ISSN 1684-8853 (print); ISSN 2541-8610 (online)

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

НАУЧНЫЙ ЖУРНАЛ

3(124)/2023

doi:10.31799/1684-8853-2023-3

ISSN 1684-8853 (print); ISSN 2541-8610 (online)

SCIENTIFIC JOURNAL

3(124)/2023

PEER REVIEWED JOURNAL

INFORMATSIONNO-UPRAVLIAIUSHCHIE SISTEMY (INFORMATION AND CONTROL SYSTEMS)

Contact information
The Editorial and Publishing Center, SUAI
67A, Bol'shaya Morskaia, 190000, Saint Petersburg, Russia
Website: http://i-us.ru/en, e-mail: ius.spb@gmail.com
Tel.: +7 - 812 494 70 02

The Journal was registered in the Ministry of Press, Broadcasting and Mass Media of the Russian Federation.
Registration Certificate JD № 77-12412 from April, 19, 2002.
Re-registration in the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR) due to change of the founder A. Vostrikov,
PI № FS 77-82226 from November, 23, 2021

Submitted for publication 10.05.23. Passed for printing 30.06.23. Signed for printing: 03.07.2023.
Format $60\times84_{1/8}$. Phototype CentSchbkCyrill BT. Digital printing.

Layout original is made at the Editorial and Publishing Center, SUAI.
67A, Bol'shaya Morskaia, 190000, Saint Petersburg, Russia
Printed from slides at the Editorial and Publishing Center, SUAI.
67A, Bol'shaya Morskaia, 190000, Saint Petersburg, Russia

The journal is indexed in Scopus.

Free distribution.

doi:10.31799/1684-8853-2023-3 ISSN 1684-8853 (print); ISSN 2541-8610 (online)
НАУЧНЫЙ ЖУРНАЛ

3(124)/2023

РЕЦЕНЗИРУЕМОЕ ИЗДАНИЕ

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

Учредитель
А. А. Востриков
Издатель
Санкт-Петербургский государственный университет аэрокосмического приборостроения

Адрес редакции: 190000, г. Санкт-Петербург,
ул. Большая Морская, д. 67, лит. А, ГУАП, РИЦ
Тел.: (812) 494-70-02, эл. адрес: ius.spb@gmail.com,
сайт: http://i-us.ru

Журнал входит в БД Scopus и в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук.

Сдано в набор 10.05.23. Подписано в печать 30.06.23. Дата выхода в свет: 03.07.2023.
Формат 60×841/8. Гарнитура CentSchbkCyrill BT. Печать цифровая.
Усл. печ. л. 8,5. Уч.-изд. л. 11,7. Тираж 1000 экз (1-й завод 50 экз.). Заказ № 307.
Оригинал-макет изготовлен в редакционно-издательском центре ГУАП.
190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А.
Отпечатано в редакционно-издательском центре ГУАП.
190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А.
Распространяется бесплатно.

Журнал зарегистрирован в Министерстве РФ по делам печати, телерадиовещания и средств массовых коммуникаций.
Свидетельство о регистрации ПИ № 77-12412 от 19 апреля 2002 г.
Перерегистрирован в Роскомнадзоре.
Свидетельство о регистрации ПИ № ФС77-82226 от 23 ноября 2021 г.

УДК 519.614
doi:10.31799/1684-8853-2023-3-2-15
EDN: JQPBFE

Научные статьи
Articles

Максимум детерминанта бициклических матриц с каймой

Н. А. Балонин[a], *доктор техн. наук, профессор, orcid.org/0000-0001-7338-4920, korbendfs@mail.ru*
М. Б. Сергеев[a], *доктор техн. наук, профессор, orcid.org/0000-0002-3845-9277*
[a]*Санкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ*

Введение: *матрицы максимального детерминанта хорошо известны и исследованы для четных порядков 2t и 4t, где их структура чаще всего бициклическая, которую называют адамаровой в случае деления порядка на 4. Нечетные порядки исследованы значительно меньше в силу того, что сложность структуры оптимальных матриц неограниченно возрастает.* ***Цель:*** *заменить заведомо сложный объект гиперобъектом, состоящим из бициклической матрицы с каймой, оптимальным на множестве матриц такой фиксированной структуры. Выявить связь точек Гаусса на сечениях тел вращения с количеством и видами матриц максимума детерминанта с фиксированной структурой для нечетных порядков. Определить верхнюю и нижнюю границы значений максимумов детерминанта для бициклических матриц с каймой.* ***Результаты:*** *уточнена излишне оптимистическая граница Элича – Войтаса на случай матриц фиксированной структуры для порядков 4t + 1 (адамариды) и 4t – 1 (мерсенниды). Кроме границ снизу и сверху, приведены более близкие к значениям детерминантов экстремальных матриц кусочно-гладкие квадратичные аппроксимации. Приведены алгоритмы углубленного поиска матриц расширенного семейства Адамара с использованием орбит и компрессий бинарных последовательностей, а также результаты поиска, подтверждающие приведенные оценки границ.* ***Практическая значимость:*** *матрицы максимального детерминанта имеют практическое значение для задач помехоустойчивого кодирования, сжатия и маскирования видеоинформации.*

Ключевые слова *– точки Гаусса, проблема Гаусса, параболоид на решетке, ортогональные матрицы, матрицы Адамара, матрицы максимального детерминанта, бициклические матрицы с каймой.*

Для цитирования: Балонин Н. А., Сергеев М. Б. Максимум детерминанта бициклических матриц с каймой. *Информационно-управляющие системы*, 2023, № 3, с. 2–15. doi:10.31799/1684-8853-2023-3-2-15, EDN: JQPBFE
For citation: Balonin N. A., Sergeev M. B. Maximum determinant two circulant matrices with border. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 2–15 (In Russian). doi:10.31799/1684-8853-2023-3-2-15, EDN: JQPBFE

Введение

Изучению матриц максимального детерминанта посвящено много исследований [1–9], поскольку они представляют не только значительный научный и соревновательный интерес, но и практический при обработке информации [10, 11].

Матрицы максимального детерминанта – квадратные матрицы произвольных порядков n с элементами 1 и –1, сходные с матрицами Адамара [12] порядков, кратных 4, но не обязательно ортогональные.

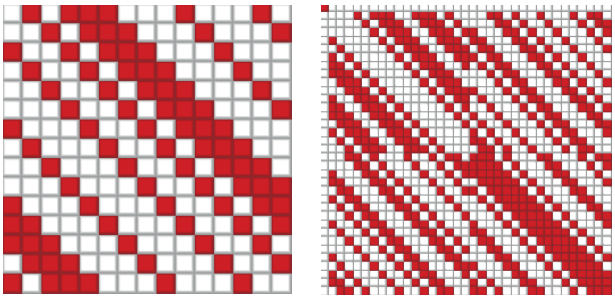
На ряде порядков, кратных (помимо степеней 2) простым числам 5 и 13, открытых Марком Голеем [13, 14], матрицы Адамара упрощаются до бициклической конструкции вида $\mathbf{T}=\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^{\mathrm{T}} & -\mathbf{A}^{\mathrm{T}}\end{pmatrix}$, где $\mathbf{A}$ и $\mathbf{B}$ – циклические матрицы, заданные верхними строками – последовательностями a и b.

Легко заметить отсутствие $\mathbf{T}$ порядков, кратных четным степеням чисел Мерсенна 3, 7: нет бициклов Адамара размера $36 = 4 \times 9$ и $72 = 8 \times 9$, однако есть решение на порядке $144 = 16 \times 9$. Это наводит на мысль, что мерсенновы порядки не критичны, но требуют их увеличения степенями двойки для того, чтобы стала возможной ортогонализация строк и столбцов выбором последовательностей a и b.

Различие чисел Ферма и Мерсенна или последовательностей вида $4t + 1$ и $4t – 1$, которым они соответственно принадлежат, сказывается на матрицах Адамара, содержащих их в качестве множителей порядков: существование бициклических матриц во втором случае гарантировано дополнительным пространством для ортогонализации строк и столбцов. Умножение порядка на степени двойки дает бициклической матрице желаемое качество. Чем больше порядок, тем больше увеличение, что выводит такой прием из ряда интересных для практики.

Это простое наблюдение поясняет свойства не только матриц Адамара, но и матриц максимального детерминанта, которые конструируются добавлением каймы к бициклической матрице $\mathbf{T}$ как математическому объекту. На рис. 1 приведены портреты моноциклической матрицы порядка 15 и бициклической матрицы с каймой порядка 35.

Если нет бициклической матрицы Адамара, нет и возможности построить на ее основе матри-



■ ***Рис. 1.*** Портреты моноциклической и бициклической матриц с каймой

■ ***Fig. 1.*** Monocycle and two circulant matrix portraits

цу максимума детерминанта. Более того, внешний вид такой строго экстремальной по детерминанту матрицы неограниченно усложняется с ростом порядка [1, 2, 15, 16]. Возможная альтернатива состоит в оптимизации детерминанта на фиксированной структуре в виде бициклической матрицы с каймой.

Для голеевских пар a и b существует алгоритм перемножения их между собой, предложенный Тюриным [17] и развитый в работе [18] для умножений не голеевских пар. У последних свойства быть множителями пар нет. Первая отличная от голеевских бициклическая матрица имеет порядок $68 = 4 \times 17$, вторая — 100 [13]. Такое повышение основы до 17 говорит о том, что ортогональность возможна на специфичном порядке, а не только на порядках с простыми основами 2, 5 и 13.

На множестве экстремальных по детерминанту матриц четных порядков, не кратных 4, свойство образовывать пары сохраняется. Это так называемые матрицы оптимального дизайна OD [14] — не ортогональные матрицы, но которые можно сделать ортогональными за счет уменьшения значения отрицательного элемента до некоторого $-b$, где $|b| \le 1$. Их можно отнести к расширенному семейству матриц Адамара. Эти порядки распадаются на разрешимые и неразрешимые на заданной структуре по признаку простоты числа, характеризующего порядок, его повышенная величина уже не имеет значения.

Отсюда видно, что максимальные по детерминанту матрицы нечетных порядков распадаются на два различающихся между собой семейства.

Определение 1. Семейство матриц на порядках $4t + 1$, получаемых из бициклических матриц (в ряде случаев — матриц Адамара) добавлением каймы, называют «адамариды».

Определение 2. Семейство матриц на порядках $4t - 1$, получаемых из бициклических матриц (в ряде случаев – порядков, равных числам Мерсенна) добавлением каймы, называют «мерсенниды».

Бициклические матрицы с каймой не являются строго оптимальными по детерминанту, исключая порядки, равные первым пяти числам Ферма [19] 3, 5, 17, 257, 65 537. Описываемая нами проблема неразрешимости экстремальной задачи матрицей общего вида неизвестной структуры сходна с известным результатом Матиясевича о неразрешимости диофантовых уравнений формулами, которые можно было бы как-либо систематизировать и описать [20].

Целями работы являются:

— исследование бициклических матриц с каймой, имеющих максимальный детерминант на множестве таких матриц;

— определение связи точек Гаусса на сечениях тел вращения с количеством и видами бициклических матриц с каймой нечетных порядков, имеющих максимум детерминанта;

— определение уточненных границ значений максимумов детерминанта бициклических матриц с каймой.

Уравнения орнаментов и инварианты матриц

Рост неопределенности вида строго оптимальной матрицы оправдывает переход к более узкому классу матриц, экстремальных на бициклической структуре с каймой. Матрицы Адамара и матрицы OD существуют не всегда, тогда как экстремальные на бициклической структуре с каймой матрицы существуют для всех нечетных порядков.

Определение 3. Матрица $[\mathbf{T}] = \begin{pmatrix} -1 & \mathbf{e}^{\mathrm{T}} \\ \mathbf{e} & \mathbf{T} \end{pmatrix}$ называется экстремальной бициклической матрицей (все элементы вектора $\mathbf{e}$ равны 1), если она имеет максимум детерминанта на множестве матриц такой структуры, т. е. является *гиперобъектом* над объектом $\mathbf{T}$.

Кроме экстремальных бициклических матриц, будем выделять так же устроенные матрицы большого детерминанта (МБД) на фиксированной структуре. К ним относятся матрицы, детерминант которых несколько меньше максимального. Чем выше порядок матрицы, тем чаще будут встречаться более доступные, чем экстремальные, МБД.

Экстремальные бициклические матрицы и МБД нечетных порядков дополняют собой матрицы Адамара и OD, хорошо вписываясь в контент задачи на оптимальность.

Бициклические матрицы отличаются инвариантами $k_1 = (v - x)/2$, $k_2 = (v - y)/2$, задающими число -1 в последовательностях a и b длины $v = n/2$. Здесь x и y — координаты точек Гаусса на характеристической окружности — сечении чаши параболоида $x^2 + y^2 = h$ на высоте h [15,

16]. В данном случае $h = n$ для матриц Адамара и $h = 2n - 2$ для матриц OD. Матрицы обоих типов могут и не существовать, например, нет бициклических матриц Адамара порядков 36 и 72. Эти порядки не разрешимы, хотя точки Гаусса имеются [13, 15].

Выход из ситуации, когда нет точек Гаусса для заданного порядка экстремальной бициклической матрицы, прост — они берутся на той высоте h, на которой они существуют. Тогда задача сводится к выяснению: является ли это кольцо среза параболоида ближайшим или оно отстоит подальше.

Другими словами, в отличие от матриц Адамара и OD, инварианты k_1 и k_2 определяются двумя срезами (или порядками), а не одним. Например, на порядке $n = 32$ адамарид отличается от матрицы Адамара тем, что у него $h = 36$. Экстремальная бициклическая матрица порядка 32 не ортогональна и заимствует для своего построения координаты точек Гаусса у соседней окружности с $h = 36$. Если $x^2 + y^2 = 36$, то $x = 0$, $y = 6$ и $k_1 = 8$, $k_2 = 5$. Заметим, отсутствие бициклических матриц Адамара порядка 36 не мешает использовать этот срез для расчета параметров матрицы порядка 32.

Правило. Для адамаридов отстающий порядок среза h связан с порядками регулярных матриц Адамара порядков $4 \times q^2$, где q — целое число. Он начинается таким порядком и отстает от текущего порядка, сменяясь по достижении очередной регулярной точки.

К сожалению, в этой последовательности притяжений встречаются сбои, связанные с негативным влиянием множителей, равных, как уже отмечалось, квадратам чисел Мерсенна. Именно из-за него нет бициклической матрицы Адамара порядка 36.

Для мерсеннидов тоже есть опорные порядки, *кольца Барбы* разрешимы крайними точками Гаусса с одной из координат, равной 1. Они выделены синим цветом на рис. 2 и по смыслу противопоставлены кольцам с точкой посередине, характерным для побочного семейства субоптимальных бициклических матриц.

■ ***Рис. 2.*** Кольца на высотах Барбы $h = 2, 10, 26, 50, 82, ...$
■ ***Fig. 2.*** Circles for Barba heights $h = 2, 10, 26, 50, 82, ...$

Смещение наводящего решение среза легко объясняется тем, что детерминант матрицы с каймой зависит от детерминанта матрицы без каймы и эксцесса S [21] — суммы элементов матрицы или, что бывает удобнее, ее инверсии. Чем более разнесены срезы, тем выше эксцесс, компенсирующий падение отмеченного детерминанта. Для ориентации в размерах семейств смежных порядков, сидящих на одних и тех же точках Гаусса, нужно познакомиться с построением системы чисел.

Треугольные и квадратные числа

С целыми числами связаны пары чисел q и $q + 1$, их квадраты q^2 и $(q + 1)^2$, а также их суммы, произведения, половинные произведения и т. п. Треугольные числа T: 0, 1, 3, 6, 10, 15 … — это значения последовательных сумм целых чисел 0, 1, 2, …, q. Квадратные числа Q: 0, 1, 4, 9, 16, 25, 36 … — это квадраты целых чисел.

Эти и другие фигурные числа известны с глубокой древности. Заметим, что $T = q(q + 1)/2$, где q — последнее число суммируемой цепочки. Треугольные числа связаны с произведением $L = q(q + 1)$ и суммой квадратов $B = q^2 + (q + 1)^2$ двух соседних чисел как $L = 2T$ и $B = 4T + 1$. L и B (с точностью до 1) являются четвертью и половиной квадрата величины $8T + 1$. Гаусс активно использовал тот факт, что каждый второй квадрат с точностью до 1 равен треугольному числу, умноженному на 8: $Q = 8T + 1 = 1, 9, 25, 49, …$ и т. п.

Согласно теореме Гаусса любое целое число представимо суммой не более чем трех треугольных чисел. Геометрическая интерпретация теоремы состоит в том, что уравнение сферы $x^2 + y^2 + z^2 = h$, где h — целое число (квадрат радиуса), нелинейным преобразованием координат $x^2 = 8T_x + 1$, $y^2 = 8T_y + 1$, $z^2 = 8T_z + 1$ сводится к уравнению плоскости $T_x + T_y + T_z = (h - 3)/8$.

Заметим, что среди квадратов, которые не входят в последовательность $8T + 1$, с точностью до 1 встречаются числа $Q = F - 1$, где $F = (3), 5, 17, 257, 65\,537, …$ — числа Ферма. Все числа $F = Q + 1$ (для $Q \neq 8T + 1$, $Q = 0, 4, 16, 36, …$ и т. п.) будем называть *фермитами* и, не внося путаницы, можем обозначать той же буквой F, так как все они, за вычетом 1, порождают квадраты, стоящие между квадратами, порожденными треугольными числами. В них входят четные степени двойки, куда попадает число 64 (но не 32), и произведения на них, например, 36 — учетверенный квадрат $4q^2$.

Констатируем: квадраты Q — это 8T + 1 или F – 1, являющиеся соседями. Соседство, в свою очередь, то дополнительное качество чисел, которое отражается на матрицах таких порядков.

Границы детерминантов Барбы и Элича

Спустя три десятилетия после введения Адамаром верхней границы значения детерминанта $n^{n/2}$ для экстремальных матриц произвольной структуры, ее поправил Барба, получив менее оптимистичную оценку детерминанта для матриц нечетных порядков.

Согласно теореме Барбы на границе $(2n - 1)^{1/2}(n - 1)^{(n-1)/2}$ есть рациональные точки (точки Гаусса), достижимые детерминантами целочисленных матриц порядков $B = q^2 + (q + 1)^2$, где числа B — числа Барбы [22].

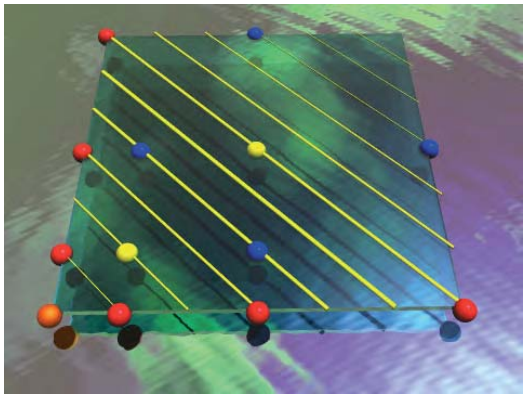
Границы Барбы достигают блочно-составные матрицы из матриц Адамара меньших порядков. В отличие от матриц Адамара порядков Q = F – 1, связанных с числами Ферма и построенных на голеевских парах, матрицы Адамара прочих порядков не имеют общего для всех них алгоритма построения. Рациональные точки, выделенные Барбой, имеют, скорее, значение для теории, а не для практики.

Формула Барбы и следствие из нее в виде семейства матриц заведомо сложных конструкций не подрывает приоритет экстремальных матриц порядков чисел Ферма.

Парадокс состоит в том, что иррациональные детерминанты матриц порядков чисел Ферма не достигают рациональных точек той же границы, поэтому на них длительное время не обращали внимания, хотя это основные сугубо оптимальные (т. е. достигающие строгого максимума) бициклические матрицы. Следующий шаг был сделан спустя полвека написавшими независимые обзоры Эличем и Войтасом [4, 5].

Согласно теореме Элича границу $2(n - 1) \times (n - 2)^{(n-2)/2}$ детерминантов матриц четных порядков, не кратных 4, достигают матрицы OD четных порядков, равных значениям B + 1 или 2B. Первый основной случай дает сечение $x^2 + y^2 = h$ с $h = 2n - 2 = 2B = 4L + 2 = 8T + 2$, который описывает кольца Барбы $x^2 + y^2 = 8T + 2$ — сечения параболоида на высотах на 1 больших квадратов нечетных целых чисел: 2, 10, 26, 50 и т. п. (см. рис. 2).

Любое такое кольцо содержит точку Гаусса с координатой $x = 1$, поскольку следующее из него уравнение $y^2 = 8T + 1$ всегда разрешимо квадратом целого числа. С подачи Гаусса точки (поверхности чаши гиперболоида, включая точки с целыми координатами) легко сводимы к точкам плоскости с решеткой треугольных чисел 0, 1, 3, 6, 10, … на ней. Так поступают с глобусом, изготавливая на его основе карту.



■ ***Рис. 3.*** Точки Гаусса на решетке треугольных чисел
■ ***Fig. 3.*** Gauss points on triangle numbers lattice

Рассмотрим преобразование квадратичного уравнения к линейной форме учетом $x^2 = 8x + 1$, $y^2 = 8y + 1$. Базовое уравнение упростится до вида $x + y = (h - 2)/8$. Создав координатную решетку на треугольных числах T = 0, 1, 3, 6, 10, ..., проведем параллельные линии через любые две симметрично расположенные на осях точки с координатами из ряда T, как показано на рис. 3. Это и есть множество решений уравнения $x + y = T$ для $h = 8T + 2$.

Кольцам отвечают уравнения прямых линий $y = T - x$, проходящих через одинаково расположенные точки на двух осях $x = 0$, $y = T$ или $y = 0$, $x = T$. Иногда на прямую линию попадает срединная точка, например, для T = 6 имеем $x = y = 3$. Это случай прохождения прямых параллельных линий через точки со значениями T на осях, но линии можно проводить и через любые две симметричные относительно диагонали точки с координатами из ряда 0, 1, 3, 6, 10, … .

Границы детерминантов адамаридов и мерсеннидов

Решение целочисленных уравнений дает инварианты k_1 и k_2, определяющие количество –1 в строках циклических блоков **A** и **B**. Заметим, что порядок бициклической матрицы $n = 2v$ может намного превышать значение влияющего на высоту $h = 2n - 2 = 2B$ порядка $n = B + 1$.

Для решения можно выбрать любое сечение параболоида, дающее точки Гаусса. Детерминант матрицы с каймой $\det([\mathbf{T}]) = (1 + S)\det(\mathbf{T})$ определяется через значение эксцесса S матрицы $\mathbf{T}^{-1}$ и детерминант матрицы без каймы **T**.

Перекос S в знаках при понижении n растет линейно, тогда как значение det(**T**) падает и оптимум произведения приходится на число B.

В целом порядки Барбы связаны с орнаментальными инвариантами бициклических матриц с каймой. Для порядков $4t$ справа стоит показатель $(h - 2)/8$, $h = n + \delta = 4t$. К решениям в треугольных числах задача не сводится, она решается в целых числах относительно исходных координат.

