается функция $\mathrm{Ext}_i\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{m-i}$, полученная из исходной отрезанием последних i битов.

В **главе 4** формулируются и доказываются различные аналоги теоремы Мучника для сложности с ограничением на память.

Определение 6. Колмогоровской сложностью слова x относительно слова y *с ограничением на память* s называется длина кратчайшего слова p, такого что $U(p, y) = x$ и при этом $U(p, y)$ использует не больше s ячеек памяти. Это число будем обозначать $\mathrm{C}^s(x|y)$.

Благодаря тому, что существуют явные конструкции экстракторов, становится возможным использование разработанной техники для расширения теоремы на случай сложности с ограничением на память. В диссертации доказывается следующая теорема:

Теорема 7. *Пусть даны двоичные слова a и b, а также числа n, k и s, для которых верно* $\mathrm{C}^s(a) < n$ *и* $\mathrm{C}^s(a|b) < k$. *Пусть числа d и q таковы, что при любом $l \leqslant k$ существует $(l,$ 0.25$)$-экстрактор* $\mathrm{Ext}_l\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^l$, *вычислимый на памяти q (для некоторой заранее фиксированной универсальной многоленточной машины Тьюринга). Тогда существует такое слово p, что:*

- $\mathrm{C}^{O(s+q+n^2)}(a|p, b) \leqslant d + O(\log n)$;

[27] Н. Алон, Дж. Спенсер. Вероятностный метод в комбинаторике. — М.: БИНОМ, 2011. — 320 с.

- $|p| \leqslant k + O(\log n)$;
- $\mathrm{C}^{O(q)}(p|a) \leqslant d + O(\log n)$.

С использованием наилучших известных конструкций экстракторов[28] можно получить безусловное следствие, в котором в правых частях неравенств вместо $d + O(\log n)$ стоят $O(\log^3 n)$.

Далее, для улучшения результата применяется техника «наивной дерандомизации». Идея состоит в следующем: нужно найти хеш-функцию, имеющую мало коллизий внутри каждого множества $S_{b,s} = \{x \mid \mathrm{C}^s(x|b) < k\}$. Множеств такого вида немного относительно числа всех множеств размера 2^k: экспонента вместо двойной экспоненты. В то же время конструкция с экстракторами гарантирует малость числа коллизий для всех множеств размера 2^k. В работе показано, как можно обойтись лишь одной функцией, полученной в результате работы генератора псевдослучайных чисел Нисана–Вигдерсона. Для этого нужно совершить такую последовательность шагов:

1. Доказать, что случайная функция удовлетворяет условию малого числа коллизий с отделённой от нуля вероятностью;
2. Доказать, что свойство малости числа коллизий распознаётся схемой из функциональных элементов полиномиального размера и константной глубины;
3. Воспользоваться свойством генератора Нисана–Вигдерсона «обманывать» все такие схемы и сделать вывод, что среди его значений встречается код нужной функции;
4. Показать, что нужный аргумент генератора можно найти на полиномиальной памяти.

В результате доказывется следующая теорема:

Теорема 8. *Пусть даны двоичные слова a и b, а также числа n, k и s, для которых верно $|b| < n$, $\mathrm{C}^s(a) < n$ и $\mathrm{C}^s(a|b) < k$. Тогда существует такое слово p, что:*

- $\mathrm{C}^{O(s)+\mathrm{poly}(n)}(a|p, b) \leqslant O(\log\log s) + O(\log n)$;
- $|p| \leqslant k + O(\log n)$;

[28] R. Raz, O. Reingold, S. Vadhan. Extracting All the Randomness and Reducing the Error in Trevisan's Extractor // Proceedings of the 30th Annual ACM Symposium on the Theory of Computing. — 1999. — P. 149–158.

- $\mathrm{C}^{O(s)+\mathrm{poly}(n)}(p|a) \leqslant O(\log\log s) + O(\log n)$.

Комбинацией двух подходов получается безусловная теорема, в которой все невязки равны $O(\log n)$.

Теорема для двух условий также распространяется на сложность с ограничением на память. Как и для одного условия, можно действовать двумя способами: использовать явную конструкцию префиксного экстрактора или использовать метод «наивной дерандомизации». Комбинацией двух подходов получена следующая теорема.