Кроме границ Элича и Барбы, которые описывают границы детерминантов не структурированных матриц, введем две новые, более реальные границы значений бициклических матриц с каймой. Матрицы порядков $4t + 1$ на основе бициклических матриц порядков $4t$ приводят к тому, что детерминант матрицы без каймы не превышает границы Адамара $n^{n/2}$, и так как эксцесс обратной матрицы растет не быстрее $\sqrt{n}$, то из $\det([\mathbf{T}]) = (1 + \mathrm{S})\det(\mathbf{T})$ следует, что $\det([\mathbf{T}]) \le (1 + n^{1/2})n^{n/2}$. Согласно оценке Брента [23] $\det([\mathbf{T}]) \ge 2n^{n/2}$.

Граница Ферма. Будем называть функцию $d(\mathrm{F}) = (1 + n^{1/2})n^{n/2}$ границей Ферма, поскольку матрицы на порядках чисел Ферма [19] касаются детерминантами ее, а не задранной вверх границы Барбы.

Граница Мерсенна. Будем называть функцию $d(\mathrm{M}) = n^{(n + 1)/2}$ границей Мерсенна, поскольку множество $4t - 1$ включает в себя числа Мерсенна. Впервые она выделена в работе [16], ее обозначают Nick's bound (граница Ника): матрицы порядков $4t - 1$, построенные на бициклических матрицах порядков $n = 4t - 2$, имеют детерминант $\det([\mathbf{T}]) \le n^{(n+1)/2}$. Эта граница нарушается на порядке $n = 18$: в выражении справа использованы оба порядка (бицикла n и бицикла с каймой $n + 1$).

Поделив формулы двух границ на $n^{n/2}$ и сопоставляя оставшиеся $n^{1/2} + 1$ и $n^{1/2}$, приходим к заключению, что $d(\mathrm{F}) = d(\mathrm{M}) + d(\mathrm{M})/\mathrm{S}$, где величина $\mathrm{S} = n^{1/2}$ у матриц Адамара описывает максимально возможный эксцесс. Эксцессы прямой и обратных матриц различаются множителем, равным порядку, в научной литературе [21, 23] чаще упоминается первый, но для формул границ удобен второй. Эксцесс возрастает с ростом порядка, так что границы Ферма и Мерсенна сближаются сколь угодно близко и различие между бициклическими матрицами исчезает. Кроме того, граница Барбы $d(\mathrm{B}) = d(\mathrm{M}) \times (2 + 1/\mathrm{S}^2)^{1/2}$, т. е. обе они стремятся стать меньше границы Барбы на $\sqrt{2}$ (на значение гипотенузы квадрата со стороной 1).

Граница Ника $d(\mathrm{M})$ — довольно важная универсальная формула: с ее помощью, а также с помощью эксцесса S выражаются остальные границы в виде $\mathrm{A}(d(\mathrm{M}) \pm \mathrm{K}d(\mathrm{M})/\mathrm{S})$, где A — амплитуда, K — крутизна склона границы, обычно 1 или 1/4.

Следуя Бренту [23], амплитуду A имеет смысл выражать в долях $\sqrt{2}$, основания натурального логарифма e, π или их инверсий и комбинаций. Например, граница Барбы $d(\mathrm{B})$ с ростом порядка стремится к $\sqrt{2}/\mathrm{e}$, т. е. $\sim 0{,}85$ от границы Адамара $n^{n/2}$.

Крайнее значение $d(\mathrm{M})$ меньше этого трансцендентного числа на $\sqrt{2}$, плата за отказ от произвольного вида экстремальной матрицы конкретна и выражается красивой формулой в виде $\sqrt{(1/\mathrm{e})}$ ($\sim 0{,}6$ от границы Адамара).

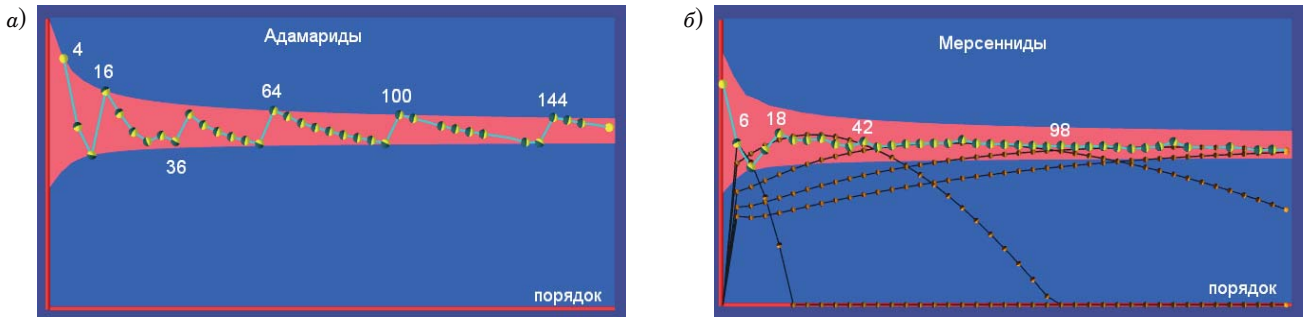
Детерминанты экстремальных матриц слегка осциллируют между опорными порядками в виде порядков рациональных матриц (у адамаридов) или порядков, наведенных кольцами Барбы (у мерсеннидов). Амплитуда A принимает значение 1, когда $d(\mathrm{F}) = d(\mathrm{M}) + d(\mathrm{M})/\mathrm{S}$, или меньше 1, настраиваемое так, чтобы осцилляции не выходили за указанные константами пределы *в начале*, где размах колебаний максимален. Например, для границы адамаридов снизу $\mathrm{A} = 3/\pi$. У мерсеннидов детерминанты осциллируют несколько ниже, выбирая A для границ сверху и снизу как $(\mathrm{e} + 1)/4$ и $(\mathrm{e} - 1)/2$, причем низ легко согласуется с более осторожной оценкой Брента [23] выбором $(2/\pi)^{1/2}$.

Политика выбора крутизны склонов K одинакова для обоих типов матриц, для верхнего склона 1, для нижнего $\mathrm{K} = 1/4$ (вычитаем $\mathrm{K}d(\mathrm{M})/\mathrm{S}$).

Осцилляции детерминантов описываются квадратичной функцией порядка [16]: для адамаридов $\det([\mathbf{T}]) \approx 0{,}8 + 0{,}2(n/n^* - 2)^2 d(\mathrm{F})$, где $n^* = 4q^2$ — наводящий порядок (ближний снизу порядок регулярных матриц Адамара). Придерживаясь той же практики, что и при описании границ, заменим первые два настраиваемых коэффициента на близкие к ним $(2/\pi)^{1/2}$ и $1 - (2/\pi)^{1/2}$ (рис. 4, *а*).

Как видно, точки найденных относительных детерминантов (по отношению к границе Адамара) оседают на вычисленной для них кусочно-квадратичной функции с нарушениями, вносимыми порядками с множителями в виде квадрата числа Мерсенна (или, в общем, $4t - 1$). Таков, например, порядок 36 регулярной матрицы Адамара, которая не может быть бициклической, на соседнем порядке 32 экстремальный бицикл тоже не ортогонален и за счет этой свободы образует, напротив, локальный уход графика вверх.

Хотя оценка максимума детерминанта удовлетворительна, желательно получить ее с учетом эксцесса S. Допустим, порядок n расположен между порядками двух регулярных матриц $h = 4q^2 \le n \le h_2 = 4(q + 1)^2$, $q = 1, 2, 3, \ldots$ или между двумя соседними учетверенными квадратами $\mathrm{Q} = 8\mathrm{T} + 1$ либо $\mathrm{Q} = \mathrm{F} - 1$. Тогда $(n - h)/(h_2 - h)$ меняется от 0 до 1 линейно. Если эксцесс $\mathrm{S} \le n^{1/2}$, то определитель $\det(\mathbf{T}) \le n^{n/2}$. Мы касаемся обеих зависимостей на старте $n = h$, ес-



■ ***Рис. 4.*** Границы детерминантов адамаридов (*а*) и мерсеннидов (*б*)
■ ***Fig. 4.*** Adamaride (*а*) and Mersennide (*б*) determinant bounds

ли $S = (1 + M)n^{1/2}$, $M = K(n - h)/(h_2 - h)$ и $K = 1/q$, а определитель отличается знаком, амплитудой и показателем степени корректирующего члена $\det(\mathbf{T}) = (1 - (M/e)^{1/2})n^{n/2}$, обеспечивающим выпуклость зубцов «пилы» $\det([\mathbf{T}]) = (1 + S)\det(\mathbf{T})$.

На порядках мерсеннидов [16] линейная оценка $\det(\mathbf{T}) = (1 - (v/(L + 1) - 1)/15)$ в долях границы Элича $d(E) = 2(n - 1)(n - 2)^{(n-2)/2}$ зависит от размера плеча бициклической матрицы $v = n/2$. Здесь $L = 2T$ — число полосок (альтернансов знака в $\mathbf{N} = \mathbf{A}^T\mathbf{A} + \mathbf{B}^T\mathbf{B}$), равное у экстремальных матриц удвоенному треугольному числу. Линейная аппроксимация берется для актуального диапазона порядков, величина наклона мало влияет на итог расчета. На старте при $T = 0$ крутизна склона 15 линейной характеристики снижается до 12.

Значение детерминанта матрицы с каймой $[\mathbf{T}]$ отличается от детерминанта матрицы без каймы эксцессом обратной матрицы, увеличенным на 1: $\det([\mathbf{T}]) = (1 + S)\det(\mathbf{T})$. S описывается аналитически точно линейной зависимостью $S = n/((8T + 1)^{1/2} + (8T + 1)^{-1/2})$. Формулу для эксцесса обратной матрицы можно писать, опираясь на другие инварианты орнамента, используя соотношения $h = 2n - 2 = 2B = 4L + 2 = 8T + 2$, в частности через число полосок $S = n/((4L + 1)^{1/2} + (4L + 1)^{-1/2})$.

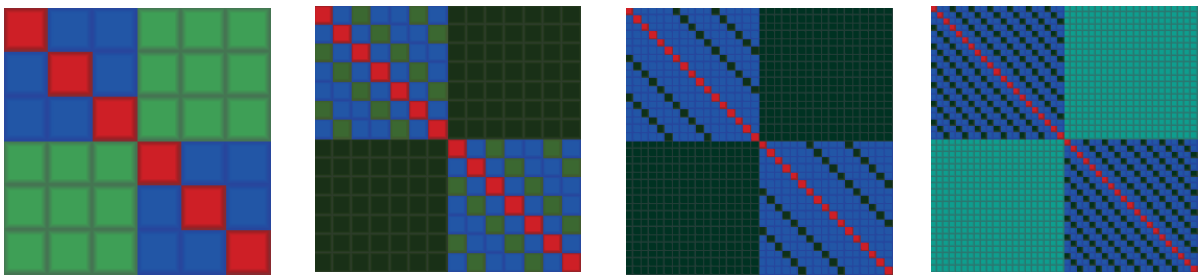
Произведение $\det(\mathbf{T})$ на $1 + S = 1 + 2v/((4L + 1)^{1/2} + (4L + 1)^{-1/2})$ дает цепочку смещающихся с увеличением L парабол детерминантов (поделенных на адамарову оценку $n^{n/2}$) (рис. 4, *б*).

Как видно, точки найденных относительных детерминантов (по отношению к границе Адамара) оседают на вычисленной для них кусочно-квадратичной функции с небольшим возрастанием на порядках, на которых число полосок L кратно размеру плеча бициклической матрицы. Матрица $\mathbf{T}^T\mathbf{T}$ содержит на диагонали (остальное нули) пару блоков симметричной циклической орнаментальной матрицы, содержащей на месте более редких элементов две полоски, остальные –2. В отличие от орнаментальных матриц ортогональных структур, эта матрица «полосатая» (рис. 5).

Кроме выделенных цветом узких границ для колебаний, мы можем выделить более широкие полосы захвата вниз для матриц тех порядков, для которых найти бициклическую матрицу затруднительно. Бициклическая структура значительно ограничивает хаос изменения знаков элементов регулярной разверткой их по высоте сдвигами, так что серьезной перспективы сдвинуть границу вверх, к границе Барбы, поиск матриц максимума детерминанта не имеет.

Для матриц Адамара высота h сечения параболоида равна порядку.

В 80-х годах прошлого века максимальные значения эксцессов матриц Адамара выяснили Фармакис и Коунис [21]. Если взглянуть на точки (поделенные на $n^{1/2}$), куда легли полу-



■ ***Рис. 5.*** Орнаментальные блоки матрицы $\mathbf{T}^T\mathbf{T}$ порядков 6, 14, 34, 42
■ ***Fig. 5.*** Ornamental blocks of matrix $\mathbf{T}^T\mathbf{T}$ for order 6, 14, 34, 42

ченные ими значения, объединенные синими связками, то окажется, что они сидят на семействе кривых, вычисленных по формуле $S \approx n/((h+1)^{1/2} + (h+1)^{-1/2})$ посредством небольшой вариации $h = n + \delta$ (δ *может принимать дробные значения*) в окрестности порядка.

Центральная оранжевая кривая, приведенная для $\delta = 0$ на рис. 6, служит осевой для колебаний максимально возможного эксцесса в зависимости от порядка.

Отсюда легко получаем часто используемое в теории эксцессов этих матриц неравенство $S \le n^{1/2}$, считая делитель стремящимся к $h^{1/2}$.

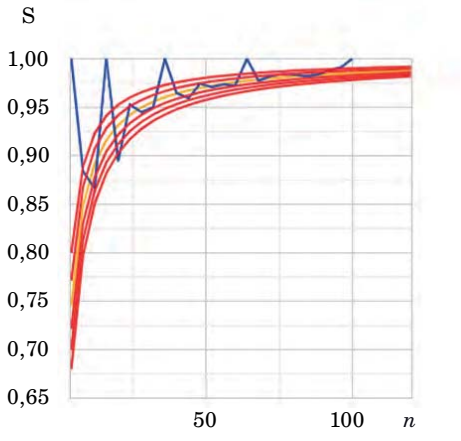
Ранее отмечалось, что некоторые свойства экстремальных по детерминанту матриц связаны не столько с простотой множителей порядков, сколько с началом числовой системы. Например, есть всего три типа последовательностей Голея, основанных на простоте чисел 2, 5, 13, но не захватывающих число 17 и большие.

В особенности это правило касается симметрий, порядок симметричных циклических матриц Адамара не превышает 4, а у бициклических — 32 [13, 15].

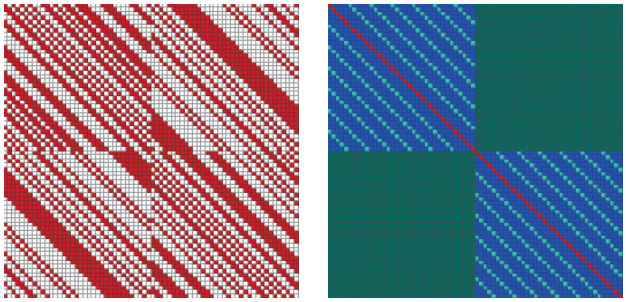
Обобщение. Для мерсеннидов максимальный порядок симметричных экстремальных бициклических матриц равен 66 (размер плеча 33).

Начиная с порядка 26, симметричные бициклические матрицы идут с шагом 8: 26, 34 (нет 42), 50, 58, 66. Для матриц порядка 42 размер плеча 42/2 = 21 — не простое число и не степень простого числа, что исключает симметрию матрицы. На порядках 74 и 82 симметричных матриц не найдено. Найденная матрица ближайшего порядка 70 с максимальным среди бициклических структур детерминантом и орнаментальная для нее матрица представлены на рис. 7.

Они иллюстрируют еще один тип симметрии, орнаментальный. Матрицы с этим типом симметрии встречаются, очевидно, когда число полосок (плюс 1) кратно размеру последовательности: здесь 35/7 = 5 описывает число промежуточных элементов между полосками. Различные типы симметрий, описанные компрессиями, орбитами, орнаментальными расстояниями между полосками и т. п., помогают искать довольно сложные в поиске экстремальные матрицы.



■ ***Рис. 6.*** Колебания максимального эксцесса
■ ***Fig. 6.*** Maximal excess oscillations



■ ***Рис. 7.*** Портреты бициклической матрицы порядка 70 и орнамент $\mathbf{T}^{\mathrm{T}}\mathbf{T}$
■ ***Fig. 7.*** Two circulant matrix of order 70 and ornament $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ portraits

Майнинг экстремальных бициклов

Существует несколько способов повышения эффективности поиска матриц — майнинга [24]. Матрица **T** без каймы порядка n с элементами 0 и 1 (вместо 1 и –1) отличается детерминантом от связанного с ней бицикла с каймой нечетного порядка $n + 1$ ровно в 2^n раза. Это обстоятельство нередко используется поисковиками — ищут именно эти матрицы.

Бициклическая матрица **N** (вдвое меньшего порядка, чем **T**) с элементами 2 и –2 и значением n на диагонали можно искать независимо, поэтому задача определения орнаментальной матрицы и значения оптимального детерминанта матрицы заметно проще нахождения ее блоков **A** и **B**. Перебор элементов не двух блоков **A** и **B**, а только субблока **N**, дает заметное ускорение и надежность вычислений, позволившие установить зависимость количества альтернансов 2 от порядка матрицы.

Метод компрессии. Самый простой и действенный состоит в замене пар элементов 1, 1 на 2; –1, –1 на –2; 1, –1 и –1, 1 на 0. Это называется компрессией последовательностей a, b в последовательности c, d вдвое. Для матриц невысоких порядков подходящую компрессию c, d можно угадать, ведь это недлинные последовательности. Декомпрессия состоит в замене элементов в обратной последовательности, причем гадательным остается лишь третий вариант, когда расширяется 0.

Почему компрессия эффективна, понять несложно, ведь фиксация c, d заметно уменьшает разброс возможных решений. Метод поиска эффективнее цепляет цель, как и при стрельбе из ружья дробью. Там, где пуля промахивается, из ствола выпускается облако дроби, которая сшибает утку. Поэтому этот метод называется еще *утиной охотой*. Коэффициент компрессии можно назначать любой, метод легко обобщается на любые коэффициенты сжатия, в особенности на случай составных размеров длин последовательностей. Если длина последовательности — простое число, образуется остаток, как в алгоритме Евклида измерения длин эталонным метром. Остаток можно сжимать иначе. Это создает широкое поле для вариантов реализации метода компрессии.

Вариацией утиной охоты является очень эффективный метод орбит.

Метод орбит. В методе орбит фиксируется орбита — сет адресов –1 (прямых или косвенных, кратных реальным адресам). Так как все адреса являются умножениями элементов орбиты на подбираемые при поиске множители, длинная орбита, как и длинная угаданная матрица компрессии, очень быстро приводит поиск к результату. Метод орбит опирается на разложимость группы, связанной с размером последовательности, на циклические подгруппы, что позволяет автоматизировать выбор орбит.

Метод фильтрации. К алгоритмам, ускоряющим поиск, относится контроль амплитуд спектров фурье-последовательностей. У ортогональных последовательностей амплитуда спектра ограничена порядком матрицы или близким к нему значением. Так как матрицы большого детерминанта не ортогональны, то границу придется поднимать, но сама идея уменьшить число последовательностей отбраковкой плохих по спектру остается привлекательной.

Все вместе это позволяет искать решения порядков, помещенные в таблицу Томаса [24], заметно быстрее перебора. Кроме того, верхний предел таблицы отодвигается выше и превышает изученный вдвое, за исключением не удобных для применения акселераторов порядков. Благодаря описанным методам найдено множество матриц бициклической структуры, не указанных в результатах ранее, включая симметричные матрицы порядков 6, 10, 14, 18, 22, 26, идущих через 4 [25].

Заключение

Представленные в статье формулы для детерминантов и эксцессов экстремальных бициклических матриц с каймой являются новыми и публикуются впервые. В силу неограниченного роста сложности структуры матриц нечетных порядков детерминанты ранее часто оценивались всего лишь границами сверху и снизу.

При ограничении структуры можно получать более информативные сведения, подтвержденные найденными конкретными матрицами.

К сходным приведенным в статье уравнениям параллельных линий на решетке четных целых чисел $2t$ приводит проблема Гольдбаха. Гольдбах заметил, что нет такого четного числа, которое не разлагалось бы в сумму двух простых $6 = 3 + 3$, $8 = 3 + 5$, $10 = 3 + 7$, $12 = 5 + 7$, ..., а это значит, что на решетке всегда найдется хотя бы одна точка, координатами которой будут простые числа.

Задача оптимизации детерминанта бициклической матрицы с каймой меняет решетку четных чисел на решетку треугольных чисел. Кусочно-квадратичные аппроксимации детерминантов адамаридов и мерсеннидов отличаются не только видом, но и тем, что вторая не подвержена «сбоям», которые вносят квадраты чисел Мерсенна и их обобщения.

В работе получена точная картина поведения детерминантов на неограниченном диапазоне порядков, при этом сами матрицы не ищутся, они даются выборочно лишь для сравнения и подтверждения расчетов. Впрочем, для всех матриц Адамара детерминанты известны, и мы можем лишь сомневаться, является ли конкретная матрица бициклической.

Библиотека бициклических матриц и бициклических матриц с каймой сайта mathscinet.ru содержит исследованные структуры, параметры и виды экстремальных матриц.

Финансовая поддержка

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации, соглашение № FSRF-2023-0003.

ПРИЛОЖЕНИЯ

Детерминант det([**T**]) приведен к границе Адамара делением на нее, оптимальные на структуре бициклические матрицы с каймой достигают с ростом порядка примерно половину ее значения, тогда как для МБД допустима и четверть. В отчетах по бинарным матрицам указываются две строки, соответствующие последовательностям a и b. Бинарные элементы последовательностей обозначаются символами « + » и «–».