Теорема 9. *Пусть даны двоичные слова a, b и c, а также числа n, k, l и s, для которых выполнено $|b| < n$, $|c| < n$, $\mathrm{C}^s(a) < n$, $\mathrm{C}^s(a|b) < k$ и $\mathrm{C}^s(a|c) < l$. Тогда существуют слова p и q, одно из которых является началом другого, для которых выполнены неравенства:*

- $\mathrm{C}^{O(s)+\mathrm{poly}(n)}(a|p,b) \leqslant O(\log n)$;
- $\mathrm{C}^{O(s)+\mathrm{poly}(n)}(a|q,c) \leqslant O(\log n)$;
- $|p| \leqslant k + O(\log n)$;
- $|q| \leqslant l + O(\log n)$;
- $\mathrm{C}^{O(s)+\mathrm{poly}(n)}(p|a) \leqslant O(\log n)$;
- $\mathrm{C}^{O(s)+\mathrm{poly}(n)}(q|a) \leqslant O(\log n)$.

Более того, метод наивной дерандомизации позволяет распространить теорему не только на полиномиальное, но и на экспоненциальное число условий.

Теорема 10. *Пусть даны числа n, s, r и $k_1, \ldots, k_r$ и двоичные слова a, b_1, $\ldots$, b_r, такие что $\mathrm{C}^s(a) < n$, $|b_i| < \mathrm{poly}(n)$ и $\mathrm{C}^s(a|b_i) < k_i$ при всех $i = 1, \ldots, r$. Тогда существуют слова p_1, $\ldots$, p_r, для которых выполнены следующие условия:*

- *Все p_i являются префиксами одного и того же слова;*
- $\mathrm{C}^{O(s)+\mathrm{poly}(n)}(a|p_i, b_i) \leqslant O(\log\log s) + O(\log n)$ *при всех* $i = 1, \ldots, r$;
- $|p_i| \leqslant k_i + O(\log n)$ *при всех* $i = 1, \ldots, r$;
- $\mathrm{C}^{O(s)+\mathrm{poly}(n)}(p_i|a) \leqslant O(\log\log s) + O(\log n)$ *при всех* $i = 1, \ldots, r$.

В **главе 5** доказывается вариант теоремы Мучника для CAM-сложности. Идея состоит в том, чтобы вместо произвольного явного экс-

трактора взять экстрактор Тревисана[29], который позволяет быстро осуществлять и кодирование, и декодирование. Похожая техника использовалась в работе Г. Бюрмана, Т. Ли и Д. ван Мелкебеека о сжатии языков.[30]

CAM-сложность определяется в модели вычислений Артура–Мерлина, сочетающей в себе недетерминизм и случайность. Под W будем понимать универсальную функцию четырёх аргументов, из которых первый понимается как текст программы, второй — как аргумент, третий — как сертификат и четвёртый — как случайные биты. Функция W возвращает либо двоичное слово, либо специальный символ ошибки $\perp$.

Определение 11. *Сложностью Артура-Мерлина* слова x относительно слова y за время t называется длина кратчайшего слова p, такого что $\Pr_r\left[\exists q W(p,y,q,r) = x \text{ и } \forall q W(p,y,q,r) \in \{x, \perp\}\right] > \frac{2}{3}$, при этом время работы $W(p,y,q,r)$ не превосходит t при всех q и r. Это число будем обозначать через $\mathrm{CAM}^t(x|y)$.

В работе доказана следующая теорема:

Теорема 12. *Для любого полинома $t_1(n)$ существует полином $t_2(n)$, для которого выполнено следующее свойство. Пусть даны числа n и k, слово a длины меньше n и слово b, такие что $\mathrm{C}^{t_1(n)}(a|b) < k$. Тогда существует слово p длины k, такое что сложности $\mathrm{CAM}^{t_2(n)}(a|b,p)$ и $\mathrm{C}^{t_2(n)}(p|a)$ имеют порядок $O(\log^3 n)$.*

Также в диссертации формулируется гипотеза об аналогичном утверждении для нескольких условий и анализируются препятствия к расширению конструкции на этот случай.