■ **Таблица П1.** Экстремальные матрицы семейства адамаридов
■ **Table П1.** Extremal matrices of the hadamarid family

$n = 4$, det([**T**] = 0,8587, $k_1 = 1, k_2 = 0, x = 0, y = 2, x^2 + y^2 = 4$ −+ ++
$n = 8$, det([**T**] = 0,6243, $k_1 = 1, k_2 = 1, x = 2, y = 2, x^2 + y^2 = 8$ −+++ −+++
$n = 12$, det([**T**] = 0,5272, $k_1 = 3, k_2 = 2, x = 0, y = 2, x^2 + y^2 = 4$ −++−−+ −+−+++
$n = 16$, det([**T**] = 0,7466, $k_1 = 4, k_2 = 2, x = 0, y = 4, x^2 + y^2 = 16$ +−−+++−− +−+++++−
$n = 20$, det([**T**] = 0,6698, $k_1 = 4, k_2 = 3, x = 2, y = 4, x^2 + y^2 = 20$ +++−+−+−−+ −+++++−++−
$n = 24$, det([**T**] = 0,6056, $k_1 = 6, k_2 = 4, x = 0, y = 4, x^2 + y^2 = 16$ +−+−−+++−−+− −−+−+++++−++
$n = 28$, det([**T**] = 0,5741, $k_1 = 7, k_2 = 5, x = 0, y = 4, x^2 + y^2 = 16$ +−+++−+−+−−−+ −++−++−+−++++−
$n = 32$, det([**T**] = 0,5941, $k_1 = 8, k_2 = 5, x = 0, y = 6, x^2 + y^2 = 36$ ++−−−−−++−+−++−+ −+++−+++−+−−++++
$n = 36$, det([**T**] = 0,5746, $k_1 = 8, k_2 = 6, x = 2, y = 6, x^2 + y^2 = 40$ +++−+−+−−++++−−+−− −++−+−−−+++−++++++
$n = 40$, det([**T**] = 0,6672, $k_1 = 9, k_2 = 7, x = 2, y = 6, x^2 + y^2 = 40$ +++−+−++−++−+++−−−−− +++−++−+−−−+++−++++−
$n = 44$, det([**T**] = 0,6303, $k_1 = 11, k_2 = 8, x = 0, y = 6, x^2 + y^2 = 36$ +−−++−+−−+−+++−−−+−+−+ ++−+++−++++++−−+−−−−++
$n = 48$, det([**T**] = 0,607, $k_1 = 12, k_2 = 9, x = 0, y = 6, x^2 + y^2 = 36$ −−++++−++−−+−−−−+−−+−+++ +−++++++−−−++++−++−−+−+−
$n = 52$, det([**T**] = 0,5898, $k_1 = 13, k_2 = 10, x = 0, y = 6, x^2 + y^2 = 36$ −++−−+−+−−++−−++++++−+−−−− ++++−+−−+++++−−++−+−+−++−−
$n = 56$, det([**T**] = 0,5764, $k_1 = 14, k_2 = 11, x = 0, y = 6, x^2 + y^2 = 36$ +++−−+++−+−+−−−++−++−++−−−−− ++−++++++−++++−+−+−−++−+−−−−
$n = 60$, det([**T**] = 0,566, $k_1 = 15, k_2 = 12, x = 0, y = 6, x^2 + y^2 = 36$ ++−−−++−+−−++−+−−−+−++++−++−−− −++++−−+−−++++++−++++−+−+−+−−−
$n = 64$, det([**T**] = 0,6797, $k_1 = 16, k_2 = 12, x = 0, y = 8, x^2 + y^2 = 64$ +−+−−−+−+−++++−+−−+++−++−++−−−−− +−−+++++−+−−+−+++++−++++−−−++−−+
$n = 68$, det([**T**] = 0,6596, $k_1 = 16, k_2 = 13, x = 2, y = 8, x^2 + y^2 = 68$ −−−+−−+−−+−+−+++−−−−++++++−++++−−+ −+−−++−++++−−−++−−+++++−+−+++−++−+

■ *Окончание табл. П1*
■ *End of Table П1*

$n = 72$, det([**T**] = 0,6371, $k_1 = 18$, $k_2 = 14$, $x = 0$, $y = 8$, $x^2 + y^2 = 64$ +–+++–++–++–––++–++–+–++–––+++–––––– ++–+–+–+–++++––+–+++++++++–++––++–––
$n = 76$, det([**T**] = 0,6209, $k_1 = 19$, $k_2 = 15$, $x = 0$, $y = 8$, $x^2 + y^2 = 64$ ++++–++++–+–+–+–+++––+–––++–+––+–––––– ++––+–++–+++++–+–++––+++–++––+++++–––
$n = 80$, det([**T**] = 0,6084, $k_1 = 18$, $k_2 = 16$, $x = 4$, $y = 8$, $x^2 + y^2 = 80$ –++–+++––++–+––++++++–––+––++––+–+–+ –++++++–++++––+––+++++–+–+–––+–++–––++–+
$n = 84$, det([**T**] = 0,596, $k_1 = 21$, $k_2 = 17$, $x = 0$, $y = 8$, $x^2 + y^2 = 64$ ––+–––+–++––+–+–––––+++–+–++++–+–+–++++–+– +––+–++–––++++++––++–+++––+––+–++–++++––+
$n = 88$, det([**T**] = 0,5862, $k_1 = 22$, $k_2 = 18$, $x = 0$, $y = 8$, $x^2 + y^2 = 64$ –+++–+–++––––++++–+––+–+––+––+++++––++–––+ –+–––++–+–+++++–++–+++++–––+++–+++–++–+––+––+––
$n = 92$, det([**T**] = 0,5777, $k_1 = 23$, $k_2 = 19$, $x = 0$, $y = 8$, $x^2 + y^2 = 64$ –+++–––+++–+––+–+–+++––+–+++–++–––+++–++–––+––– ––++++++–+++–++++++–+–––––+–+–++–++––+––+–++–
$n = 96$, det([**T**] = 0,5674, $k_1 = 24$, $k_2 = 20$, $x = 0$, $y = 8$, $x^2 + y^2 = 64$ +–+++–++++––+–+––++–+–––+++–+––––––+–+–+–++–+––+ ++––+–––+–+++––+–––++++–+–++–++–––+++––++++++–++
$n = 100$, det([**T**] = 0,6655, $k_1 = 25$, $k_2 = 20$, $x = 0$, $y = 10$, $x^2 + y^2 = 100$ ++–+++–+–+–+–++––––+–+–++–+––––––+––+––++++–++++––– –+–––++++++–++++–++–++++––––––+++––+++–+–++––++–++–
$n = 104$, det([**T**] = 0,6527, $k_1 = 25$, $k_2 = 21$, $x = 2$, $y = 10$, $x^2 + y^2 = 104$ –+––++–++++–––––––++–––+–+–+––++–++++–+–++++––+++–+– +–++––+––––+++++++––+++–+––+––++–++++–+–++++––+++–+–

■ **Таблица П2.** Экстремальные матрицы семейства мерсеннидов
■ **Table П2.** Extremal matrices of the mersennid family

$n = 2$, det([**T**] = 0,7698, L = 0, $k_1 = 0$, $k_2 = 0$, $x = 1$, $y = 1$, $x^2 + y^2 = 2$ + +
$n = 6$, det([**T**] = 0,5642, L = 0, $k_1 = 1$, $k_2 = 1$, $x = 1$, $y = 1$, $x^2 + y^2 = 2$ –++ –++
$n = 10$, det([**T**] = 0,4793, L = 2, $k_1 = 2$, $k_2 = 1$, $x = 1$, $y = 3$, $x^2 + y^2 = 10$ ++––+ ++++–
$n = 14$, det([**T**] = 0,5411, L = 2, $k_1 = 3$, $k_2 = 2$, $x = 1$, $y = 3$, $x^2 + y^2 = 10$ ––++++– ++–++–+
$n = 18$, det([**T**] = 0,5985, L = 2, $k_1 = 4$, $k_2 = 3$, $x = 1$, $y = 3$, $x^2 + y^2 = 10$ +––++++–– ++–++–+–+
$n = 22$, det([**T**] = 0,5748, L = 2, $k_1 = 6$, $k_2 = 4$, $x = –1$, $y = 3$, $x^2 + y^2 = 10$ +–+––++––+– ++–+++++–––
$n = 26$, det([**T**] = 0,574, L = 2, $k_1 = 6$, $k_2 = 5$, $x = 1$, $y = 3$, $x^2 + y^2 = 10$ +–+––++++––+– +++++–+–––++–

■ *Продолжение табл. П2*
■ *Continuation of Table П2*

$n = 30$, det([**T**] = 0,5804, L = 2, $k_1 = 7$, $k_2 = 6$, $x = 1$, $y = 3$, $x^2 + y^2 = 10$ +−−−++−−+−++−++ +−+−−++++++−−+−
$n = 34$, det([**T**] = 0,5608, L = 2, $k_1 = 8$, $k_2 = 7$, $x = 1$, $y = 3$, $x^2 + y^2 = 10$ +++−+−−−++−−−+−++ +−++++−++−+−−++−−
$n = 38$, det([**T**] = 0,5518, L = 2, $k_1 = 10$, $k_2 = 8$, $x = –1$, $y = 3$, $x^2 + y^2 = 10$ −−−−++−+++++−−−−+−+ −+−+−++−−++−++++−−+
$n = 42$, det([**T**] = 0,5694, L = 6, $k_1 = 11$, $k_2 = 8$, $x = –1$, $y = 5$, $x^2 + y^2 = 26$ +−−++−+−++−−−+−+−+−++−− −+−−+++−−−−++−+++++++
$n = 46$, det([**T**] = 0,5501, L = 6, $k_1 = 12$, $k_2 = 9$, $x = –1$, $y = 5$, $x^2 + y^2 = 26$ −++−+−+−−−++−+−−−+−+−++ −++++−−++−−−−−+++++−+++
$n = 50$, det([**T**] = 0,5559, L = 6, $k_1 = 12$, $k_2 = 10$, $x = 1$, $y = 5$, $x^2 + y^2 = 26$ ++−+++−−−−+−++−+−−−−+++−+ +++−−−+−++−−+−++−−+++−+++
$n = 54$, det([**T**] = 0,5613, L = 6, $k_1 = 13$, $k_2 = 11$, $x = 1$, $y = 5$, $x^2 + y^2 = 26$ +++−−+−+−−++−+−++−−−+−+−−++ −+++−+++−++−−−−−+++−+−−++++
$n = 58$, det([**T**] = 0,5624, L = 6, $k_1 = 14$, $k_2 = 12$, $x = 1$, $y = 5$, $x^2 + y^2 = 26$ +−++−−−−+−+++−++−+++−+−−−−++− −−++++−+++++−−−+−−++−−+−+−+++
$n = 62$, det([**T**] = 0,5637, L = 6, $k_1 = 15$, $k_2 = 13$, $x = 1$, $y = 5$, $x^2 + y^2 = 26$ +++−+−−++−−−+−+−++−+−−−++−−+−++ −++++++−−+−+−−−++−−−+−+++++−++−
$n = 66$, det([**T**] = 0,5641, L = 6, $k_1 = 16$, $k_2 = 14$, $x = 1$, $y = 5$, $x^2 + y^2 = 26$ +−+−−+−−+++++−−−++−−−+++++−−+−−+− −−+−+−+++−+−++−+++−++++−++−−−−−++
$n = 70$, det([**T**] = 0,5753, L = 6, $k_1 = 17$, $k_2 = 15$, $x = 1$, $y = 5$, $x^2 + y^2 = 26$ +−−++−−+−−−−+−+−+−+++++−−+++−++−+−− +−−−−−++++−++−+−++−−−++++−−+++−+++−
$n = 74$, det([**T**] = 0,5628, L = 6, $k_1 = 19$, $k_2 = 16$, $x = –1$, $y = 5$, $x^2 + y^2 = 26$ −+−++−+++−−−++−++−−+−−−−−+−−−+++−++−+ ++−−−−−+++−−−+++−+++−++−++−+−−+−+−+++
$n = 78$, det([**T**] = 0,5614, L = 6, $k_1 = 20$, $k_2 = 17$, $x = –1$, $y = 5$, $x^2 + y^2 = 26$ −++−−−+−−−−−++++−+−−−−+−+−−++++−++−++− ++−−+−+−++−−+−+−+−−+++−−−+−++++−++−−++
$n = 82$, det([**T**] = 0,5595, L = 6, $k_1 = 20$, $k_2 = 18$, $x = 1$, $y = 5$, $x^2 + y^2 = 26$ +−+−++−+−−+−−++++−++−−+−+++−−−−+−−−+−−+++ ++−−−++−+++−++−++++++−−+−−−+++−+−+−+−−−−+
$n = 86$, det([**T**] = 0,5573, L = 6, $k_1 = 21$, $k_2 = 19$, $x = 1$, $y = 5$, $x^2 + y^2 = 26$ −−−−−++−−+++−+−+++−+−+−−++−+−−++−−−+−+++++ −+−++−−++++−++++−+++−+−−+−++−−−++−+++−−−−−
$n = 90$, det([**T**] = 0,5557, L = 6, $k_1 = 22$, $k_2 = 20$, $x = 1$, $y = 5$, $x^2 + y^2 = 26$ +−−−+−−−+++−++−−+−−+−−−−++−+−−−−+−+++++++−+++ ++−+−+−+−−++++−−−−+++−−+++−+−+−++−++−+−−++−−
$n = 94$, det([**T**] = 0,5522, L = 6, $k_1 = 23$, $k_2 = 21$, $x = 1$, $y = 5$, $x^2 + y^2 = 26$ −−−−−++−++−++++++++−+−+−+−−+−−−++−+−−−−++−−++++− ++−++−−+−++++−+++−−+−+−+++−+−++−−−++−−+−+++−−−−

- *Окончание табл. П2*
- *End of Table П2*

$n = 98$, det([**T**] = 0,5554, L = 6, $k_1 = 24$, $k_2 = 22$, $x = 1, y = 5, x^2 + y^2 = 26$ ++–+++–+–+––++++–++–+–++++––+–+–––+–––+–++––––+– ––+–+–++++++––+++–––––++++––++––+++–++–++––++–+––
$n = 102$, det([**T**] = 0,5513, L = 12, $k_1 = 25$, $k_2 = 22$, $x = 1, y = 7, x^2 + y^2 = 50$ +++––+–––+–+––++–+++–+–++––––+––––+++–+–++++–+–––++ –+––+++––++–––+–+++––––+––+–++++–++–+++–++–++++–+–
$n = 106$, det([**T**] = 0,5522, L = 12, $k_1 = 26$, $k_2 = 23$, $x = 1, y = 7, x^2 + y^2 = 50$ –++–++––++++–++––+–+++–+–+–––––+––+–++++–––++––+–+–+ +–+–––+++++––+++–++++++++––+–––++–+–++–+–+++––++––––+

Литература

1. **Orrick W. P.** The maximal {–1, 1}-determinant of order 15. *Metrika*, 2005, no. 62, pp. 195–219. doi.org/10.1007/s00184-005-0410-3
2. **Orrick W. P., Solomon B.** Large determinant sign matrices of order 4k + 1. *Discrete Mathematics*, 2007, no. 307, pp. 226–236.
3. **Seberry J., Xia T., Koukouvinos C., Mitrouli M.** The maximal determinant and subdeterminants of ±1 matrices. *Linear Algebra and its Applications*, 2003, vol. 373, pp. 297–310. doi:10.1016/S0024-3795(03)00584-6
4. **Ehlich H.** Determinantenabschätzungen für binäre Matrizen. *Mathematische Zeitschrift*, 1964, no. 83, pp. 123–132. doi:10.1007/BF01111249
5. **Wojtas M.** On Hadamard's inequality for the determinants of order non-divisible by 4. *Colloquium Mathematicum*, 1964, vol. 12, pp. 73–83. doi:10.4064/cm-12-1-73-83
6. **Osborn J. H.** *The Hadamard maximal determinant problem: honours thesis.* University of Melbourne, 2002. 144 p. http://maths-people.anu.edu.au/~osborn/publications/pubsall.html (дата обращения: 18.01.2023).
7. **Cohn J. H. E.** On determinants with elements ±1. II. *Bulletin of the London Mathematical Societ*, 1989, vol. 21, iss. 1, pp. 36–42. doi.org/10.1112/blms/21.1.36
8. **Neubauer M. G., Radcliffe A. J.** The maximum determinant of ±1 matrices. *Linear Algebra and its Applications*, 1997, vol. 257, pp. 289–306.
9. **Hadamard J.** Resolution d'une question relative aux determinants. *Bulletin des Sciences Mathematiques*, 1893, vol. 17, pp. 240–246.
10. **Wang R.** *Introduction to Orthogonal Transforms with Applications in Data Processing and Analysis*. Cambridge University Press, 2010. 504 p.
11. **Ahmed N., Rao K. R.** *Orthogonal Transforms for Digital Signal Processing*. Springer-Verlag Berlin Heidelberg, 2012. 264 p.
12. **Jennifer S., Yamada M.** *Hadamard Matrices*: *Constructions using Number Theory and Linear Algebra*. Wiley, 2020. 352 p.
13. **Балонин Н. А., Джокович Д. Ж.** Симметрия двуциклических матриц Адамара и периодические пары Голея. *Информационно-управляющие системы*, 2015, № 3, с. 2–16. doi.org/10.15217/issn1684-8853.2015.3.2
14. **Colbourn C. J., Dinitz J. H.** *Handbook of Combinatorial Designs*. Second Edition. Chapman and Hall, CRC, 2007. 967 p.
15. **Балонин Н. А., Сергеев М. Б., Себерри Дж., Синицына О. И.** Окружности на решетках и матрицы Адамара. *Информационно-управляющие системы*, 2019, № 3, с. 2–9. doi.org/10.31799/1684-8853-2019-3-2-9
16. **Балонин Н. А., Сергеев М. Б., Себерри Дж., Синицына О. И.** Окружности на решетках и матрицы максимального детерминанта. *Информационно-управляющие системы*, 2020, № 6, с. 2–11. doi.org/10.31799/1684-8853-2020-6-2-11
17. **Turyn R. J.** Hadamard matrices, Baumert – Hall units, four symbol sequences, pulse compression and surface wave encodings. *J. Combin. Theory*, Ser. A, 1974, vol. 16, iss. 3, pp. 313–333. doi.org/10.1016/0097-3165(74)90056-9
18. **Balonin N. A., Sergeev M. B., Vostrikov A. A.** Prime Fermat numbers and maximum determinant matrix conjecture. *Информационно-управляющие системы*, 2020, № 2, с. 2–9. doi.org/10.31799/1684-8853-2020-2-2-9
19. **Matiyasevich Y.** On recursive unsolvabilitv of Hilbert's tenth problem. *Proc. of Fourth Intern. Congress on Logic, Methodology and Philosophy of Science*, Bucharest, 1971, Amsterdam, North-Holland, 1973, pp. 89–110.
20. **Farmakis N., Kounias S.** The excess of Hadamard matrices and optmimal designs. *Discrete Mathematics*, North-Holland, 1987, vol. 67, iss. 2, pp. 165–176. doi.org/10.1016/0012-365X(87)90025-2
21. **Barba G.** Intorno al teorema di Hadamard sui determinanti a valore Massimo. *Giornale di Matematiche di Battaglini*, 1933, vol. 71, pp. 70–86.
22. **Brent R. R., Osborn J. H.** General lower bounds of maximal determinants of binary matrices. *Preprint*, 2012.
23. **Сергеев А. М., Балонин Ю. Н.** Майнинг матриц. *Обработка, передача и защита информации в компьютерных системах 22: сб. докл. Второй*

Междунар. науч. конф., Санкт-Петербург, 2022, с. 169–173.

24. **Rokicki T.** *New records for maximal determinants, based on pairs of circulant matrices*. http://tomas.rokicki.com/newrec.html (дата обращения: 18.01.2023).

25. *Algorithms on line. Hadamard matrices*. http://mathscinet.ru/catalogue/index.php (дата обращения: 18.01.2023).

UDC 519.614
doi:10.31799/1684-8853-2023-3-2-15
EDN: JQPBFE

Maximum determinant two circulant matrices with border

N. A. Balonin[a], Dr. Sc., Tech., Professor, orcid.org/0000-0001-7338-4920, korbendfs@mail.ru
M. B. Sergeev[a], Dr. Sc., Tech., Professor, orcid.org/0000-0002-3845-9277
[a]Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation

Introduction: The maximum determinant matrices are well known and studied for even orders 2t and 4t, where their structure is most often a bicycle, which is called Hadamard if the order is divided by 4. Odd orders have been studied much less due to the fact that the complexity of the structure of optimal matrices increases indefinitely. **Purpose:** To replace a deliberately complex object with a hyperobject consisting of a bicycle with a border, and being optimal on the set of matrices of fixed structure. To reveal the relationship of Gauss points (on the sections of bodies of revolution) with the number and types of maximum determinant matrices with a fixed structure for odd orders. To determine the upper and lower bounds for the values of the maxima of the determinant for bordered two circulant matrices. **Results:** We present the formulas that refine the overly optimistic Elich – Wojtas' bound for the case of matrices of fixed structure with orders $4t+1$ (adamarides) and $4t–1$ (mersennides). In addition to the lower and upper bounds, piecewise-smooth quadratic approximations, which are closer to the values of the extremal matrix determinants are given. We present algorithms for mining (deep search) of matrices of the extended Hadamard family by several methods using orbits and compressions of binary sequences. Search results confirming the given bound estimates are also given. **Practical relevance:** The matrices of the maximum determinant are of direct practical importance for the problems of noise-correcting coding, compression and masking of video information

Keywords – Gauss points, Gauss problem, paraboloid lattice structure, orthogonal matrices, Hadamard matrices, maximum determinant matrices, bordered two circulant matrices.

For citation: Balonin N. A., Sergeev M. B. Maximum determinant two circulant matrices with border. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 2–15 (In Russian). doi:10.31799/1684-8853-2023-3-2-15, EDN: JQPBFE

Financial support

The work was supported financially by the Ministry of Science and Higher Education of the Russian Federation, Agreement No. FSRF-2023-0003.

References

1. Orrick W. P. The maximal {–1, 1}-determinant of order 15. *Metrika*, 2005, no. 62, pp. 195–219. doi.org/10.1007/s00184-005-0410-3
2. Orrick W. P., Solomon B. Large determinant sign matrices of order 4k + 1. *Discrete Mathematics*, 2007, no. 307, pp. 226–236.
3. Seberry J., Xia T., Koukouvinos C., Mitrouli M. The maximal determinant and subdeterminants of ±1 matrices. *Linear Algebra and its Applications*, 2003, vol. 373, pp. 297–310. doi:10.1016/S0024-3795(03)00584-6
4. Ehlich H. Determinantenabschätzungen für binäre Matrizen. *Mathematische Zeitschrift*, 1964, no. 83, pp. 123–132 (In German). doi:10.1007/BF01111249
5. Wojtas M. On Hadamard's inequality for the determinants of order non-divisible by 4. *Colloquium Mathematicum*, 1964, vol. 12, pp. 73–83. doi:10.4064/cm-12-1-73-83
6. Osborn J. H. *Hadamard maximal determinant problem: honours thesis*. University of Melbourne, 2002. 144 p. Available at: http://maths-people.anu.edu.au/~osborn/publications/pubsall.html (accessed 15 January 2023).
7. Cohn J. H. E. On determinants with elements ±1. II. *Bulletin of the London Mathematical Societ*, 1989, vol. 21, iss. 1, pp. 36–42. doi.org/10.1112/blms/21.1.36
8. Neubauer M. G., Radcliffe A. J. The maximum determinant of ±1 matrices. *Linear Algebra and its Applications*, 1997, vol. 257, pp. 289–306.
9. Hadamard J. Resolution d'une question relative aux determinants. *Bulletin des Sciences Mathematiques*, 1893, vol. 17, pp. 240–246 (In French).
10. Wang R. *Introduction to Orthogonal Transforms with Applications in Data Processing and Analysis*. Cambridge University Press, 2010. 504 p.
11. Ahmed N., Rao K. R. *Orthogonal Transforms for Digital Signal Processing*. Springer-Verlag Berlin Heidelberg, 2012. 264 p.
12. Jennifer S., Yamada M. *Hadamard Matrices*: *Constructions using Number Theory and Linear Algebra*. Wiley, 2020. 352 p.
13. Balonin N. A., Djokovic D. Z. Symmetry of two-circulant Hadamard matrices and periodic Golay pairs. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 3, pp. 2–16 (In Russian). doi.org/10.15217/issn1684-8853.2015.3.2
14. Colbourn C. J., Dinitz J. H. *Handbook of Combinatorial Designs*. Second Edition. Chapman and Hall, CRC, 2007. 967 p.
15. Balonin N. A., Sergeev M. B., Seberry J., Sinitsyna O. I. Circles on lattices and Hadamard matrices. *Informatsionno-upravliaiushchie sistemy* [*Information and Control Systems*], 2019, no. 3, pp. 2–9 (In Russian). doi.org/10.31799/1684-8853-2019-3-2-9
16. Balonin N. A., Sergeev M. B., Seberry J., Sinitsyna O. I. Circles on lattices and maximum determinant matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 6, pp. 2–11 (In Russian). doi.org/10.31799/1684-8853-2020-6-2-11
17. Turyn R. J. Hadamard matrices, Baumert – Hall units, four symbol sequences, pulse compression and surface wave encodings. *J. Combin. Theory*, Ser. A, 1974, vol. 16, iss. 3, pp. 313–333. doi.org/10.1016/0097-3165(74)90056-9
18. Balonin N. A., Sergeev M. B., Vostrikov A. A. Prime Fermat numbers and maximum determinant matrix conjecture. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 2, pp. 2–9. doi.org/10.31799/1684-8853-2020-2-2-9

19. Matiyasevich Y. On recursive unsolvabilitv of Hilbert's tenth problem. *Proc. of Fourth Intern. Congress on Logic, Methodology and Philosophy of Science*, Bucharest, 1971, Amsterdam, North-Holland, 1973, pp. 89–110.
20. Farmakis N., Kounias S. The excess of Hadamard matrices and optmimal designs. *Discrete Mathematics*, North-Holland, 1987, vol. 67, iss. 2, pp. 165–176. doi.org/10.1016/0012-365X(87)90025-2
21. Barba G. Intorno al teorema di Hadamard sui determinanti a valore Massimo. *Giornale di Matematiche di Battaglini*, 1933, vol. 71, pp. 70–86 (In Italian).
22. Brent R. R., Osborn J. H. General lower bounds of maximal determinants of binary matrices. *Preprint*, 2012.
23. Sergeev A. M., Balonin Yu. N. Majning matric [matrix mining]. *Sbornik dokladov Vtoroj Mezhdunarodnoj nauchnoj konferencii "Obrabotka, peredacha i zashchita informacii v komp'yuternyh sistemah 22"* [Proc. 2nd Int. Conf. "Processing, transmission and protection of information in computer systems 22"]. Saint-Petersburg, 2022, pp. 169–173 (In Russian).
24. Rokicki T. *New records for maximal determinants, based on pairs of circulant matrices*. Available at: http://tomas.rokicki.com/newrec.html (accessed 15 January 2023).
25. *Algorithms on line. Hadamard matrices*. Available at: http://mathscinet.ru/catalogue/index.php (accessed 15 January 2023).

Уважаемые авторы!

При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

Статьи должны содержать изложение новых научных результатов. Название статьи должно быть кратким, но информативным. В названии недопустимо использование сокращений, кроме самых общепринятых (РАН, РФ, САПР и т. п.).

Текст рукописи должен быть оригинальным, а цитирование и самоцитирование корректно оформлено.

Объем статьи (текст, таблицы, иллюстрации и библиография) не должен превышать эквивалента в 20 страниц, напечатанных на бумаге формата А4 на одной стороне через 1,5 интервала Word шрифтом Times New Roman размером 13, поля не менее двух сантиметров.

Обязательными элементами оформления статьи являются: индекс УДК, заглавие, инициалы и фамилия автора (авторов), ученая степень, звание (при отсутствии — должность), полное название организации, аннотация и ключевые слова на русском и английском языках, ORCID и электронный адрес одного из авторов. При написании аннотации не используйте аббревиатур и не делайте ссылок на источники в списке литературы. Предоставляйте подрисуночные подписи и названия таблиц на русском и английском языках.

Статьи авторов, не имеющих ученой степени, рекомендуется публиковать в соавторстве с научным руководителем, наличие подписи научного руководителя на рукописи обязательно; в случае самостоятельной публикации обязательно предоставляйте заверенную по месту работы рекомендацию научного руководителя с указанием его фамилии, имени, отчества, места работы, должности, ученого звания, ученой степени.

Простые **формулы** набирайте в Word, сложные с помощью редактора Mathtype или Equation. Для набора одной формулы не используйте два редактора; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта в Mathtype никогда не пользуйтесь вкладкой Other, Smaller, Larger, используйте заводские установки редактора, не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст; пробелы в формуле ставьте только после запятых при перечислении с помощью Ctrl+Shift+Space (пробел); не отделяйте пробелами знаки: + = – ×, а также пространство внутри скобок; для выделения греческих символов в Mathtype полужирным начертанием используйте Style → Other → bold.

Для набора формул в Word никогда не используйте вкладки: «Уравнение», «Конструктор», «Формула» (на верхней панели: «Вставка» — «Уравнение»), так как этот ресурс предназначен только для внутреннего использования в Word и не поддерживается программами, предназначенными для изготовления оригинал-макета журнала.

При наборе символов в тексте помните, что символы, обозначаемые латинскими буквами, набираются светлым курсивом, русскими и греческими — светлым прямым, векторы и матрицы — прямым полужирным шрифтом.

Подробнее см. pdf-файл «Правила подготовки рукописей» (стр. 11) на сайте https://guap.ru/ric

Иллюстрации:

— рисунки, графики, диаграммы, блок-схемы предоставляйте в виде отдельных исходных файлов, поддающихся редактированию, используя векторные программы: Visio (*.vsd, *.vsdx); Adobe Illustrator (*.ai); Coreldraw (*.cdr, версия не выше 15); Excel (*.xls); Word (*.docx); AutoCad, Matlab (экспорт в PDF, EPS, SVG, WMF, EMF); Компас (экспорт в PDF), веб-портал DRAW.IO (экспорт в PDF);

— фото и растровые — в формате *.tif, *.png с максимальным разрешением (не менее 300 pixels/inch).

Наличие подрисуночных подписей и названий таблиц на русском и английском языках обязательно (желательно не повторяющих дословно комментарии к рисункам в тексте статьи).

В редакцию предоставляются:

— сведения об авторе (фамилия, имя, отчество, место работы, должность, ученое звание, учебное заведение и год его окончания, ученая степень и год защиты диссертации, область научных интересов, количество научных публикаций, домашний и служебный адреса и телефоны, e-mail), фото авторов: анфас, в темной одежде на белом фоне, должны быть видны плечи и грудь, высокая степень четкости изображения без теней и отблесков на лице, фото можно представить в электронном виде в формате *.tif, *.png, *.jpg с максимальным разрешением — не менее 300 pixels/inch при минимальном размере фото 40×55 мм;

— экспертное заключение.

Список литературы составляется по порядку ссылок в тексте и оформляется следующим образом:

— для книг и сборников — фамилия и инициалы авторов, полное название книги (сборника), город, издательство, год, общее количество страниц, doi;

— для журнальных статей — фамилия и инициалы авторов, полное название статьи, название журнала, год издания, номер журнала, номера страниц, doi;

— ссылки на иностранную литературу следует давать на языке оригинала без сокращений;

— при использовании web-материалов указывайте адрес сайта и дату обращения.

Список литературы оформляйте двумя отдельными блоками по образцам lit.dot на сайте журнала (http://i-us.ru/paperrules): Литература и References.

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Руководство для авторов».

Контакты

Куда: 190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А, ГУАП, РИЦ
Кому: Редакция журнала «Информационно-управляющие системы»
Тел.: (812) 494-70-02
Эл. почта: ius.spb@gmail.com
Сайт: www.i-us.ru

УДК 004.934
doi:10.31799/1684-8853-2023-3-16-25
EDN: YOLUUY

Научные статьи
Articles

Система автоматического распознавания карельской речи

И. С. Кипяткова[a], *канд. техн. наук, доцент, http://orcid.org/0000-0002-1264-4458, kipyatkova@iias.spb.su*
И. А. Кагиров[a], *научный сотрудник, orcid.org/0000-0003-1196-1117*
[a]*Санкт-Петербургский Федеральный исследовательский центр РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ*

Введение: *в последнее время растет число исследований, посвященных автоматической обработке малоресурсных языков. Отсутствие или малый объем обучающих данных является существенным препятствием в развитии речевых технологий для подобных языков.* ***Цель:*** *разработать систему автоматического распознавания речи на карельском языке.* ***Результаты:*** *представлена система автоматического распознавания карельской речи. Обучены акустические модели на основе искусственных нейронных сетей с временными задержками и скрытых марковских моделей. Обучение осуществлялось на речевом корпусе, составленном из записей радиопередач и аудиоданных, полученных путем аугментации. Модель карельского языка обучалась как на письменных текстах, так и на расшифровках обучающей части речевого корпуса. Во время обучения исследовались различные коэффициенты для интерполяции языковой модели, обученной на расшифровках, с моделью языка, обученной на письменных текстах. В ходе экспериментов по распознаванию карельской речи лучший результат по показателю количество неправильно распознанных слов составил 25,81 %, что сопоставимо с общим уровнем распознавания речи для других малоресурсных языков. Собран обучающий набор данных, который включает звукозаписи на карельском языке с расшифровками, а также текстовый корпус.* ***Практическая значимость:*** *полученные решения могут играть роль в создании автоматических систем распознавания не только карельского, но и других малоресурсных языков. Разработанная система поможет исследователям карельского языка, предоставляя эффективный инструмент для записи и обработки карельского языкового материала.*

Ключевые слова *— малоресурсные языки, автоматическое распознавание речи, карельский язык, искусственные нейронные сети с временной задержкой.*

Для цитирования: Кипяткова И. С., Кагиров И. А. Система автоматического распознавания карельской речи. *Информационно-управляющие системы*, 2023, № 3, с. 16–25. doi:10.31799/1684-8853-2023-3-16-25, EDN: YOLUUY
For citation: Kipyatkova I. S., Kagirov I. A. Automatic speech recognition system for Karelian. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 16–25 (In Russian). doi:10.31799/1684-8853-2023-3-16-25, EDN: YOLUUY

Введение

Современные технологии обработки естественного языка активно развиваются и находят применение в различных областях, включая автоматическое распознавание речи. Однако отсутствие развитых речевых технологий для малоресурсных языков остается актуальной проблемой [1]. В связи с этим авторы настоящего исследования предлагают ее решение для карельского языка.

Карельский язык — это язык из финно-угорской языковой семьи, который в настоящее время используется на территории Республики Карелия (Россия). Он относится к прибалтийско-финской группе и близок к вепсскому и ижорскому языкам [2] как генетически, так и типологически. Карельский язык также имеет типологическое сходство с другими региональными языками, в том числе с финским и эстонским. Карельский язык относится к уязвимым языкам, в частности из-за уменьшения числа носителей.

Создание системы автоматической обработки естественного языка составляет сложную и нетривиальную задачу, и одним из важнейших условий ее реализации является наличие обучающих наборов данных. Так как объем электронных языковых ресурсов для карельского языка невелик, то этот вопрос является принципиальным. Также важным моментом при разработке системы автоматического распознавания речи является выбор оптимального алгоритма обучения системы. В рамках данной статьи рассмотрен ряд систем распознавания речи, созданных как для малоресурсных, так и для генетически близких (прибалтийско-финских) языков, и проводится сравнение нескольких подходов к автоматическому распознаванию.

Настоящая работа может иметь определенное значение с точки зрения разработки информационного обеспечения для малоресурсных языков [3–5], поскольку полученные в ходе работы решения помогут при создании автоматических систем распознавания речи не только для карельского, но и других языков.

Аналитический обзор систем автоматического распознавания речи для прибалтийско-финских и малоресурсных языков

Несколько работ, появившихся за последнее время, позволяют оценить эффективность подходов к распознаванию речи, применявшихся для родственных карельскому языков, — финского и эстонского. Несмотря на то, что оба этих языка не являются малоресурсными и примененные методики не могут быть напрямую перенесены на карельский материал, информация о методах является достаточно ценной в силу схожести языков.

В статье [6] описан речевой корпус, состоящий из записей заседаний Парламента Финляндии и снабженный автоматическими расшифровками. Авторы провели анализ качества распознавания речи на этом корпусе, используя несколько метрик и методов оценки. Сравнивались различные модели распознавания речи, включая скрытые марковские модели (СММ), гибридные модели, объединяющие искусственные нейронные сети (ИНС) и СММ — ИНС/СММ, а также модели с архитектурой кодер-декодер с механизмом внимания (Encoder-Decoder with the Attention Mechanism). Из проведенных экспериментов следует, что наилучшие результаты достигаются при использовании гибридных ИНС/СММ моделей. Кроме того, было выяснено, что наличие предобученных на других языках моделей может привести к улучшению качества распознавания речи на финском языке.

В работе [7] представлена система распознавания речи на эстонском языке, обученная на речевом корпусе объемом 268,5 ч. Акустическое моделирование осуществлялось факторизованными ИНС с временными задержками (Factorized Time Delay Neural Networks, TDNN-F) совместно с СММ. Для обучения акустической модели применялся критерий безрешеточной максимизации взаимной информации (Lattice-Free Maximum Mutual Information) [8]. Гиперпараметры настройки обучения взяты из библиотеки Kaldi Switchboard. Декодирование речи выполнялось с использованием 4-граммной модели языка, кроме того, осуществлялась переоценка гипотез распознавания при помощи нейросетевой модели языка. В результате экспериментов количество неправильно распознанных слов (Word Error Rate, WER) составило 8,1 % на тестовом наборе данных, содержащем записи ток-шоу и телефонных интервью, 12,9 % — при распознавании записей с конференций и 22,7 % – на зашумленных записях.

Развитие описанной выше системы распознавания эстонской речи представлено в работе [9]. Авторы расширили корпус эстонской речи до 761 ч и исследовали интегральный (end-to-end) подход с использованием предобученной модели wav2vec2.0 [10]. Авторам удалось снизить значение WER до 6,9 % на тестовом наборе данных, собранных из записей радио- и телепередач. Следует отметить, что при большом объеме обучающих данных интегральные системы показывают лучшую производительность с точки зрения скорости и точности распознавания речи, однако для их обучения требуется существенно больший объем данных, и при недостатке обучающих данных точность таких систем ниже, чем систем, построенных из отдельных компонентов [11].

В статье [12] отмечается, что в целом интегральный подход к автоматическому распознаванию речи [13, 14] оказывается рациональным для бесписьменных или находящихся под угрозой исчезновения языков, поскольку зачастую сбор переводов на высокоресурсный язык оказывается легче, чем транскрибирование записей исходного языка [15]. Тем не менее создание высококачественной интегральной системы с небольшим количеством исходных параллельных данных представляет собой проблему при отсутствии доступа к параллельным корпусам языковых данных. В том случае, если дополнительных ресурсов для исходного языка нет, хорошим подходом оказывается применение метода переноса знаний (transfer learning), суть которого сводится к использованию результатов обучения родительской модели языка, полученных на большом наборе данных, для инициализации весов в дочерней модели, обученной на данных целевого малоресурсного языка. Например, в работе [16] продемонстрировано, что предварительное обучение системы автоматического распознавания речи на материале английского и французского языков позволяет существенно улучшить точность распознавания для испанского языка. Применение подобного подхода для малоресурсных языков описано в работах [17, 18].

Несмотря на широкое распространение интегрального подхода к распознаванию речи, при разработке систем распознавания для малоресурсных языков чаще всего используют стандартный подход, при котором система строится из трех отдельных компонентов (моделей): акустической, языковой и лексической (словаря), — поскольку при этом подходе требуется меньше данных для обучения модели. Такой подход применялся, например, в исследовании [19] для распознавания речи на малоресурсном сингальском языке (о. Шри-Ланка). Полученные авторами результаты показывают, что применение гибридных ИНС/СММ акустических моделей превосходит использование статистических СММ на

7,48 % по показателю WER на тестовом наборе данных. Наименьшее значение WER (35,16 %) получено при использовании архитектуры ИНС с временными задержками (Time Delay Neural Network, TDNN) [20] для создания акустической модели.

Авторами работы [21] представлены результаты экспериментов по многоязычному распознаванию речи на малоресурсных языках (10 языков из набора, предложенного в рамках соревнования OpenASR20 (https://sat.nist.gov/openasr20), а также на североамериканских языках кри и инуитских. В работе исследовалось применение TDNN-F в гибридных ИНС/СММ акустических моделях и показано, что в этом случае значения WER ниже, чем при применении двунаправленных ИНС с долгой кратковременной памятью (LSTM). Тем не менее представленные значения WER достаточно большие и варьируются от 48 до 69,6 % в зависимости от языка [21]. Похожий результат был получен для сомалийского языка [22].

В различных исследованиях, проводимых для русского языка, также было установлено, что использование гибридных акустических моделей на основе TDNN по точности распознавания речи превосходит использование СММ с гауссовыми смесями, а также гибридных ИНС/СММ с другими архитектурами нейронных сетей [23, 24].

Практически обязательным этапом создания системы распознавания речи для малоресурсных языков является аугментация данных — метод создания дополнительных данных путем изменения (модификации) собранных обучающих данных. К распространенным методам аугментации речевых данных относятся изменение частоты основного тона, темпа речи, преобразование голоса, изменение спектрограммы, синтез речи [25–27]. Для расширения набора текстовых данных может выполняться контекстная аугментация, аугментация на основе замены символов или слов, а также обратный перевод [28–30]. Чаще всего наилучшие результаты дает применение сразу нескольких методов аугментации [31]. Подробный обзор методов, применяемых при разработке систем автоматического распознавания речи для решения проблемы недостаточного количества обучающих данных, представлен в работе [32].

По результатам проведенного в ходе данного исследования обзора было принято решение использовать стандартный подход к построению системы распознавания речи для карельского языка, гибридные ИНС/СММ акустические модели, а также методы аугментации речевых данных для расширения обучающего речевого корпуса.

Речевые и текстовые данные для обучения системы распознавания карельской речи

Обучение системы автоматического распознавания речи производится с использованием двух корпусов: речевого и текстового. В качестве речевого корпуса в рамках выполненного исследования использовались записи радиопередач на карельском языке, представляющие собой интервью с двумя и более дикторами (в общей сложности 15, из них 6 мужчин и 9 женщин). Были проведены обработка и аннотирование данного корпуса. Одна из проблем, возникших в ходе работы, связана с наложениями речи, т. е. с фразами, содержащими одновременную речь двух дикторов. Другую проблему составили фоновые шумы, которые сильно ухудшили качество записей. Хотя для создания корпуса использовались только записи студийного качества, фоновый шум в некоторых случаях все же имел место. Все записи с фоновым шумом и наложениями речи были удалены из базы данных.

Одной из особенностей современного карельского языка является переключение кодов [33], которое представляет собой спонтанный переход говорящего с одного языка или диалекта на другой. В данное время среди носителей карельского языка в России распространено карельско-русское двуязычие [34], поэтому переход на русский язык и обратно вполне естественен. Поскольку в настоящем исследовании не ставилась задача обработки явления переключения кодов, фразы, содержащие слова на русском языке, не были включены в корпус.

Итоговый объем речевого корпуса составил 3,5 ч, общее число записанных фраз — 3819. Корпус разбит на обучающую и тестовую части в соотношении 9:1 (табл. 1).

Дополнительно для расширения обучающей части речевого корпуса проведена аугментация

■ ***Таблица 1.*** Характеристики речевого корпуса
■ ***Table 1.*** Speech corpus features

Параметр	Значение
Количество дикторов	15 (6 муж., 9 жен.)
Общая продолжительность слитной речи	3,5 ч
Общий объем данных	2,2 Гб
Количество фраз	3819
Частота дискретизации аудио	16 000 Гц
Квантование сигнала	16 бит
Соотношение обучающей/тестовой части	9:1

■ ***Таблица 2.*** Характеристики текстового корпуса
■ ***Table 2.*** Text corpus features

Параметр	Значение
Общий объем	5 млн словоупотреблений
Соотношение объема материала по источникам, %:	
книги	22,6
периодические издания	73,1
тексты из корпуса «ВепКар»	3,8
расшифровки аудиоматериала	0,5
Нормализация	Сегментация текста по предложениям Замена заглавных букв на строчные Удаление знаков препинания

данных, для чего был использован инструментарий Sox (http://sox.sourceforge.net/sox.html), с помощью которого выполнено изменение темпа речи и высоты голоса диктора, что позволило увеличить объем обучающего речевого материала в два раза.

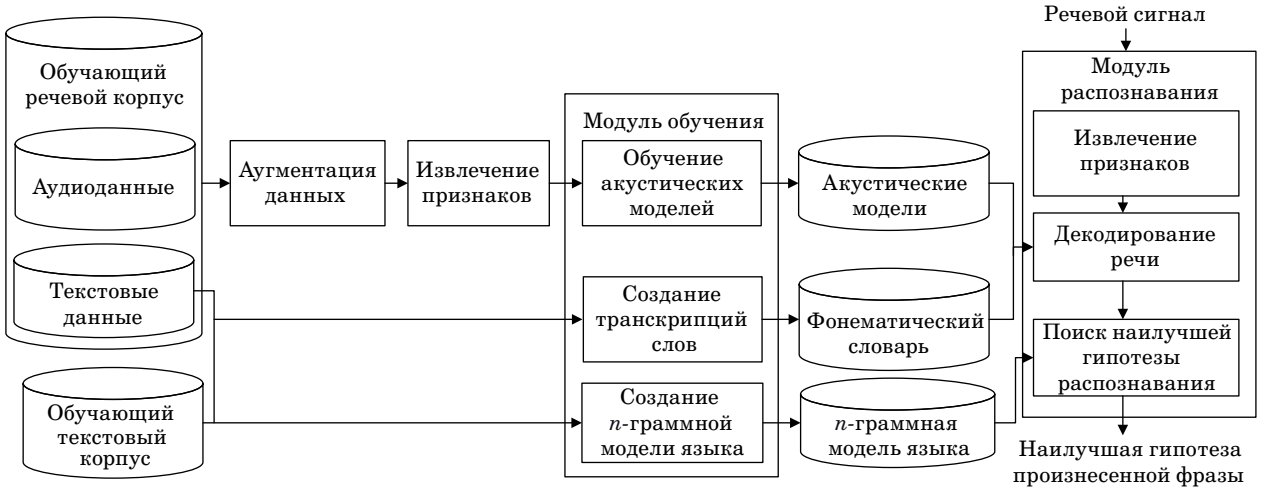
Собранный в ходе данного исследования текстовый корпус включает материалы из таких источников, как печатные издания, периодика на ливвиковском наречии, тексты из открытого корпуса на вепсском и карельском языках («ВепКар» – http://dictorpus.krc.karelia.ru/ru), а также расшифровки аудиозаписей обучающей части речевого корпуса (табл. 2). Все тексты приведены в формат .txt. Текстовая часть базы данных получена с частичным применением полуавтоматического распознавания текста.

Более подробно процесс сбора и обработки речевого и текстового корпусов описан в работе [35].

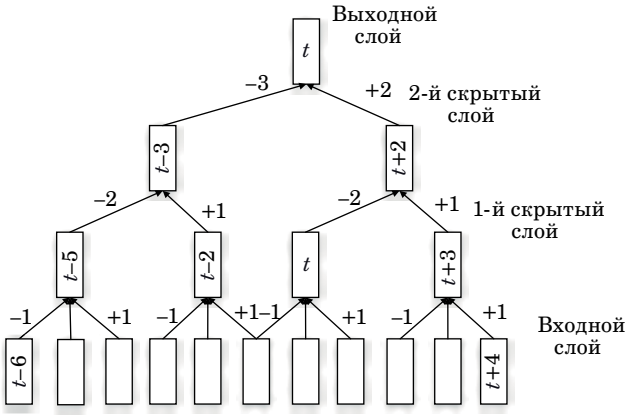
Система распознавания карельской речи

Структура системы распознавания карельской речи представлена на рис. 1.