Наконец, в **главе 6** техника «наивной дерандомизации» применяется к теории колмогоровских экстракторов, в результате доказывается теорема о существовании оптимальных колмогоровских экстракторов с ограничением на память.

Теорема 13. *Существует полином $p(n)$, такой что для любой функции $s(n) > p(n)$, конструируемой по памяти, и любых функций $1 < k(n) < n$ и $1 < \delta(n) < k(n) - O(\log n)$, вычислимых на памяти*

[29] L. Trevisan. Construction of Extractors Using Pseudo-Random Generators // Journal of the ACM. — 2001. — V. 48. №4. — P. 860–879.

[30] H. Buhrman, T. Lee, D. van Melkebeek. Language Compression...

$s(n)$, существует вычислимое на памяти $O(s(n))$ семейство функций $\mathrm{KExt}_n\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^{m(n)}$, *где* $m = 2k(n) - O(\log n)$, *такое что для некоторой константы* $\mu > 1$ *для всех* n *и всех слов* x *и* y *длины* n *из условий* $\mathrm{C}^s(x) > k(n)$, $\mathrm{C}^s(y) > k(n)$ *и* $\mathrm{C}^{\mu s}(x,y) > \mathrm{C}^s(x) + \mathrm{C}^s(y) - \delta(n)$ *следует, что* $\mathrm{C}^s(\mathrm{KExt}_n(x,y)) > m(n) - \delta(n) - O(\log n)$. *Также существует аналогичное семейство функций* SKExt_n *для* $m = k(n) - O(\log n)$, *такое что при тех же условиях на* x *и* y *выполнено* $\mathrm{C}^s(\mathrm{SKExt}_n(x,y)|x) > m(n) - \delta(n) - O(\log n)$ *и* $\mathrm{C}^s(\mathrm{SKExt}_n(x,y)|y) > m(n) - \delta(n) - O(\log n)$.

Функция KExt_n *называется колмогоровским экстрактором, а функция* SKExt_n *— усиленным колмогоровским экстрактором.*

Следуя работам М. Зиманда о колмогоровских экстракторах для сложности без ограничений[31,32], в диссертации вводятся понятия пёстрых и равномерно пёстрых таблиц. Затем они модифицируются для случая сложности с ограничением на память, и применяется описанная в главе 4 техника наивной дерандомизации. Также доказывается вариант теоремы для малых значений s: в этом случае ограничение μs заменяется на $\mu s + \mathrm{poly}(n)$.

Заключение

А.Н. Колмогоров[33] говорил о трёх подходах к понятию «количество информации»: комбинаторном, вероятностном и алгоритмическом. Каждый из этих подходов представляет интерес и довольно подробно изучен, но особенно интересны соотношения между разными подходами. В диссертации установлены новые связи между комбинаторными и алгоритмическими утверждениями и разработан новый способ распространения утверждений об обычной колмогоровской сложности на сложность с ограничением на ресурсы. Однако множество вопросов остаются нерешёнными. Прежде всего: каковы пределы этого подхода? Какие утверждения о колмогоровской сложности можно переложить для сложности с ограничением на ресурсы посредством комбинаторных утверждений? Насколько универсален метод «наивной дерандомизации»? Можно ли

[31] M. Zimand. Extracting the Kolmogorov Complexity...
[32] M. Zimand. Impossibility of Independence Amplification...
[33] А.Н. Колмогоров. Три подхода...

доказать какую-либо общую теорему на этот счёт? Эти вопросы являются предметом дальнейшего изучения.

Ещё одно важное направление развития — построение явных конструкций использованных в диссертации комбинаторных объектов. Например, построение явных экстракторов с оптимальными параметрами стало бы прорывом сразу во многих областях. Было бы интересно построить экстрактор с оптимальными параметрами, вычислимый на полиномиальной памяти: тогда бы для соответствующего аналога теоремы Мучника не нужно было бы наивной дерандомизации. Возможно, наоборот, методом наивной дерандомизации можно построить такой экстрактор. Можно ли при помощи каких-либо явных экстракторов доказать теорему Мучника для обычной сложности с ограничением на память вместо CAM-сложности? Возможно, это как-то связано со стандартными теоретико-сложностными предположениеми? Также отличным результатом стало бы построение полиномиально вычислимых пёстрых таблиц: возможно, это позволило бы доказать существование колмогоровских экстракторов для сложности с ограничением на время (хотя бы для CAM-сложности). Открытым вопросом остаётся справедливость аналога теоремы Мучника с двумя условиями для CAM-сложности.