В качестве акустической модели использовалась гибридная модель с TDNN, аналогичная архитектуре, показавшей наилучшие результаты для русской речи [24]. Обучение осуществлялось с помощью библиотеки nnet2 для рецепта swbd (s5c) инструментария Kaldi [8], при этом был применен стандартный метод обратного распространения ошибки (backpropagation), в качестве функции потерь применялась перекрестная энтропия. Для сокращения времени обучения использовалась технология увеличения скорости обучения, которая предполагает, что веса элементов в скрытом слое обучаются только на некоторых временных шагах, а не на каждом временном шаге [20]. На рис. 2 показан пример архитектуры для TDNN при временном контексте [–6, 4], интервал состоит из целых чисел,



■ ***Рис. 1.*** Структура системы распознавания карельской речи
■ ***Fig. 1.*** Outline of the Karelian speech recognition system



■ ***Рис. 2.*** Пример архитектуры TDNN с применением технологии объединения для временного контекста

■ ***Fig. 2.*** An example of TDNN architecture with sub-sampling

соответствующих временным шагам. Входной слой объединяет фреймы в интервале {–1, 0, 1} (более компактно это можно записать как [–1, 1]). Для скрытого слоя объединение осуществляется для временных шагов {–2, 1}, это означает, что объединяются фрейм, располагающийся на временном контексте за два фрейма до текущего, и фрейм, располагающийся через один фрейм после текущего. На втором скрытом слое объединяется фрейм, располагающийся за три фрейма до текущего, и фрейм, располагающийся через два фрейма после текущего ({–3, 2}).

Входными данными для нейронной сети были мел-частотные кепстральные коэффициенты, при этом для адаптации к речи диктора к ним был добавлен 100-мерный *i*-вектор, использование которого, как показано в работе [36], позволяет снизить WER. В данной работе были обучены ИНС с применением активационной функции p-norm [37], которая вычисляется следующим образом:

$$y = \|\mathbf{x}\|_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}},$$

где вектор **x** представляет небольшую группу входных данных (мини-батч). Величина p выбирается опытным путем, в работе [37] показано, что p=2 дает лучшие результаты. Выходными данными нейронной сети являются апостериорные вероятности контекстно-зависимых моделей фонем.

Для ИНС с активационной функцией p-norm вместо параметра размерности скрытого слоя используются два параметра: число входов и число выходов. Отношение числа входов к числу выходов должно быть целым числом. Обычно используется отношение 5 или 10 [37]. Были созданы ИНС с различным числом скрытых слоев, различным временным контекстом и различными индексами объединяемых элементов. В процессе обучения коэффициент скорости обучения уменьшался от 0,02 до 0,004 в течение 15 эпох, затем пять эпох обучение происходило с постоянным коэффициентом скорости обучения, равным 0,004.

В словарь системы распознавания вошли все слова из расшифровок обучающей части речевого корпуса и слова из остальных текстов, которые встретились не менее двух раз. Это связано с тем, что часть текстов была представлена в виде графических данных, которые были преобразованы в текстовый формат путем полуавтоматического распознавания текста, поэтому в некоторых текстах могли возникнуть ошибки. Таким образом, слова, которые встретились только один раз, зачастую являлись именно словами с ошибками. Итоговый размер словаря составил 143,5 тыс. слов. Фонематические транскрипции создавались автоматически с помощью специально разработанного программного модуля, выполняющего преобразование графема-фонема для поданного на вход списка слов на карельском языке. Более подробно процесс создания фонематического словаря представлен в [35].

Триграммная модель языка обучена с помощью программных средств SRILM [38]. Стоит отметить, что оптимально было бы обучать модель языка на расшифровках спонтанной речи, однако объем таких данных обычно не велик, поэтому чаще всего для этой цели используются письменные тексты. В то же время письменная речь сильно отличается от разговорной, что снижает качество моделей. В ходе исследования триграммная модель языка создавалась двумя способами. При первом способе модель обучалась сразу на всех текстах, включая расшифровки обучающей части речевого корпуса. Второй способ состоял в том, что вначале отдельно обучались две модели языка: одна обучалась на расшифровках обучающей части речевого корпуса, вторая — на остальных текстах. Затем была выполнена линейная интерполяция созданных моделей, при этом коэффициент интерполяции модели, обученной на расшифровках, задавался выше, чем для модели, обученной на письменных текстах. Были проведены эксперименты с использованием разных значений весового коэффициента интерполяции.

Результаты экспериментов по распознаванию карельской речи

Для декодирования речевого сигнала использовался декодер Kaldi на основе взвешенных

конечных преобразователей [8]. Оценка работы системы распознавания речи проводилась по показателю WER. Вначале были проведены эксперименты по автоматическому распознаванию карельской речи с моделью языка, обученной сразу на всех текстовых данных. При использовании акустической модели на основе гауссовых смесей значение WER составило 40,00 %. Результаты экспериментов по автоматическому распознаванию карельской речи с гибридными акустическими моделями на базе архитектуры TDNN с различным временным контекстом и различным отношением числа входов к числу выходов представлены в табл. 3. Наилучшие результаты получены при использовании архитектуры TDNN с пятью скрытыми слоями и временным контекстом [–8, 4] и отношением числа входов к числу выходов, равным 1000/100. Увеличение длины контекста и отношения числа входов/выходов привело к ухудшению качества распознавания, что может быть вызвано переобучением ИНС.

Затем были проведены эксперименты с использованием модели языка, созданной вторым способом, — путем линейной интерполяции модели, обученной на расшифровках аудиоданных, с моделью, обученной на текстах с различными коэффициентами интерполяции. В ходе этих экспериментов использовалась акустическая модель, которая дала наилучшие результаты в предыдущем эксперименте (модель с отношением числа входов к числу выходов, равным 1000/100, и временным контекстом [–8, 4]). Результаты экспериментов представлены в табл. 4, где также указаны значения коэффициента неопределенности (perplexity) для каждой модели языка, вычисленные по текстам из расшифровок тестовой части речевого корпуса. Количество внесловарных слов (out-of-vocabulary words) составило 6 %. Скорость распознавания речи без использования графического процессора на компьютере с многоядерным процессором с тактовой частотой 4 ГГц составила около 1,3 RTF (real-time factor).

Наилучшие результаты были получены при использовании коэффициента интерполяции 0,7 для модели, обученной на расшифровках. Следует отметить, что текстовые данные, полученные из расшифровок, наиболее адекватно отражают разговорный язык, поскольку в них представлена спонтанная речь, на которую не накладываются стилистические правила, характерные для литературного языка.

■ ***Таблица 3.*** Значения WER, полученные с применением гибридных акустических моделей с различными параметрами
■ ***Table 3.*** WER for hybrid acoustic models with different parameters

Длина контекста	Контекст послойно					WER, %, для		
	1	2	3	4	5	500/50[1]	1000/100[1]	2000/200[1]
[–6, 4]	[–1, 1]	{–2, 1}	{–3, 2}	{0}	–	29,46	29,32	31,22
[–6, 6]	[–1, 1]	{–2, 2}	{–3, 3}	{0}	–	31,08	31,49	29,86
[–7, 7]	[–1, 1]	{–2, 2}	{–4, 4}	{0}	–	31,35	31,62	29,05
[–7, 7]	[–2, 2]	{–1, 1}	{–2, 2}	{–2, 2}	{0}	30,14	30,27	30,00
[–8, 4]	[–1, 1]	{–3, 1}	{–4, 2}	{0}	–	30,81	**28,51**	28,65
[–8, 5]	[–2, 2]	{–1, 1}	{–2, 1}	{–3, 1}	{0}	29,46	28,92	28,78
[–8, 8]	[–2, 2]	{–1, 1}	{–2, 2}	{–3, 3}	{0}	29,32	29,32	29,46

[1]Количество входов/выходов.

■ ***Таблица 4.*** Результаты экспериментов с применением различных моделей языка
■ ***Table 4.*** Experimental results with different language models

Способ обучения модели языка	Весовой коэффициент модели языка, обученной на расшифровках	Коэффициент неопределенности	WER, %
На всех текстах сразу	–	4030,06	28,51
Путем линейной интерполяции	0,5	2118,88	26,35
	0,6	2083,11	26,08
	0,7	2095,15	**25,81**
	0,8	2172,63	26,89

Заключение

В работе представлена система автоматического распознавания речи для ливвиковского наречия карельского языка. Для повышения точности работы системы выполнена аугментация речевых данных, кроме того, отдельно обучена модель языка на расшифровках обучающей части речевого корпуса, которая затем интерполировалась с моделью языка, обученной на текстовых данных. Ошибка распознавания слов, полученная в ходе проведенных экспериментов, составила 25,81 %, что, конечно, хуже, чем современные результаты, получаемые для языков с большими ресурсами данных, но находится на уровне мировых результатов для других малоресурсных языков.

Другим важным результатом явилось создание обучающего набора данных, состоящего из звукозаписей на карельском языке, их текстовых транскрипций и собственно текстового корпуса. В представленном исследовании набор данных использован для обучения языковой и акустической моделей, однако его можно использовать и в других исследованиях по карельскому языку в области обработки естественного языка.

Одна из главных проблем, возникших при создании системы автоматического транскрибирования карельского языка, состоит в нехватке данных. Авторы статьи решили эту проблему, во-первых, созданием собственного набора данных и, во-вторых, применением аугментации.

Однако, несмотря на достигнутые результаты, существует потребность в дальнейших исследованиях и улучшении разработанной системы. Так, проблема переключения кодов осталась за рамками настоящего исследования. В общем и целом расширение обучающего набора данных существенно повысит качество разработанной системы. В дальнейшем планируется исследовать метод переноса знаний при обучении акустических моделей, а также применить нейросетевой подход к обучению модели языка.

Результаты настоящего исследования представляют собой вклад в развитие технологий автоматической обработки карельского языка и могут быть использованы при создании подобных систем для иных малоресурсных языков.

Финансовая поддержка

Работа выполнена при финансовой поддержке фонда РНФ (проект № 22-21-00843).

Литература

1. **Joshi P., Santy S., Budhiraja A., Bali K., Choudhury M.** The state and fate of linguistic diversity and inclusion in the NLP world. *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 6282–6293. doi:10.48550/arXiv.2004.09095
2. **Афанасьева А. А., Муллонен И. И.** Кареловепсский диалог на карте южной Карелии. *Acta Linguistica Petropolitana*, 2020, no. 16(3), с. 9–28. doi:10.30842/alp2306573716301
3. **Krauwer S.** The basic language resource kit (BLARK) as the first milestone for the language resources roadmap. *Proc. of Intern. Workshop on Speech and Computer (SPECOM-2003)*, 2003, pp. 8–15.
4. **Berment V.** *Méthodes pour informatiser des langues et des groupes de langues «peu dotées»*: Doct. Diss. Grenoble, 2004. 278 p. https://theses.hal.science/tel-00006313/document (дата обращения: 28.04.2023).
5. **Cieri Ch., Maxwell M., Strassel S., Tracey J.** Selection criteria for low resource language programs. *Proc. of the Tenth Intern. Conf. on Language Resources and Evaluation (LREC'16)*, 2016, pp. 4543–4549.
6. **Virkkunen A., Rouhe A., Phan N., Kurimo M.** Finnish parliament ASR corpus. *Language Resources and Evaluation*, 2023. https://doi.org/10.1007/s10579-023-09650-7 (дата обращения: 28.04.2023)
7. **Alumäe T., Tilk O., Asadullah.** Advanced rich transcription system for Estonian speech. *Frontiers in Artificial Intelligence and Applications; Ebook*, 2018, vol. 307. doi:10.3233/978-1-61499-912-6-1
8. **Povey D., Ghoshal A., Boulianne G., Burget L., Glembek O., Goel N., Hannemann M., Motlíček P., Qian Y., Schwarz P., Silovský J., Stemmer G., Vesel K.** The Kaldi speech recognition toolkit. *Proc. of 2011 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 2011, pp. 1–4.
9. **Olev A., Alumäe T.** Estonian speech recognition and transcription editing service. *Baltic Journal of Modern Computing*, 2022, vol. 10(3), pp. 409–421. doi:10.22364/bjmc.2022.10.3.14
10. **Baevski A., Zhou Y., Mohamed A., Auli M.** wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, 2020, vol. 3, pp. 12449–12460. doi:10.48550/arXiv.2006.11477
11. **Марковников Н. М., Кипяткова И. С.** Исследование методов построения моделей кодер-декодер для распознавания русской речи. *Информационно-управляющие системы*, 2019, № 4, с. 45–53. doi:10.31799/1684-8853-2019-4-45-53
12. **Stoian M., Bansal S., Goldwater Sh.** Analyzing ASR pretraining for low-resource speech-to-text translation. *Proc. of IEEE Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2020)*,

2020, pp. 7909–7913. doi:10.1109/ICASSP40776.2020.9053847

13. **Sperber M., Neubig G., Niehues J., Waibel A.** Attention passing models for robust and data-efficient end-to-end speech translation. *Transactions of the Association for Computational Linguistics*, 2019, vol. 7, pp. 313–325. doi:10.1162/tacl_a_00270
14. **Salesky E., Sperber M., Waibel A.** Fluent translations from disfluent speech in end-to-end speech translation. *Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, vol. 1 (Long and Short Papers), pp. 2786–2792. doi:10.18653/v1/N19-1285
15. **Godard P., Adda G., Adda-Decker M., Benjumea J., Besacier L., Cooper-Leavitt J., Kouarata G.-N., Lamel L., Maynard H., Mueller M., Rialland A., Stueker S., Yvon F., Zanon-Boito M.** A very low resource language speech corpus for computational language documentation experiments. *Proc. of the Eleventh Intern. Conf. on Language Resources and Evaluation (LREC 2018)*, 2018, pp. 3366–3370.
16. **Bansal S., Kamper H., Livescu K., Lopez A., Goldwater S.** Pre-training on high-resource speech recognition improves low resource speech-to-text translation. *Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, vol. 1 (Long and Short Papers), pp. 58–68. doi:10.18653/v1/N19-1006
17. **Wet de F., Kleynhans N., Compernolle van D., Sahraeian R.** Speech recognition for under-resourced languages: Data sharing in hidden Markov model systems. *South African Journal of Science*, 2017, vol. 113, no. 1–2, pp. 1–9. doi:10.17159/sajs.2017/20160038
18. **Woldemariam Y.** Transfer learning for less-resourced semitic languages speech recognition: The case of Amharic. *Proc. of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, 2020, pp. 61–69.
19. **Karunathilaka H., Welgama V., Nadungodage T., Weerasinghe R.** Low-resource Sinhala speech recognition using deep learning. *Proc. of 2020 20th Intern. Conf. on Advances in ICT for Emerging Regions (ICTer)*, 2020, pp. 196–201. doi:10.1109/ICTer51097.2020.9325468
20. **Peddinti V., Povey D., Khudanpur S.** A time delay neural network architecture for efficient modeling of long temporal contexts. *Proc. of INTERSPEECH-2015*, 2015, pp. 3214–3218. doi:10.21437/Interspeech.2015-647
21. **Gupta V., Boulianne G.** Progress in multilingual speech recognition for low resource languages Kurmanji Kurdish, Cree and Inuktut. *Proc. of the 13th Conf. on Language Resources and Evaluation (LREC 2022)*, 2022, pp. 6420–6428.
22. **Biswas A., Menon R., Westhuizen van der E., Niesler Th.** Improved low-resource Somali speech recognition by semi-supervised acoustic and language model training. *Proc. of INTERSPEECH-2019*, 2019. *arXiv preprint*, 2019. arXiv:1907.03064 (дата обращения: 28.04.2023). doi:10.21437/Interspeech.2019-1328
23. **Обухов Д. С.** Разработка современной системы распознавания русскоязычной телефонной речи. *Управление большими системами*, 2021, № 89, с. 106–122. doi:10.25728/ubs.2021.89.4
24. **Kipyatkova I.** Improving Russian LVCSR using deep neural networks for acoustic and language modeling. *Proc. of the 20th Intern. Conf. on Speech and Computer SPECOM-2018*, 2018, vol. 11096, pp. 291–300. doi:10.1007/978-3-319-99579-3_31
25. **Park D. S., Chan W., Zhang Y., Chiu Ch.-Ch., Zoph B., Cubuk E. D., Le Q. V.** Specaugment: A simple data augmentation method for automatic speech recognition. *Proc. of INTERSPEECH-2019*, 2019, pp. 2613–2617. doi:10.21437/Interspeech.2019-2680
26. **Kaneko T., Kameoka H., Tanaka K., Hojo N.** CycleGAN-VC3: Examining and improving CycleGAN-VCs for mel-spectrogram conversion. *Proc. of INTERSPEECH-2020*, 2020, pp. 2017–2021. doi:10.21437/Interspeech.2020-2280
27. **Rebai I., BenAyed Y., Mahdi W., Lorré J. P.** Improving speech recognition using data augmentation and acoustic model fusion. *Procedia Computer Science*, 2017, vol. 112, pp. 316–322. doi:10.1016/j.procs.2017.08.003
28. **Şahin G. G.** To augment or not to augment? A comparative study on text augmentation techniques for low-resource NLP. *Computational Linguistics*, 2022, vol. 48, no. 1, pp. 5–42. doi:10.1162/coli_a_00425
29. **Kobayashi S.** Contextual augmentation: Data augmentation by words with paradigmatic relations. *Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* 2018, vol. 2 (Short Papers), pp. 452–457. doi:10.18653/v1/N18-2072
30. **Sennrich R., Haddow B., Birch A.** Improving neural machine translation models with monolingual data. *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics*, 2016, vol. 1 (Long Papers), pp. 86–96. doi:10.18653/v1/P16-1009
31. **Gokay R., Yalcin H.** Improving low resource Turkish speech recognition with data augmentation and TTS. *Proc. of 2019 16th Intern. Multi-Conf. on Systems, Signals and Devices (SSD)*, 2019, pp. 357–360. doi:10.1109/SSD.2019.8893184
32. **Кипяткова И. С., Кагиров И. А.** Аналитический обзор методов решения проблемы малых наборов данных при создании систем автоматического распознавания речи для малоресурсных языков. *Информатика и автоматизация*, 2022, № 21(4), с. 678–709. doi:10.15622/ia.21.4.2
33. **Ковалева С. В., Родионова А. П.** *Традиционное и новое в лексике и грамматике карельского языка (по данным социолингвистического исследования).* Петрозаводск, КарНЦ РАН, 2011. 138 с.

https://www.booksite.ru/fulltext/koval/text.pdf (дата обращения: 28.04.2023).
34. **Karjalainen H., Ulriikka P., Riho G., Svetlana K.** Karelian in Russia: ELDIA Case-Specific Report, with contributions by Reetta Toivanen, Anneli Sarhimaa and Eva Kühhirt (Studies in European Language Diversity 26). Research consortium ELDIA, 2013. https://phaidra.univie.ac.at/detail/o:314612 (дата обращения: 28.04.2023)
35. **Кипяткова И. С., Родионова А. П., Кагиров И. А., Крижановский А. А.** Подготовка речевых и текстовых данных для создания системы автоматического распознавания карельской речи. *Ученые записки Петрозаводского государственного университета*, 2023, № 45(5), с. 82–91. (В печати.)
36. **Saon G., Soltau H., Nahamoo D., Picheny M.** Speaker adaptation of neural network acoustic models using i-vectors. *Proc. of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2013, pp. 55–59. doi:10.1109/ASRU.2013.6707705
37. **Zhang X., Trmal J., Povey D., Khudanpur S.** Improving deep neural network acoustic models using generalized maxout networks. *Proc. of 2014 IEEE Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 215–219. doi:10.1109/ICASSP.2014.6853589
38. **Stolcke A., Zheng J., Wang W., Abrash V.** SRILM at sixteen: update and outlook. *Proc. of IEEE Automatic Speech Recognition and Understanding Workshop*, 2011, pp. 5–9.

UDC 004.934
doi:10.31799/1684-8853-2023-3-16-25
EDN: YOLUUY

Automatic speech recognition system for Karelian

I. S. Kipyatkova[a], PhD, Tech., Associate Professor, http://orcid.org/0000-0002-1264-4458, kipyatkova@iias.spb.su
I. A. Kagirov[a], Research Fellow, orcid.org/0000-0003-1196-1117
[a]St. Petersburg Federal Research Center of the RAS, 39, 14th Line, 199178, Saint-Petersburg, Russian Federation

Introduction: There has been a growth in the number of studies devoted to automatic processing of low-resource languages. The lack of training data is a significant obstacle to the development of speech technologies for such languages. **Purpose:** To develop an automatic speech recognition system for Karelian. **Results:** we present a system for automatic speech recognition in Karelian. We have trained acoustic models based on artificial neural networks with time delays and hidden Markov models. We have trained the system with the use of a speech corpus composed of radio broadcast recordings and audio data modified with augmentation techniques. Both written texts and transcripts of a training part of the speech corpus have been involved. We have explored various coefficients to interpolate a language model trained on transcripts with a language model trained on written texts. The best value of the word error rate was 25.81%, which is comparable with the results for other low-resource languages. We have collected a training data set, which includes sound recordings of the Karelian language with transcripts, as well as a text corpus. **Practical relevance:** The results can be of a certain significance for the development of automatic recognition systems not only for Karelian but for other low-resource languages as well. In addition, the developed system may be useful for the researchers of the Karelian language, providing them with an effective tool for recording and processing the Karelian language data.

Keywords — low-resource languages, automatic transcription, the Karelian language, Time Delay Neural Network.

For citation: Kipyatkova I. S., Kagirov I. A. Automatic speech recognition system for Karelian. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 16–25 (In Russian). doi:10.31799/1684-8853-2023-3-16-25, EDN: YOLUUY

Financial support
This work was supported financially by the Russian Science Foundation (project No. 22-21-00843).