Можно сказать, что теория колмогоровской сложности с ограничением на ресурсы всё ещё находится в стадии накопления фактов. Довести количество известных результатов до критической массы и уложить их в стройную теорию остаётся задачей будущих исследований.

Благодарности

Прежде всего, автор выражает глубокую благодарность своему научному руководителю Николаю Константиновичу Верещагину за постановку задачи о колмогоровских экстракторах, постоянное внимание и поддержку в работе. Автор благодарит своих соавторов и наставников Андрея Ромащенко и Александра Шеня за знакомство с тематикой, постановку задачи о применении экстракторов к задаче условного кодирования, многочисленные обсуждения, советы и конструктивную критику по тематике диссертации. Автор выражает отдельную благодарность Андрею Альбертовичу Мучнику (1958–2007) за обсуждение результатов автора, развивающих идеи Андрея Альбертовича. Безвременная кончи-

на Андрея Альбертовича стала тяжёлой утратой для российской науки.

Автор благодарит своих коллег Виктора Булатова, Эдуарда Гирша, Брюно Дюрана (Bruno Durand), Илью Ирхина, Дмитрия Ицыксона, Руслана Ишкуватова, Юрия Притыкина, Михаила Раскина и Андрея Румянцева за обсуждение отдельных разделов работы. Автор благодарит Ронена Шалтиела (Ronen Shaltiel) за полезное замечание, позволившее существенно упростить доказательство одного из утверждений. Также автор благодарит многочисленных анонимных рецензентов, сделавших немало полезных замечаний по текстам статей автора, и участников семинаров в МГУ и МФТИ за внимание и интерес к докладам автора.

Автор благодарит оргкомитет Турнира Городов и лично Сергея Дориченко за включение в состав варианта осеннего тура 2005 года задачи, возникшей в ходе работы над результатами диссертации.

Наконец, автор глубоко признателен своей семье за постоянную поддержку.

Работы автора по теме диссертации

1. **Musatov, D.V.** On Extracting Space-Bounded Kolmogorov Complexity / D.V. Musatov. // Theory of Computing Systems. — 2014. — DOI 10.1007/s00224-014-9563-7.

2. **Musatov, D.V.** Space-Bounded Kolmogorov Extractors / D.V. Musatov // Proceedings of the 7th Computer Science Symposium in Russia. — 2012. — LNCS, Vol. 7353. — P. 266–277.

3. **Musatov, D.V.** Improving the Space-Bounded Version of Muchnik's Conditional Complexity Theorem via "Naive" Derandomization / D.V. Musatov // Theory of Computing Systems. — 2014. — Vol. 55, no. 2. — P. 299–312.

4. **Musatov, D.V.** Improving the Space-Bounded Version of Muchnik's Conditional Complexity Theorem via "Naive" Derandomization / D.V. Musatov // Proceedings of the 6th Computer Science Symposium in Russia. — 2011. — LNCS, Vol. 6651. — P. 64–76.

5. **Musatov, D.V.** Variations on Muchnik's Conditional Complexity Theorem / D.V. Musatov, A.E. Romashchenko, A. Shen // Theory of Computing Systems. — 2011. — Vol. 49, no. 2. — P. 227–245.

   Соискателю принадлежат доказательства всех теорем из раздела 3.

6. **Musatov, D.V.** Variations on Muchnik's Conditional Complexity Theorem / D.V. Musatov, A.E. Romashchenko, A. Shen // Proceedings of the 4th Computer Science Symposium in Russia. — 2009. — LNCS, Vol. 5675. — P. 250–262.

   Соискателю принадлежат доказательства всех теорем из раздела 3.

7. **Мусатов Д.В.** Упрощённое доказательство теоремы Мучника об условной колмогоровской сложности с ограничением на память / Д.В. Мусатов // Труды 55-ой научной конференции МФТИ. — М.–Долгопрудный–Жуковский: МФТИ, 2012. — С. 30–31.