References

1. Joshi P., Santy S., Budhiraja A., Bali K., Choudhury M. The state and fate of linguistic diversity and inclusion in the NLP world. *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 6282–6293. doi:10.48550/arXiv.2004.09095
2. Afanasjeva A. A., Mullonen I. I. A Karelian-Veps dialogue on the map of southern Karelia. *Acta Linguistica Petropolitana*, 2020, no. 16(3), pp. 9–28 (In Russian). doi:10.30842/alp2306573716301
3. Krauwer S. The basic language resource kit (BLARK) as the first milestone for the language resources roadmap. *Proc. of Intern. Workshop on Speech and Computer (SPECOM-2003)*, 2003, pp. 8–15.
4. Berment V. *Méthodes pour informatiser des langues et des groupes de langues «peu dotées»*: Doct. Diss. Grenoble, 2004, 278 p. (In French). Available at: https://theses.hal.science/tel-00006313/document (accessed 28 April 2023).
5. Cieri Ch., Maxwell M., Strassel S., Tracey J. Selection criteria for low resource language programs. *Proc. of the Tenth Intern. Conf. on Language Resources and Evaluation (LREC'16)*, 2016, pp. 4543–4549.
6. Virkkunen A., Rouhe A., Phan N., Kurimo M. Finnish parliament ASR corpus. *Language Resources and Evaluation*, 2023. Available at: https://doi.org/10.1007/s10579-023-09650-7 (accessed 28 April 2023).
7. Alumäe T., Tilk O., Asadullah. Advanced rich transcription system for Estonian speech. *Frontiers in Artificial Intelligence and Applications; Ebook*, 2018, vol. 307. doi:10.3233/978-1-61499-912-6-1
8. Povey D., Ghoshal A., Boulianne G., Burget L., Glembek O., Goel N., Hannemann M., Motlíček P., Qian Y., Schwarz P., Silovský J., Stemmer G., Vesel K. The Kaldi speech recognition toolkit. *Proc. of 2011 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 2011, pp. 1–4.
9. Olev A., Alumäe T. Estonian speech recognition and transcription editing service. *Baltic Journal of Modern Computing*, 2022, vol. 10(3), pp. 409–421. doi:10.22364/bjmc.2022.10.3.14

10. Baevski A., Zhou Y., Mohamed A., Auli M. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, 2020, vol. 3, pp. 12449–12460. doi:10.48550/arXiv.2006.11477
11. Markovnikov N. M., Kipyatkova I. S. Encoder-decoder models for recognition of Russian speech. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2019, no. 4, pp. 45–53 (In Russian). doi:10.31799/1684-8853-2019-4-45-53
12. Stoian M., Bansal S., Goldwater Sh. Analyzing ASR pretraining for low-resource speech-to-text translation. *Proc. of IEEE Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2020)*, 2020, pp. 7909–7913. doi:10.1109/ICASSP40776.2020.9053847
13. Sperber M., Neubig G., Niehues J., Waibel A. Attention passing models for robust and data-efficient end-to-end speech translation. *Transactions of the Association for Computational Linguistics*, 2019, vol. 7, pp. 313–325. doi:10.1162/tacl_a_00270
14. Salesky E., Sperber M., Waibel A. Fluent translations from disfluent speech in end-to-end speech translation. *Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, vol. 1 (Long and Short Papers), pp. 2786–2792. doi:10.18653/v1/N19-1285
15. Godard P., Adda G., Adda-Decker M., Benjumea J., Besacier L., Cooper-Leavitt J., Kouarata G.-N., Lamel L., Maynard H., Mueller M., Rialland A., Stueker S., Yvon F., Zanon-Boito M. A very low resource language speech corpus for computational language documentation experiments. *Proc. of the Eleventh Intern. Conf. on Language Resources and Evaluation (LREC 2018)*, 2018, pp. 3366–3370.
16. Bansal S., Kamper H., Livescu K., Lopez A., Goldwater S. Pre-training on high-resource speech recognition improves low resource speech-to-text translation. *Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, vol. 1 (Long and Short Papers), pp. 58–68. doi:10.18653/v1/N19-1006
17. Wet de F., Kleynhans N., Compernolle van D., Sahraeian R. Speech recognition for under-resourced languages: Data sharing in hidden Markov model systems. *South African Journal of Science*, 2017, vol. 113, no. 1–2, pp. 1–9. doi:10.17159/sajs.2017/20160038
18. Woldemariam Y. Transfer learning for less-resourced semitic languages speech recognition: The case of Amharic. *Proc. of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, 2020, pp. 61–69.
19. Karunathilaka H., Welgama V., Nadungodage T., Weerasinghe R. Low-resource Sinhala speech recognition using deep learning. *Proc. of 2020 20th Intern. Conf. on Advances in ICT for Emerging Regions (ICTer)*, 2020, pp. 196–201. doi:10.1109/ICTer51097.2020.9325468
20. Peddinti V., Povey D., Khudanpur S. A time delay neural network architecture for efficient modeling of long temporal contexts. *Proc. of INTERSPEECH-2015*. 2015, pp. 3214–3218. doi:10.21437/Interspeech.2015-647
21. Gupta V., Boulianne G. Progress in multilingual speech recognition for low resource languages Kurmanji Kurdish, Cree and Inuktut. *Proc. of the 13th Conf. on Language Resources and Evaluation (LREC 2022)*, 2022, pp. 6420–6428.
22. Biswas A., Menon R., Westhuizen van der E., Niesler Th. Improved low-resource Somali speech recognition by semi-supervised acoustic and language model training. *Proc. of INTERSPEECH-2019*, 2019. *arXiv preprint*, 2019. arXiv:1907.03064 (accessed 28 April 2023). doi:10.21437/Interspeech.2019-1328
23. Obukhov D. S. Speech recognition system for Russian language telephone speech. *Large-Scale Systems Control*, 2021, no. 89, pp. 106–122 (In Russian). doi:10.25728/ubs.2021.89.4
24. Kipyatkova I. Improving Russian LVCSR using deep neural networks for acoustic and language modeling. *Proc. of the 20th Intern. Conf. on Speech and Computer SPECOM-2018*, 2018, vol. 11096, pp. 291–300. doi:10.1007/978-3-319-99579-3_31
25. Park D. S., Chan W., Zhang Y., Chiu Ch.-Ch., Zoph B., Cubuk E. D., Le Q. V. Specaugment: A simple data augmentation method for automatic speech recognition. *Proc. of INTERSPEECH-2019*, 2019, pp. 2613–2617. doi:10.21437/Interspeech.2019-2680
26. Kaneko T., Kameoka H., Tanaka K., Hojo N. CycleGAN-VC3: Examining and improving CycleGAN-VCs for mel-spectrogram conversion. *Proc. of INTERSPEECH-2020*, 2020, pp. 2017–2021. doi:10.21437/Interspeech.2020-2280
27. Rebai I., BenAyed Y., Mahdi W., Lorré J. P. Improving speech recognition using data augmentation and acoustic model fusion. *Procedia Computer Science*, 2017, vol. 112, pp. 316–322. doi:10.1016/j.procs.2017.08.003
28. Şahin G. G. To augment or not to augment? A comparative study on text augmentation techniques for low-resource NLP. *Computational Linguistics*, 2022, vol. 48, no. 1, pp. 5–42. doi:10.1162/coli_a_00425
29. Kobayashi S. Contextual augmentation: Data augmentation by words with paradigmatic relations. *Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2018, vol. 2 (Short Papers), pp. 452–457. doi:10.18653/v1/N18-2072
30. Sennrich R., Haddow B., Birch A. Improving neural machine translation models with monolingual data. *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics*, 2016, vol. 1 (Long Papers), pp. 86–96. doi:10.18653/v1/P16-1009
31. Gokay R., Yalcin H. Improving low resource Turkish speech recognition with data augmentation and TTS. *Proc. of 2019 16th Intern. Multi-Conf. on Systems, Signals and Devices (SSD)*, 2019, pp. 357–360. doi:10.1109/SSD.2019.8893184
32. Kipyatkova I. S., Kagirov I. A. Analytical review of methods for solving data scarcity issues regarding elaboration of automatic speech recognition systems for low-resource languages. *Informatics and Automation*, 2022, no. 21(4), pp. 678–709 (In Russian). doi:10.15622/ia.21.4.2
33. Kovaleva S. V., Rodionova A. P. *Tradicionnoe i novoe v leksike i grammatike karel'skogo jazyka (Po dannym sociolingvisticheskogo issledovaniya)* [Traditional and Innovative in the Vocabulary and Grammar of Karelian (based on a sociolinguistic research)]. Petrozavodsk, KarNC RAN Publ., 2011. 138 p. (In Russian). Available at: https://www.booksite.ru/fulltext/koval/text.pdf (accessed 28 April 2023).
34. Karjalainen H., Ulriikka P., Riho G., Svetlana K. Karelian in Russia: ELDIA Case-Specific Report, with contributions by Reetta Toivanen, Anneli Sarhimaa and Eva Kühhirt (Studies in European Language Diversity 26). Research consortium ELDIA, 2013. Available at: https://phaidra.univie.ac.at/detail/o:314612 (assessed 28 April 2023).
35. Kipyatkova I. S., Rodionova A. P., Kagirov I. A., Krizhanovsky A. A. Speech and text data preparation for development of an automatic speech recognition system for the Karelian language. *Proceedings of Petrozavodsk State University*, 2023, no. 45(5), pp. 82–91 (In Russian; in print.)
36. Saon G., Soltau H., Nahamoo D., Picheny M. Speaker adaptation of neural network acoustic models using i-vectors. *Proc. of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2013, pp. 55–59. doi:10.1109/ASRU.2013.6707705
37. Zhang X., Trmal J., Povey D., Khudanpur S. Improving deep neural network acoustic models using generalized maxout networks. *Proc. of 2014 IEEE Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 215–219. doi:10.1109/ICASSP.2014.6853589
38. Stolcke A., Zheng J., Wang W., Abrash V. SRILM at sixteen: update and outlook. *Proc. of IEEE Automatic Speech Recognition and Understanding Workshop*, 2011, pp. 5–9.

УДК 654.1
doi:10.31799/1684-8853-2023-3-26-38
EDN: ZZTURK

Научные статьи
Articles

Метод перцептуальной цветокоррекции

А. А. Мотыко[a], *канд. техн. наук, доцент, orcid.org/0000-0003-4241-4298*
Н. А. Обухова[a], *доктор техн. наук, профессор, orcid.org/0000-0003-1953-2085, natalia172419@yandex.ru*
[a]*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ», Профессора Попова ул., 5, Санкт-Петербург, 197376, РФ*

Введение: *в конвейере обработки графической информации широко используются цветокорректирующие преобразования, которые применяются при характеризации сенсора, балансе белого, художественной обработке, согласовании цветовых охватов дисплеев и т. д. Однако в современных методах, применяемых для преобразования координат между различными цветовыми пространствами, свойства зрения учитываются недостаточно, что приводит к ошибкам цветопередачи.* ***Цель:*** *разработать метод оценки параметров цветокорректирующего преобразования с учетом свойств зрительного восприятия человека.* ***Результаты:*** *предложенный метод состоит из двух этапов. На первом этапе выполняется оценка параметров цветокорректирующего преобразования методом наименьших квадратов. На втором этапе полученная оценка уточняется путем численной минимизации целевой функции, построенной на базе перцептуального критерия (основанного на формуле для расчета цветовых различий CIE DE 2000). Экспериментальное исследование показало, что данный метод позволяет значительно повысить точность цветопередачи. Средняя ошибка снижается на 10–50 % в зависимости от используемой модели цветокорректирующего преобразования. Кроме того, на 10–15% снижается доля цветов со значительной с точки зрения человеческого восприятия ошибкой цветопередачи. Метод может использовать различные модели цветокорректирующих преобразований, что делает его универсальным и дает широкие возможности применения в конвейере обработки графической информации.* ***Практическая значимость:*** *разработанный метод оценки параметров цветокорректирующего преобразования с учетом свойств зрения человека может применяться при синтезе телевизионных систем в задачах, где для пользователя важна точная цветопередача, например в цифровой фотографии, телемедицине и пр.*

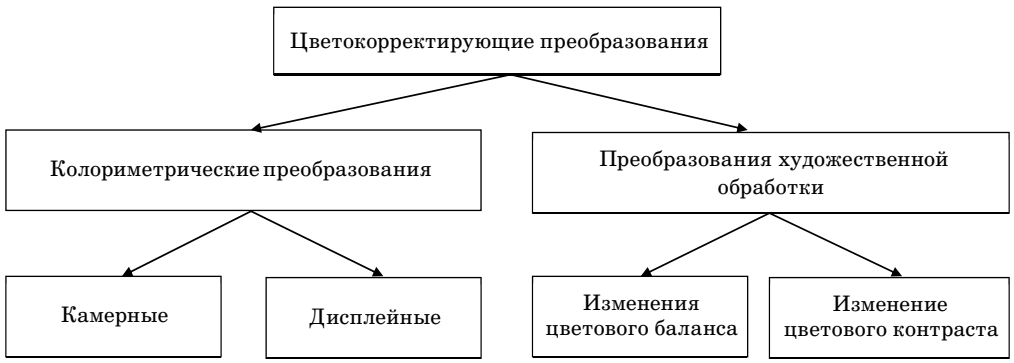
Ключевые слова *— цветокоррекция, CIE DE 2000, цветокорректирующие преобразования, перцептуальная цветокоррекция, формула цветового различия.*

Для цитирования: Мотыко А. А., Обухова Н. А. Метод перцептуальной цветокоррекции. *Информационно-управляющие системы*, 2023, № 3, с. 26–38. doi:10.31799/1684-8853-2023-3-26-38, EDN: ZZTURK
For citation: Motyko A. A., Obukhova N. A. The method for perceptual color correction. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 26–38 (In Russian). doi:10.31799/1684-8853-2023-3-26-38, EDN: ZZTURK

Введение

Цветокорректирующие преобразования используют на различных этапах работы с видеосигналом [1]. Они являются обязательной частью конвейера обработки сигнала при формировании изображения камерой, выполняются в процессе вывода графической информации на дисплей, применяются в рамках обработки снимка для повышения качества, для достижения желаемого художественного эффекта.

Схема классификации цветокорректирующих преобразований по их назначению приведена на рис. 1.



■ ***Рис. 1.*** Классификация цветокорректирующих преобразований
■ ***Fig. 1.*** The color-correcting transformations classification

Задача колориметрических преобразований — обеспечение точной цветопередачи в соответствии с моделью стандартного наблюдателя. Колориметрические преобразования применяются в камерном и дисплейном конвейерах обработки сигнала. При формировании изображения камерой выполняется переход от зависимых от устройства RGB-координат в независимые XYZ-координаты, цветовая адаптация, переход из неперцептуальных пространств в перцептуальные (например, из XYZ в L*a*b*). В дисплейной части колориметрические преобразования служат для согласования цветовых охватов устройств отображения.

Преобразования художественной обработки осуществляют цветокоррекцию, повышающую субъективное качество изображения, например, выполняют манипуляции с цветовым тоном и насыщенностью цветов палитры, чтобы сделать фотографию более привлекательной для наблюдателя, или повышают локальный цветовой контраст с этой же целью. Помимо использования в массовом сегменте при обработке фотографий, указанные преобразования находят применение и в специализированных телевизионных системах. Например, возможность увеличения локального цветового контраста в видеоэндоскопии является обязательной функцией, используемой для улучшения заметности мелких особенностей тканей при проведении обследования врачом.

Колориметрические преобразования, как правило, опираются на информацию о спектральном составе освещения и характеристиках сцены, спектральной чувствительности сенсора. В цветокорректирующих преобразованиях художественной обработки цвета́ преобразуют исходя из некоторых эмпирик (определяемых производителем), направленных на повышение привлекательности изображений в том или ином виде.

Цветовые пространства обычно трехмерные, что связано с трихроматическим зрением человека, поэтому цвет (в частности, элемента разложения изображения) кодируется тремя числами. Для колориметрических цветокорректирующих преобразований характерно то, что результат изменения цвета для каждого пикселя не зависит от его окружения. Таким образом:

$$\mathbf{I}_2(x, y) = f(\mathbf{I}_1(x, y)),$$

где $\mathbf{I}_2$ — цветовые координаты пикселя в целевом цветовом пространстве; (x, y) — координаты пикселя; f — функция, осуществляющая цветокорректирующее преобразование; $\mathbf{I}_1$ — цветовые координаты пикселя в исходном цветовом пространстве.

Задача синтеза цветового преобразования состоит в определении функции f, которая представляет собой некоторую модель с набором параметров $\mathbf{A}$. Для идентификации параметров используют набор цветов с известными координатами в исходном и целевом пространстве. Набор может быть задан в цифровом виде (спектры цветов или координаты) или напечатан в виде таблицы, так называемой диаграммы цветов (ColorChecker). Известны и широко применяются диаграммы Macbeth [2], SG [3], DC [4] и др. При оценке параметров модели преобразования предполагается, что набор цветов O, входящих в цветовое пространство K (прообраза), берется в качестве эталона, а координаты цветов из набора P, входящих в цветовое пространство L (образа), пересчитываются с помощью f так, чтобы в результате минимально отличаться от эталонных цветов O. Таким образом, при синтезе цветокорректирующего преобразования необходимо для функции f определить вид модели и в результате некоторой процедуры идентифицировать параметры $\mathbf{A}$ этой модели, которые обеспечивают минимальное различие между цветами наборов O и P.

Для процедуры оценки параметров модели необходимо сформулировать показатель качества цветового преобразования (или целевую функцию), отражающий ошибку (суммарную, или среднюю) — разницу между истинными значениями цветов и получившимися в результате отображения, при этом важным фактором является используемая метрика, определяющая расстояние между цветами:

$$E_{color} = \frac{1}{C}\sum_{i=1}^{C} E_{color_i} =$$

$$= \frac{1}{C}\sum_{i=1}^{C} M(l_i - f(\mathbf{A}, k_i)),\ l_i \in L,\ k_i \in K,$$

где C – число соответствующих друг другу цветов в пространствах прообраза K и образа L; M — метрика цветового различия; f — функция цветокорректирующего преобразования с параметрами $\mathbf{A}$. В общем случае чем больше число пар цветов C, чем полнее эти пары покрывают пространства K и L, тем более точно оцениваются параметры модели.

Существует множество моделей для отображения цветовых пространств, подавляющее большинство которых можно разделить на две группы. К первой группе относятся линейные регрессионные модели. Два основных варианта реализации модели первого порядка содержат девять или 12 коэффициентов [5], но также широко используются полиномиальные модели второго и более высоких порядков (линейные по параме-

трам), например, различные полиномиальные модели и так называемые рут-полиномиальные приведены в работе [6]. Вторая группа основана на нелинейных моделях. Она включает различные варианты: модель на основе сплайнов [7], радиально-базисных функций [8], гауссовых процессов [9], трехмерных таблиц поиска [10], также к данной группе можно отнести нейросетевые модели [11].

Простейшая модель линейного преобразования координат цвета из (X_K, Y_K, Z_K) в (X_L, Y_L, Z_L) (из одного цветового пространства в другое), в частности реализованная в известном в телевидении устройстве — матричном цветокорректоре (с добавлением условия неискаженной передачи белого цвета), использует матрицу параметров размера 3 × 3:

$$\begin{aligned}\begin{pmatrix} X_L & Y_L & Z_L \end{pmatrix} &\approx \begin{pmatrix} \hat{X}_L & \hat{Y}_L & \hat{Z}_L \end{pmatrix} = \\ = \begin{pmatrix} X_K & Y_K & Z_K \end{pmatrix} &\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \\ &= \begin{pmatrix} X_K & Y_K & Z_K \end{pmatrix} \mathbf{A}_{33}.\end{aligned}$$

При использовании более сложной модели матрица параметров может быть расширена, перестав при этом быть квадратной. Например, при расширении матрицы за счет добавления единичного вектора преобразование будет учитывать смещение яркости. В этом случае преобразование цветовых пространств выглядит следующим образом:

$$\begin{aligned}\begin{pmatrix} X_L & Y_L & Z_L \end{pmatrix} &\approx \begin{pmatrix} \hat{X}_L & \hat{Y}_L & \hat{Z}_L \end{pmatrix} = \\ = \begin{pmatrix} 1 & X_K & Y_K & Z_K \end{pmatrix} &\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{pmatrix} = \\ &= \begin{pmatrix} 1 & X_K & Y_K & Z_K \end{pmatrix} \mathbf{A}_{43}.\end{aligned}$$

Пусть

$$\mathbf{L} = \begin{pmatrix} X_L^1 & Y_L^1 & Z_L^1 \\ \vdots & \vdots & \vdots \\ X_L^N & Y_L^N & Z_L^N \end{pmatrix};$$

$$\left[1\mathbf{K}\right] = \begin{pmatrix} 1 & X_K^1 & Y_K^1 & Z_K^1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_K^N & Y_K^N & Z_K^N \end{pmatrix}.$$

Тогда при наличии достаточного числа цветов (N) можно найти решение, применяя метод наименьших квадратов (МНК) и псевдообращение Мура — Пенроуза [12]:

$$\mathbf{A}_{43} = \left(\left[1\mathbf{K}\right]^{\mathrm{T}}\left[1\mathbf{K}\right]\right)^{-1}\left[1\mathbf{K}\right]^{\mathrm{T}}\mathbf{L}.$$

При использовании сложных моделей (например, на основе полиномиальных функций второго и выше порядков) общая идея заключается в расширении линейного преобразования путем добавления членов высшего порядка (квадратов, кубов, перемножений цветовых координат и пр.) для достижения лучшей точности преобразования. Ниже приведены примеры полиномиальных моделей:

$$p_5 = \begin{bmatrix} X & Y & Z & 1 & XYZ \end{bmatrix};$$

$$p_6 = \begin{bmatrix} X & Y & Z & XY & YZ & XZ \end{bmatrix};$$

$$p_9 = \begin{bmatrix} X & Y & Z & XY & YZ & XZ & (X)^2 & (Y)^2 & (Z)^2 \end{bmatrix};$$

$$p_{11} = \begin{bmatrix} X & Y & Z & XY & YZ & XZ & (X)^2 & (Y)^2 & (Z)^2 & XYZ & 1 \end{bmatrix}.$$

Результат коррекции для некоторого цвета, например в случае использования p_6, вычисляется по формуле

$$\begin{pmatrix} X_B & Y_B & Z_B \end{pmatrix} = \begin{pmatrix} X & Y & Z & XY & YZ & XZ \end{pmatrix} \mathbf{A}_{63},$$

где $\mathbf{A}_{63}$ — матрица параметров полиномиальной модели цветокорректирующего преобразования.

Обычно предполагают, что ошибка преобразования цвета E_{color_i} рассчитывается как евклидова норма разностного вектора координат в пространстве образа. Для каждого цвета из имеющегося набора определяют норму разностного вектора между истинными координатами в целевом пространстве и получившимися координатами в результате преобразования. Преимущества такой оценки понятны. В этом случае определение параметров модели преобразования может быть легко реализовано аналитически, а в случае использования численной оптимизации евклидова норма удобна для градиентных методов.

Такой подход полностью оправдан, только если коррекция проводится для целей автоматического анализа изображений. Если предполагается, что пользователем результата цветокоррекции является человек (что соответствует действительности в большинстве случаев), то метрика M должна учитывать свойства человеческого зрения. Множество цветовых пространств, таких как XYZ, RGB, CMYK, ориентировано на работу аппаратуры, а не на учет нюансов восприятия цвета человеком. Это отражается в том, что одинаковое (в смысле евклидовой нормы) изменение значений координат цвета в различных

областях подобных пространств производит разное ощущение изменения цвета. На диаграмме цветности (являющейся проекцией XYZ) данное свойство хорошо иллюстрируют так называемые эллипсы МакАдама [13] (рис. 2). Они показывают области, в пределах которых цвета кажутся неразличимыми для человека. Видно, что размер этих областей и их элонгация существенно различаются в пределах видимого спектра.

Для оценки качества цветокоррекции, рассчитанной на человека, необходимо учитывать особенности восприятия им цвета. Для моделирования зрительной системы созданы так называемые перцептуальные пространства, самым известным из которых является L*a*b*, принятое Международной комиссией по освещению CIE [15]. В L*a*b*, как и в других перцептуальных пространствах, яркостная составляющая отделена от цветовой, что позволяет работать с цветом напрямую. При создании L*a*b* стояла задача сделать данное пространство униформным (равномерным), т. е. как можно более линейным с точки зрения человеческого восприятия изменения цвета. Несмотря на то, что в результате пространство L*a*b* получилось гораздо более равномерным, чем RGB или другое неперцептуальное пространство, данная цель не была в полной мере достигнута (например, отмечено нарушение равномерности в области желтых цветов).

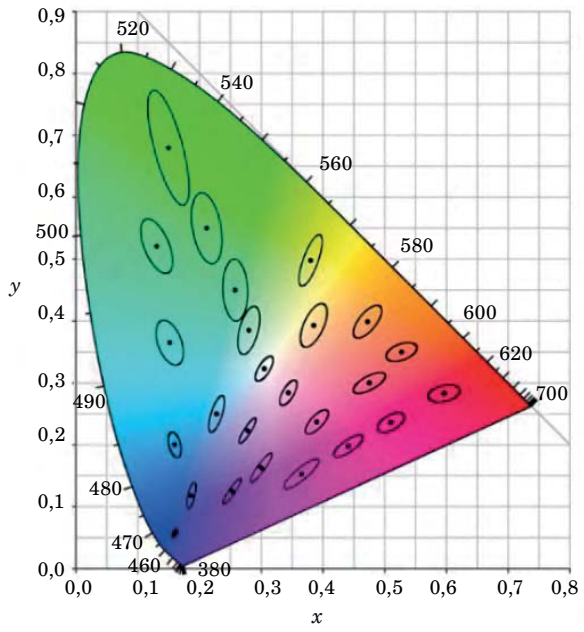
Исторически первое выражение для оценки цветового различия с учетом свойств человеческого зрения, разработанное CIE в 1976 г. [16], представляет собой евклидову норму в координатах L*a*b*:

$$\Delta E^*_{ab} = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2}.$$

В связи с неполной равномерностью L*a*b* работы по совершенствованию формулы для оценки цветовых различий были продолжены. В 1994 г. была представлена формула CIE94, в 2000 г. появилась формула CIE DE 2000 [17]. Вычисление значений с помощью этой формулы оказалось достаточно сложным и громоздким, но в то же время обеспечило вполне приемлемые результаты, и на сегодня CIE DE 2000 широко используется для оценки цветовых различий.

Адекватность оценки с помощью CIE DE 2000 особенностям зрительного восприятия была достигнута путем введения значительных нелинейных коррекций, в результате чего идентификация параметров моделей цветокорректирующих преобразований, построенных на метрике с использованием этой формулы, значительно затруднилась. Например, при использовании CIE DE 2000 оценка параметров не может быть получена аналитически с помощью МНК. Решения оптимизационных задач численными методами, выполняющими минимизацию целевой функции, использующей метрику на базе CIE DE 2000, сопряжены с трудностями, так как рельеф функции (обладающей разрывами [18] и являющейся мультимодальной) получается сложным для алгоритмов поиска минимума.

Нужно отметить, что для оценки перцептуальной разницы между цветами существует и альтернативный CIE DE 2000 подход. Наряду с совершенствованием формулы для оценки цветовых различий в колориметрии непрерывно ведется работа по созданию все более равномерных перцептуальных пространств. В результате появляются пространства (например, OkLab, опубликованный в 2020 г., https://bottosson.github.io/posts/oklab/), в которых, как сообщается, ощущаемое цветовое различие практически соответствует обычной евклидовой норме [19]. Но несмотря на появление новых перцептуально-равномерных пространств, на данный момент для расчета воспринимаемых цветовых различий в промышленности повсеместно используется именно CIE DE 2000. Поэтому целесообразно рассмотреть применение для перцептуальной цветокоррекции метрики на базе этой формулы.



■ ***Рис. 2.*** Эллипсы МакАдама (увеличенные в 10 раз) [14]

■ ***Fig. 2.*** McAdam ellipses (magnified 10 times) [14]

Метод перцептуальной цветокоррекции

Схема колориметрического конвейера расчета ошибки цветового преобразования на базе

формулы CIE DE 2000 представлена на рис. 3. Для наглядного описания метода в данной работе рассмотрен случай отображения из пространства RGB-камеры в XYZ (как один из самых важных в конвейере), однако предложенный принцип актуален и для других колориметрических процедур.

Ветвь наблюдателя включает формирование XYZ-координат цвета с помощью функций соответствия стандартного наблюдателя и известных спектральных характеристик цветов:

$$X = \frac{K}{N}\int_{\lambda} R(\lambda) I(\lambda) \bar{x}(\lambda) \mathrm{d}\lambda;$$

$$Y = \frac{K}{N}\int_{\lambda} R(\lambda) I(\lambda) \bar{y}(\lambda) \mathrm{d}\lambda; \qquad (1)$$

$$Z = \frac{K}{N}\int_{\lambda} R(\lambda) I(\lambda) \bar{z}(\lambda) \mathrm{d}\lambda;$$

$$N = \int_{\lambda} I(\lambda) \bar{y}(\lambda) \mathrm{d}\lambda,$$

где $R(\lambda)$ — спектральная характеристика отражения материала объекта; $I(\lambda)$ — спектральное распределение освещения, падающего на объект; $\bar{x}, \bar{y}, \bar{z}$ — функции цветового соответствия стандартного колориметрического наблюдателя; K — масштабирующий множитель (обычно равный 1 или 100), интеграл принято считать в интервале от 380 до 780 нм.

Спектральные характеристики отражения для цветов могут быть получены в общем случае двояко: либо с помощью использования диаграммы цветов (где они заданы для определенного набора), либо с помощью гиперспектральной камеры, формирующей в каждом пикселе изображения спектральный отклик с числом отсчетов, существенно превышающим стандартный триплет RGB (обычно используется около 200 отсчетов). Далее выполняют хроматическую адаптацию (учет изменения освещенности) [20] от источника белого света типа D65 (стандартный для XYZ) к источнику типа D50 (стандартный для L*a*b*), что необходимо для перехода из XYZ в L*a*b*:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\mathrm{D50}} = \mathbf{M}_{CA} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\mathrm{D65}}.$$

Существуют различные алгоритмы для расчета коэффициентов матрицы адаптации. Например, по Брэдфорду [21] переход от D65 к D50 осуществляют посредством матрицы

$$\mathbf{M}_{CA_{\mathrm{D65}\to\mathrm{D50}}} = \begin{bmatrix} 1{,}0478112 & 0{,}0228866 & -0{,}0501270 \\ 0{,}0295424 & 0{,}9904844 & -0{,}0170491 \\ -0{,}0092345 & 0{,}0150436 & 0{,}7521316 \end{bmatrix}.$$

Затем выполняют преобразование из XYZ в L*a*b*:

$$L = 116 f_y - 16;$$

$$a = 500(f_x - f_y);$$

$$b = 200(f_x - f_y);$$

где

$$f_x = \begin{cases} \sqrt[3]{x_r}, \text{ если } x_r > e \\ \dfrac{k x_r + 16}{116} \text{ иначе} \end{cases};$$

$$f_y = \begin{cases} \sqrt[3]{y_r}, \text{ если } y_r > e \\ \dfrac{k y_r + 16}{116} \text{ иначе} \end{cases};$$

■ ***Рис. 3.*** Оценка точности преобразования цветового пространства с учетом особенностей восприятия на базе CIE DE 2000

■ ***Fig. 3.*** The CIE DE 2000 — based color space conversion accuracy estimation

$$f_z = \begin{cases} \sqrt[3]{z_r}, \text{ если } z_r > e \\ \dfrac{kz_r + 16}{116} \text{ иначе} \end{cases},$$

$$x_r = \frac{X}{X_r},$$

$$y_r = \frac{Y}{Y_r},$$

$$z_r = \frac{Z}{Z_r},$$

$$e = 0{,}008856,\ k = 903{,}3,$$

(X_r, Y_r, Z_r) — значения для белого цвета D50: X_r = 96,4212, Y_r = 100, Z_r = 82,5188. Значения для точки белого даны для случая, когда X, Y, Z находятся в диапазоне [0..100]. Для диапазона [0..1] эти значения должны быть разделены на 100.

Ветвь камеры оперирует RGB-координатами (зависимыми от конкретного сенсора), получаемыми с помощью спектральных характеристик фотоприемника:

$$r = \frac{100}{N}\int_{\lambda} R(\lambda) I(\lambda) r(\lambda) \mathrm{d}\lambda;$$

$$g = \frac{100}{N}\int_{\lambda} R(\lambda) I(\lambda) g(\lambda) \mathrm{d}\lambda$$

$$b = \frac{100}{N}\int_{\lambda} R(\lambda) I(\lambda) b(\lambda) \mathrm{d}\lambda,$$

для которых выполняют нормализацию

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix}_{WB} = \mathbf{D}_{WB} \begin{bmatrix} r \\ g \\ b \end{bmatrix},$$

где $\mathbf{D}_{WB}$ — диагональная матрица, обеспечивающая баланс белого [1]:

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix}_{WB} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

значения r, g, b находятся в диапазоне [0..1].

На следующем шаге выполняется цветокорректирующее преобразование (из пространства RGB в XYZ), параметры которого необходимо оценить, и далее аналогично ветви наблюдателя осуществляется переход в L*a*b*. На основе полученных координат цветов в L*a*b* для ветвей наблюдателя и камеры выполняют расчет цветового различия по CIE DE 2000. В результате целевая функция для задачи минимизации формализуется в виде

$$\Delta E = \sum_i w_i \Delta E_{00},$$

где w — весовые коэффициенты (в общем случае все равные единице); ΔE_{00} — цветовое различие по формуле CIE DE 2000 для цветов из заданного набора.

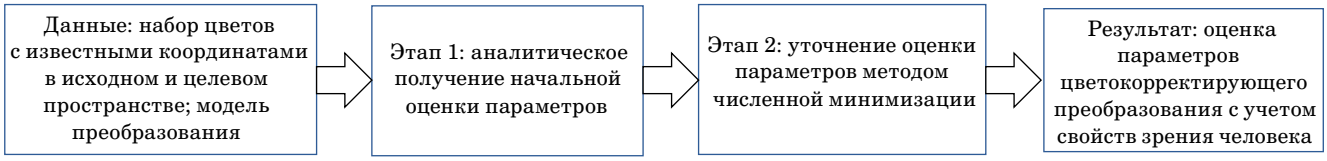
В задаче оптимизации важна норма, используемая для расчета целевой функции. Так, L1 соответствует оптимизации по средней ошибке, L2 — по среднеквадратичной и т. д. Чем выше норма, тем больше решение тяготеет к минимизации относительно больших ошибок, при этом повышается средняя. На практике в основном используют нормы L1 и L2 и крайне редко норму выше L4. Поэтому целевую функцию для задачи оптимизации можно переписать с учетом используемой нормы p:

$$\Delta E = \sum_i w_i \left\| \Delta E_{00} \right\|^p.$$

Оценку параметров цветокорректирующего преобразования с учетом перцептуальных свойств возможно реализовать путем численной минимизации целевой функции, основанной на метрике на базе CIE DE 2000 на множестве параметров преобразования **A**. Нахождение минимума целевой функции численным методом требует наличия стартовой точки — некоторых начальных значений матрицы **A**. С точки зрения точности и скорости решения задачи минимизации, чем ближе стартовая точка к положению глобального минимума, тем лучше. Более того, при значительном удалении начального приближения от глобального минимума есть риск получить неудовлетворительное решение (попадание алгоритма минимизации в локальный минимум).

Предлагаемый метод перцептуальной цветокоррекции рассматривает ее как двухэтапную процедуру (рис. 4).

На первом этапе параметры преобразования (в рассматриваемом частном случае — для перехода из RGB- в XYZ-пространство) рассчитываются с помощью евклидовой метрики МНК. Найденное решение, являясь аналитически полученным глобальным минимумом для евклидовой метрики, служит стартовой точкой для алгоритма численной минимизации целевой функции, основанной на перцептуальной метрике. В качестве алгоритма минимизации был выбран BFGS [22].



■ ***Рис. 4.*** Предлагаемый метод перцептуальной цветокоррекции
■ ***Fig. 4.*** The proposed method of perceptual color correction

Результаты экспериментального исследования

Для оценки эффективности предложенного метода было выполнено следующее исследование. Рассчитывались коэффициенты цветового преобразования для задачи отображения координат из пространства RGB-камеры (использованы характеристики Sony DXC 930) в пространство XYZ. При этом сравнивались результаты, полученные МНК (т. е. с применением евклидовой метрики, без последующего уточнения по перцептуальной) и предложенным методом. В качестве исходных данных спектров цветов были использованы следующие диаграммы: Macbeth (часто применяемый на практике набор, содержащий спектральные характеристики 24 цветов) и SG (содержащий большое число цветов — 140, достаточно равномерно покрывающих видимый диапазон, что обеспечивает валидность полученных результатов).

Моделирование проводилось для ряда источников освещения, в частности для белого света следующих типов: A, E, D50, D65, флуоресцентного (FL3.11) и диодного (LED-RGB100) источника. Рассчитывалась средняя ошибка для двух рассматриваемых методов и относительный выигрыш предлагаемого метода. Следует отметить, что для оценки воспринимаемой человеком точности цветового преобразования ориентироваться только на среднюю ошибку нельзя. Специфика восприятия состоит в том, что цветовые различия менее 1.0 по формуле CIE DE 2000 неразличимы для человека. Поэтому 1.0 по CIE DE 2000 называют «порогом восприятия». В то же время на практике широко применяют второй порог, равный 3.0 по CIE DE 2000, называемый «приемлемым порогом». Считается, что разница в цветах менее 3.0 в целом оценивается человеком как допустимая для практических задач. Следовательно, ошибка передачи какого-либо цвета более 3.0 по CIE DE 2000 значительно влияет на воспринимаемую точность цветопередачи. Поэтому важно при оценке точности цветового преобразования проводить анализ, для какого числа цветов были превышены соответствующие пороги, что и было сделано в рамках исследования методов.

■ ***Таблица 1.*** Результаты исследования для линейной модели первого порядка с девятью коэффициентами, L2, Macbeth (24 цвета)
■ ***Table 1.*** Research results for a first-order linear model with 9 coefficients, L2, Macbeth (24 colors)

Тип освещения	Средняя ошибка, CIE DE 2000		Уменьшение средней ошибки, %	Доля цветов с ошибкой цветопередачи, превышающей порог восприятия, %		Уменьшение доли цветов с ошибкой цветопередачи, превышающей порог восприятия, %	Доля цветов с ошибкой цветопередачи, превышающей приемлемый порог, %		Уменьшение доли цветов с ошибкой цветопередачи, превышающей приемлемый порог, %
	после МНК	после предложенного метода		после МНК	после предложенного метода		после МНК	после предложенного метода	
A	2,11	1,64	22	79	66	13	25	12	13
E	1,89	1,64	13	88	75	13	8	4	4
D50	1,96	1,66	15	88	67	21	8	13	–4
D65	1,78	1,59	10	79	58	21	8	4	4
FL3.11	0,89	0,79	10	38	29	8	4	0	4
LED-RGB100	0,51	0,49	5	13	13	0	0	0	0

Очевидно, что вид и сложность модели преобразования влияют на получаемые результаты. Исследование было проведено для двух линейных моделей первого порядка с девятью и 12 коэффициентами (табл. 1–4). При расчете целевой функции использовалась норма $p = 2$ (для обеспечения одинаковой нормы у двух методов).

Анализ таблиц показывает, что применение предложенного метода по сравнению с результатом цветокоррекции без учета перцептуальной метрики позволяет значительно снизить среднюю ошибку цветового преобразования, а также уменьшить долю цветов со значимой ошибкой с точки зрения восприятия.

При использовании матрицы преобразования размера 3 × 3 для диаграммы Macbeth средняя ошибка снижается в среднем на 12 %, от пяти до 22. При этом для естественных типов освещения (A, D50, D65) удается в среднем на 15 % снизить долю цветов, для которых ошибка цветового преобразования превышает порог восприятия. Доля цветов, превышающих приемлемый порог в целом, остается на том же уровне.

■ ***Таблица 2.*** Результаты исследования для линейной модели первого порядка с 12 коэффициентами, L2, Macbeth (24 цвета)

■ ***Table 2.*** Research results for a first-order linear model with 12 coefficients, L2, Macbeth (24 colors)

Тип освещения	Средняя ошибка, CIE DE 2000		Уменьшение средней ошибки, %	Доля цветов с ошибкой цветопередачи, превышающей порог восприятия, %		Уменьшение доли цветов с ошибкой цветопередачи, превышающей порог восприятия, %	Доля цветов с ошибкой цветопередачи, превышающей приемлемый порог, %		Уменьшение доли цветов с ошибкой цветопередачи, превышающей приемлемый порог, %
	после МНК	после предложенного метода		после МНК	после предложенного метода		после МНК	после предложенного метода	
A	2,55	1,64	35	79	67	12	25	13	12
E	2,18	1,59	27	75	67	8	17	4	13
D50	2,39	1,65	31	75	63	12	21	17	4
D65	2,20	1,56	28	71	63	8	17	4	13
FL3.11	1,23	0,79	35	50	29	21	4	0	4
LED-RGB100	0,61	0,49	18	21	13	8	0	0	0

■ ***Таблица 3.*** Результаты исследования для линейной модели первого порядка с девятью коэффициентами, L2, SG (140 цветов)

■ ***Table 3.*** Research results for a first-order linear model with 9 coefficients, L2, SG (140 colors)

Тип освещения	Средняя ошибка, CIE DE 2000		Уменьшение средней ошибки, %	Доля цветов с ошибкой цветопередачи, превышающей порог восприятия, %		Уменьшение доли цветов с ошибкой цветопередачи, превышающей порог восприятия, %	Доля цветов с ошибкой цветопередачи, превышающей приемлемый порог, %		Уменьшение доли цветов с ошибкой цветопередачи, превышающей приемлемый порог, %
	после МНК	после предложенного метода		после МНК	после предложенного метода		после МНК	после предложенного метода	
A	1,71	1,52	11	59	51	8	17	11	6
E	1,83	1,72	6	78	64	14	12	15	–3
D50	1,80	1,66	8	66	64	2	15	15	0
D65	1,74	1,64	6	62	61	1	14	14	0
FL3.11	0,83	0,80	3	29	26	3	2	1	1
LED-RGB100	0,48	0,46	3	19	14	5	0	0	0

■ ***Таблица 4.*** Результаты исследования для линейной модели первого порядка с 12 коэффициентами, L2, SG (140 цветов)
■ ***Table 4.*** Research results for a first-order linear model with 12 coefficients, L2, SG (140 colors)

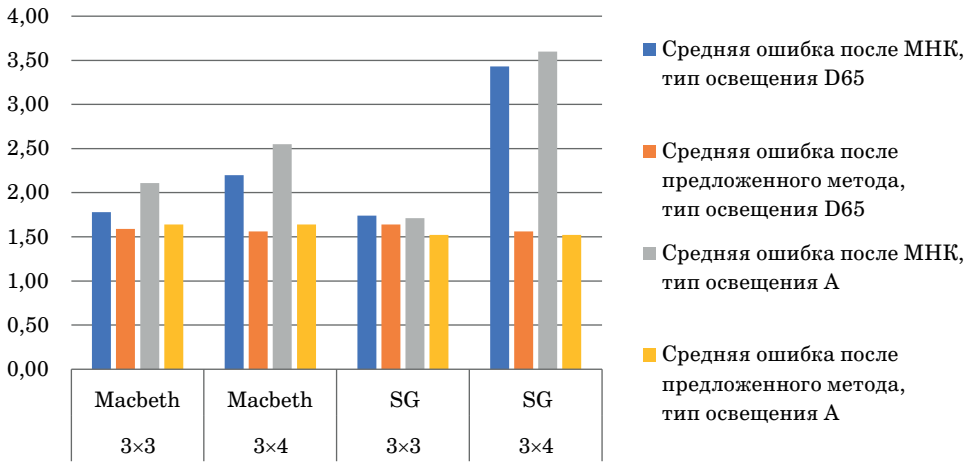
Тип освещения	Средняя ошибка, CIE DE 2000		Уменьшение средней ошибки, %	Доля цветов с ошибкой цветопередачи, превышающей порог восприятия, %		Уменьшение доли цветов с ошибкой цветопередачи, превышающей порог восприятия, %	Доля цветов с ошибкой цветопередачи, превышающей приемлемый порог, %		Уменьшение доли цветов с ошибкой цветопередачи, превышающей приемлемый порог, %
	после МНК	после предложенного метода		после МНК	после предложенного метода		после МНК	после предложенного метода	
A	3,60	1,52	58	76	51	25	29	12	17
E	3,35	1,61	52	76	63	13	28	15	13
D50	3,57	1,60	55	76	63	13	40	16	24
D65	3,43	1,56	55	74	61	13	29	14	15
FL3.11	1,95	0,78	60	54	27	27	16	1	15
LED-RGB100	0,84	0,46	45	29	14	15	0	0	0

При использовании матрицы преобразования размера 3 × 3 для диаграммы SG выигрыш по средней ошибке и доле цветов с существенной ошибкой оказывается относительно небольшим, около 5–6 %. Это связано с тем, что, с одной стороны, значительно большее число цветов позволяет точнее оценить параметры преобразования на первом шаге процедуры, а с другой стороны, преобразование размера 3 × 3 не обеспечивает достаточной гибкости для большого повышения точности цветопередачи за счет уточнения по перцептуальной метрике на втором шаге.

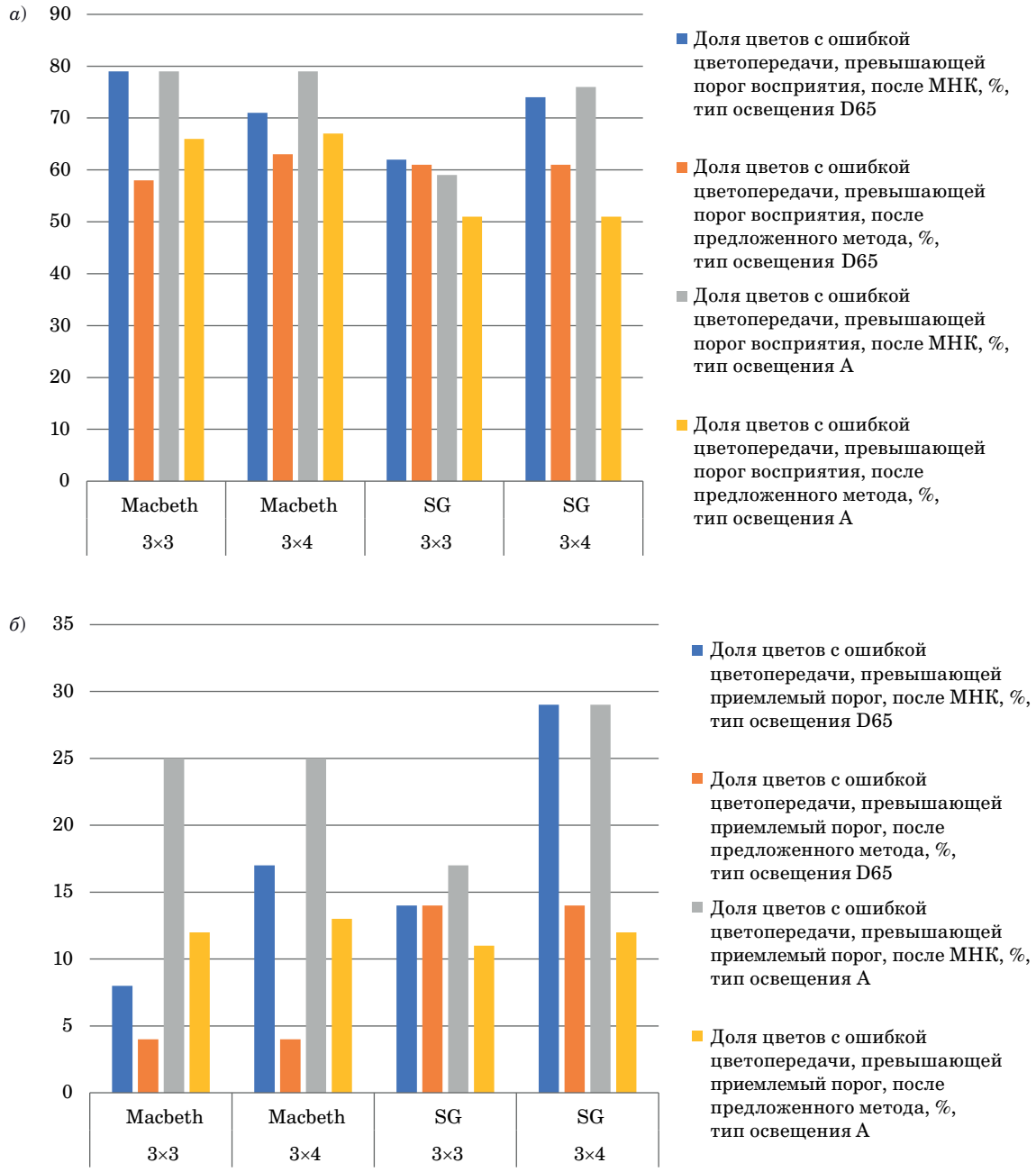
При использовании матрицы преобразования размера 3 × 4 для диаграммы Macbeth средняя ошибка снижается в среднем на 30 %, от 18 до 35. При этом в среднем на 10 % снижаются доли цветов, превышающих как порог восприятия, так и приемлемый порог.

При использовании матрицы преобразования 3 × 4 для диаграммы SG удается получить внушительный выигрыш в снижении средней ошибки — в среднем более чем на 50 %. При этом в среднем на 15–18 % снижаются доли цветов, превышающих как порог восприятия, так и приемлемый порог.

Диаграммы, обобщающие результаты для двух типов освещения D65 и A, представляющих наибольший практический интерес, приведены на рис. 5, 6, *а* и *б*.



■ ***Рис. 5.*** Средние ошибки цветового преобразования для типов освещений D65 и A
■ ***Fig. 5.*** Average color conversion errors for illumination types D65 and A



■ ***Рис. 6.*** Доля цветов, превышающих порог восприятия (*а*) и приемлемый порог (*б*), для типов освещений D65 и A

■ ***Fig. 6.*** Proportion of colors exceeding the perception threshold (*а*) and the acceptable threshold (*б*) for illuminant types D65 and A

Заключение

Полученные результаты показывают, что предложенный метод имеет значительную практическую ценность. Даже при использовании для оценки параметров достаточно простого линейного преобразования, содержащего девять коэффициентов, и небольшой диаграммы цветов Macbeth удается ощутимо повысить точность преобразования с точки зрения восприятия человеком. При использовании более сложного преобразования (хотя и далеко не самой высокой размерности из применяемых сегодня) выигрыш получается значительным как по снижению средней ошибки, так и по уменьшению доли цветов, превышающих порог восприятия и приемлемый порог.

Следует отметить и универсальность предложенного метода с точки зрения нормы, используемой при минимизации с учетом перцеп-

■ ***Таблица 5.*** Результаты исследования для линейной модели первого порядка с 12 коэффициентами, L1, Macbeth (24 цвета)

■ ***Table 5.*** Research results for a first-order linear model with 12 coefficients, L1, Macbeth (24 colors)

Тип освещения	Средняя ошибка, CIE DE 2000		Уменьшение средней ошибки, %	Доля цветов с ошибкой цветопередачи, превышающей порог восприятия, %		Уменьшение доли цветов с ошибкой цветопередачи, превышающей порог восприятия, %	Доля цветов с ошибкой цветопередачи, превышающей приемлемый порог, %		Уменьшение доли цветов с ошибкой цветопередачи, превышающей приемлемый порог, %
	после МНК	после предложенного метода		после МНК	после предложенного метода		после МНК	после предложенного метода	
A	2,55	1,58	38	79	54	25	25	21	4
E	2,18	1,48	32	75	50	25	17	13	4
D50	2,39	1,53	36	75	50	25	21	8	13
D65	2,20	1,46	34	71	42	29	17	13	4
FL3.11	1,23	0,74	40	50	21	29	4	4	0
LED-RGB100	0,61	0,47	22	21	17	4	0	0	0

туальной метрики. В табл. 5 приведены данные результата исследования предложенного метода с использованием L1 нормы при минимизации.

Видно (в сравнении с табл. 2), что выигрыш в уменьшении средней ошибки преобразования дополнительно увеличивается в среднем на 5 %. При этом удается улучшить результат преобразования с точки зрения числа цветов, ошибка на которых не превышает порог восприятия, но доля цветов, ошибка преобразования для которых превышает приемлемый порог, оказывается большей, чем при использовании L2 нормы. Это соответствует ожиданиям, так как L2 норма способствует уменьшению максимальных ошибок, а L1, наоборот, средней.

В заключение можно сделать следующие выводы.

1. Полученные результаты экспериментального исследования позволяют говорить о том, что предложенный метод перцептуальной цветокоррекции обеспечивает значительное повышение точности цветокорректирующего преобразования. Например, при использовании линейной модели первого порядка, содержащей 12 коэффициентов, средняя ошибка снижается на 30 % для диаграммы Macbeth и на 50 % для диаграммы SG. При этом в среднем на 10 % для диаграммы Macbeth и на 15–18 % для диаграммы SG снижаются доли цветов, превышающих как порог восприятия, так и приемлемый порог.

2. Метод не ограничен выбором нормы при построении целевой функции и позволяет использовать нормы Lp для любого p. Это дает возможность гибкой настройки для конкретной задачи.

3. Метод может быть успешно применен в широком спектре колориметрических задач (как в камерной, так и в дисплейной части конвейера обработки сигнала), в которых имеет смысл учитывать свойства зрения: характеризация сенсоров, преобразования цветовых охватов и пр.

Литература

1. **Yeejin L., Keigo H.** Lossless white balance for improved lossless CFA image and video compression. *IEEE Transactions on Image Processing*, 2022, no. 31, pp. 3309–3321, doi:10.1109/TIP.2022.3169687
2. **Hemrit Gh., Finlayson G. D., Gijsenij A., Gehler P., Bianco S., Funt B., Drew M., Shi L.** Rehabilitating the ColorChecker dataset for illuminant estimation. *Proc. IS&T 26th Color and Imaging Conf.*, 2018, pp.350–353.doi:10.2352/ISSN.2169-2629.2018.26.350
3. **Pedro D. Marrero Fernandez, Fidel A. Guerrero-Pena, Tsang Ing Ren, Jorge J. G. Leandro.** Fast and robust multiple ColorChecker detection using deep convolutional neural networks. *Image and Vision Computing*, 2019, no. 81, pp. 15–24. doi:10.1016/j.imavis.2018.11.001
4. *GretagMacbeth ColorChecker DC*. https://chromaxion.com/information/colorchecker_dc.html (дата обращения: 21.01.2023).
5. **Обухова Н. А., Мотыко А. А., Поздеев А. А.** Цифровая обработка эндоскопических изображений для систем поддержки врачебных решений. *Известия вузов России. Радиоэлектроника*, 2018, № 6, с. 54–65. doi:10.32603/1993-8985-2018-21-6-54-65

6. **Zhao Y., Ferguson S., Zhou H., Elliott C., Rafferty K.** Color alignment for relative color constancy via non-standard references. *IEEE Transactions on Image Processing*, 2022, no. 31, pp. 6591–6604. doi:10.1109/TIP.2022.3214107
7. **Bellavia F., Colombo C**. Dissecting and reassembling color correction algorithms for image stitching. *IEEE Transactions on Image Processing*, 2018, no. 2, pp. 735–748. doi:10.1109/TIP.2017.2757262
8. **Wu P.-Ch., Lin C.** A green and practical color photograph printing technology based on color difference model and human perception. *IEEE Access*, 2022, no. 10, pp. 649–666. doi:10.1109/ACCESS.2021.3138270
9. **Ye S., Lu Sh.-P., Munteanu A.** Color correction for large-baseline multiview video. *Signal Processing: Image Communication*, 2017, no. 43, pp. 40–50. doi:10.1016/j.image.2017.01.004
10. **Zeng H., Cai J., Li L., Cao Z., Zhang L.** Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, no. 4, pp. 2058–2073. doi:10.1109/TPAMI.2020.3026740
11. **Fei W., Wei W., Dan W., Guowang G.** Color constancy via multi-scale region-weighed network guided by semantics. *Frontiers in Neurorobotics*, 2022, no. 16, pp. 1–15. doi: 10.3389/fnbot.2022.841426
12. **Alebrahim R., Thamburaja P., Srinivasa A., Reddy J. N.** A robust Moore – Penrose pseudoinverse-based static finite-element solver for simulating non-local fracture in solids. *Computer Methods in Applied Mechanics and Engineering*, 2023, no. 403, pp. 1–26. doi: 10.1016/j.cma.2022.115727
13. **MacAdam D. L.** Visual sensitivities to color differences in daylight. *Journal of Optical Socient of America*, 1942, no. 49, pp. 247–274.
14. **Chevallier E., Farup I.** Interpolation of the MacAdam ellipses. *SIAM Journal on Imaging Sciences*, 2018, no. 3, pp. 1979–2000. doi:10.1137/17M1147184
15. **Durmus Alp.** CIELAB color space boundaries under theoretical spectra and 99 test color samples. *Color Research & Application*, 2020, no. 5, pp. 796–802. doi:10.1002/col.22521
16. **Yi-Tun Lin, Graham D. Finlayson.** A rehabilitation of pixel-based spectral reconstruction from RGB images. *Sensors*, 2023, no. 8, pp. 1–20. doi:10.3390/s23084155
17. **Mirjalili F., Luo M. R., Cui G., Morovic J.** Color-difference formula for evaluating color pairs with no separation: ΔENS. *Journal of the Optical Society of America A*, 2019, no. 5, pp. 789–799. doi:10.1364/JOSAA.36.000789
18. **Hu W., Davis W.** The effect of control resolution on the usability of color-tunable lighting systems. *LEUKOS – Journal of Illuminating Engineering Society of North America*, 2018, no. 15, pp. 1–14. doi:10.1080/15502724.2018.1477053
19. **Hellwig L., Stolitzka D.** An advanced color model for evaluating new display technologies. *Information Display*, 2023, no. 3, pp. 11–15. doi:10.1002/msid.1392
20. **Scott A. Burns.** Chromatic adaptation transform by spectral reconstruction. *Color Research & Application*, 2019, no. 5, pp. 682–693. doi:10.1002/col.22384
21. **Akazawa T., Kinoshita Y., Kiya H.** Multi-color balance for color constancy. *Proc. IEEE Intern. Conf. on Image Processing (ICIP)*, 2021, pp. 1369–1373. doi:10.1109/ICIP42928.2021.9506528
22. **Shashi Kant Mishra, Geetanjali Panda, Suvra Kanti Chakraborty, Mohammad Esmael Samei, Bhagwat Ram**. On q-BFGS algorithm for unconstrained optimization problems. *Advances in Difference Equations*, 2020, no. 638, pp. 1–24. doi:10.1186/s13662-020-03100-2

UDC 654.1
doi:10.31799/1684-8853-2023-3-26-38
EDN: ZZTURK

The method for perceptual color correction

A. A. Motyko[a], PhD, Tech., Associate Professor, orcid.org/0000-0003-4241-4298
N. A. Obukhova[a], Dr. Sc., Tech., Professor, orcid.org/0000-0003-1953-2085, natalia172419@yandex.ru
[a]Saint-Petersburg Electrotechnical University «LETI», 5, Prof. Popov St., 197376, Saint-Petersburg, Russian Federation

Introduction: Color correction transformations are widely used in the graphics processing pipeline, for example in sensor characterization, white balance, artistic processing, gamut mappings of displays, and so on. However, the current methods of coordinate transformation between different color spaces do not sufficiently take into account the human visual properties, which leads to color reproduction errors. **Purpose:** To develop a method for the estimation of color correction parameters taking into account human visual perception properties. **Results:** The proposed method consists of two stages. At the first stage the estimation of color correction transformation parameters is carried out by the least squares method. At the second stage the obtained estimation is specified by means of the numerical minimization of the target function constructed on the perceptual criterion (based on the CIE DE 2000 formula for calculation of color differences). Experimental study has shown that this method allows to significantly increase the accuracy of color reproduction. The average error is reduced by 10 to 50% depending on the color correction model used. In addition, the number of colors with significant for human perception color reproduction errors is reduced by 10 to 15%. The method can use different models of color correction transformations, which makes it universal and gives wide possibilities of implementation in the pipeline of graphic information

processing. **Practical relevance:** The developed method for the estimation of color correction parameters taking into account human visual perception properties can be applied for the synthesis of TV systems in tasks, where accurate color rendering is important for the user, for example, in digital photography, telemedicine, etc.

Keywords — color correction, CIE DE 2000, color correction transformations, perceptual color correction, color difference formula.

For citation: Motyko A. A., Obukhova N. A. The method for perceptual color correction. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 26–38 (In Russian). doi:10.31799/1684-8853-2023-3-26-38, EDN: ZZTURK

References

1. Yeejin L., Keigo H. Lossless white balance for improved lossless CFA image and video compression. *IEEE Transactions on Image Processing*, 2022, no. 31, pp. 3309–3321, doi:10.1109/TIP.2022.3169687
2. Hemrit Gh., Finlayson G. D., Gijsenij A., Gehler P., Bianco S., Funt B., Drew M., Shi L. Rehabilitating the ColorChecker dataset for illuminant estimation. *Proc. IS&T 26th Color and Imaging Conf.*, 2018, pp. 350–353. doi:10.2352/ISSN.2169-2629.2018.26.350
3. Pedro D. Marrero Fernandez, Fidel A. Guerrero-Pena, Tsang Ing Ren, Jorge J. G. Leandro. Fast and robust multiple ColorChecker detection using deep convolutional neural networks. *Image and Vision Computing*, 2019, no. 81, pp. 15–24. doi:10.1016/j.imavis.2018.11.001
4. *GretagMacbeth ColorChecker DC*. Available at: https://chromaxion.com/information/colorchecker_dc.html (accessed 21 January 2023).
5. Obukhova N. A., Motyko A. A., Pozdeev A. A. Endoscopic images digital processing for clinical decision support systems. *Journal of the Russian Universities: Radioelectronics*, 2018, no. 6, pp. 54–65 (In Russian). doi:10.32603/1993-8985-2018-21-6-54-65
6. Zhao Y., Ferguson S., Zhou H., Elliott C., Rafferty K. Color alignment for relative color constancy via non-standard references. *IEEE Transactions on Image Processing*, 2022, no. 31, pp. 6591–6604. doi:10.1109/TIP.2022.3214107
7. Bellavia F., Colombo C. Dissecting and reassembling color correction algorithms for image stitching. *IEEE Transactions on Image Processing*, 2018, no. 2, pp. 735–748. doi:10.1109/TIP.2017.2757262
8. Wu P.-Ch., Lin C. A green and practical color photograph printing technology based on color difference model and human perception. *IEEE Access*, 2022, no. 10, pp. 649–666. doi:10.1109/ACCESS.2021.3138270
9. Ye S., Lu Sh.-P., Munteanu A. Color correction for large-baseline multiview video. *Signal Processing: Image Communication*, 2017, no. 43, pp. 40–50. doi:10.1016/j.image.2017.01.004
10. Zeng H., Cai J., Li L., Cao Z., Zhang L. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, no. 4, pp. 2058–2073. doi:10.1109/TPAMI.2020.3026740
11. Fei W., Wei W., Dan W., Guowang G. Color constancy via multi-scale region-weighed network guided by semantics. *Frontiers in Neurorobotics*, 2022, no. 16, pp. 1–15. doi: 10.3389/fnbot.2022.841426
12. Alebrahim R., Thamburaja P., Srinivasa A., Reddy J. N. A robust Moore — Penrose pseudoinverse-based static finite-element solver for simulating non-local fracture in solids. *Computer Methods in Applied Mechanics and Engineering*, 2023, no. 403, pp. 1–26. doi: 10.1016/j.cma.2022.115727
13. Mac Adam D. L. Visual sensitivities to color differences in daylight. *Journal of Optical Socient of America*, 1942, no. 49, pp. 247–274.
14. Chevallier E., Farup I. Interpolation of the MacAdam ellipses. *SIAM Journal on Imaging Sciences*, 2018, no. 3, pp. 1979–2000. doi:10.1137/17M1147184
15. Durmus Alp. CIELAB color space boundaries under theoretical spectra and 99 test color samples. *Color Research & Application*, 2020, no. 5, pp. 796–802. doi:10.1002/col.22521
16. Yi-Tun Lin, Graham D. Finlayson. A rehabilitation of pixel-based spectral reconstruction from RGB images. *Sensors*, 2023, no. 8, pp. 1–20. doi:10.3390/s23084155
17. Mirjalili F., Luo M. R., Cui G., Morovic J. Color-difference formula for evaluating color pairs with no separation: ΔENS. *Journal of the Optical Society of America A*, 2019, no. 5, pp. 789–799. doi:10.1364/JOSAA.36.000789
18. Hu W., Davis W. The effect of control resolution on the usability of color-tunable lighting systems. *LEUKOS — Journal of Illuminating Engineering Society of North America*, 2018, no. 15, pp. 1–14. doi:10.1080/15502724.2018.1477053
19. Hellwig L., Stolitzka D. An advanced color model for evaluating new display technologies. *Information Display*, 2023, no. 3, pp. 11–15. doi:10.1002/msid.1392
20. Scott A. Burns. Chromatic adaptation transform by spectral reconstruction. *Color Research & Application*, 2019, no. 5, pp. 682–693. doi:10.1002/col.22384
21. Akazawa T., Kinoshita Y., Kiya H. Multi-color balance for color constancy. *Proc. IEEE Intern. Conf. on Image Processing (ICIP)*, 2021, pp. 1369–1373. doi:10.1109/ICIP42928.2021.9506528
22. Shashi Kant Mishra, Geetanjali Panda, Suvra Kanti Chakraborty, Mohammad Esmael Samei, Bhagwat Ram. On q-BFGS algorithm for unconstrained optimization problems. *Advances in Difference Equations*, 2020, no. 638, pp. 1–24. doi:10.1186/s13662-020-03100-2

УДК 004.021
doi:10.31799/1684-8853-2023-3-39-50
EDN: AUBKNL

Научные статьи
Articles

Моделирование динамики информационных потоков в маршрутах вычислительных сетей

А. А. Карандашев[a]***,*** *аспирант, orcid.org/0000-0002-0358-0165*
В. Л. Оленев[a]***,*** *канд. техн. наук, доцент, orcid.org/0000-0002-1817-2754, Valentin.Olenev@guap.ru*
Г. С. Бритов[a]***,*** *канд. техн. наук, доцент, orcid.org/0000-0002-0452-523X*
[a]*Санкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ*

Введение: *при разработке современных узкоспециализированных программных систем по проектированию сетей с червячной коммутацией необходимо составлять таблицы маршрутизации. Для этого требуется получить эффективные конфигурации беступиковых маршрутов, прокладываемых на начальных стадиях проектирования сети.* ***Цель:*** *провести моделирование динамики инерционных процессов накопления данных в узлах-приемниках вычислительных сетей при передаче пакетов из узлов-передатчиков.* ***Результаты:*** *получены три примера схем сетей, которые позволяют рассмотреть классические сетевые структуры с разной степенью вычислительной нагрузки. Для каждой сети построены матрицы, описывающие ее структуру. Показаны результаты моделирования, проведенного при помощи специально разработанных сценариев в математическом программном обеспечении. Представлены результирующие графики весовой и переходной характеристик, а также целый ряд характеристик каждого исследуемого маршрута. Проведенное моделирование позволило убедиться в отсутствии информационного коллапса в беступиковых маршрутах. Отдельное исследование двух одновременно функционирующих путей передачи данных в сети с червячной коммутацией показывает устойчивый характер динамических процессов, что говорит о корректности предложенных моделей беступиковых червячных маршрутов.* ***Практическая значимость:*** *результаты исследования могут быть использованы при проектировании вычислительных сетей в специализированном программном обеспечении. Анализ динамических характеристик призван помочь в принятии решений о выборе беступиковых маршрутов, предлагаемых разработчиком сети с червячной коммутацией.*

Ключевые слова *— динамика передачи данных, информационные потоки, беступиковые маршруты, червячная коммутация, весовая характеристика, переходная характеристика, процесс накопления данных, модели маршрутов.*

Для цитирования: Карандашев А. А., Оленев В. Л., Бритов Г. С. Моделирование динамики информационных потоков в маршрутах вычислительных сетей. *Информационно-управляющие системы*, 2023, № 3, с. 39–50. doi:10.31799/1684-8853-2023-3-39-50, EDN: AUBKNL

For citation: Karandashev A. A., Olenev V. L., Britov G. S. Modeling the dynamics of information flows in the routes of computer networks. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 39–50 (In Russian). doi:10.31799/1684-8853-2023-3-39-50, EDN: AUBKNL

Введение

Моделирование динамики маршрутов вычислительных сетей (ВС) представляет собой актуальную задачу при построении ВС и изучении их динамических характеристик. Так, в работе [1] рассмотрена проблема динамики маршрутов для современных сетевых структур. Проблемы или исследования, повлекшие за собой высокие темпы развития сетей, а также применение новых технических решений в этой сфере, описываются в работах [2–5]. Однако добавление новой аппаратуры и увеличение количества каналов передачи данных не всегда улучшает работу сети, а иногда даже сказывается негативно [6]. Кроме того, изменение классических сетей и замена оборудования дорого обходятся при использовании классических протоколов [7–9]. Поэтому разработка узкоспециализированных протоколов может помочь в решении проблем, которые не решить простым добавлением каналов передачи или добавлением новых сетевых устройств [10]. При проектировании подобных систем с применением не только новых устройств, но и стандартов, важно проводить моделирование и оценивать характеристики получаемых сетей, чтобы удешевить и ускорить процесс разработки, не потеряв при этом в надежности получаемого продукта. Важно уметь правильно выделить основные характеристики разрабатываемых сетей и корректно оценить их. Так, в работах [11–13] исследованы проблемы моделирования, выделения параметров оценки и методов мониторинга ВС. Вместе с этим в работах [14–16] проведены параллели между социальными и вычислительными сетями и выделены метрики для оценки динамики и характеристик информационных потоков в социальных сетях. Согласно исследованиям [17, 18] задача определения и оценивания характеристик также актуальна для сферы информационной безопасности, для выявления атак в реальном времени и безопасности энергосистем.

Тем не менее множество узкоспециализированных систем по-прежнему остаются недостаточно изученными и не имеют четко выделенных характеристик для изучения соответствующих динамических процессов. Примером могут быть ВС с червячной коммутацией, для которых рассматриваемая задача до сих пор не имеет решения даже с использованием нейронных сетей, актуальность применения которых при разработке современных сетевых структур рассмотрена в работах [19–22].

Целью статьи является представление полученных динамических характеристик инерционных процессов, описывающих накопление данных информационных потоков в узлах-приемниках ВС при передаче пакетов из узлов-передатчиков. В связи с поставленной целью необходимо решить следующие задачи:

1) выделить отличительные особенности сетей с червячной коммутацией и беступиковой маршрутизацией;

2) разработать динамические модели узлов и беступиковых маршрутов в ВС с червячной коммутацией;

3) представить набор динамических характеристик процессов накопления данных от узла-источника к узлу-приемнику сети;

4) провести моделирование динамики информационных потоков.

В представленном исследовании используются формализмы теории сетей и теории систем управления.

Сети с червячной коммутацией и беступиковая маршрутизация

В сетях с червячной коммутацией используется особый способ маршрутизации пакетов — «на лету». В таких сетях данные не буферизируются в коммутаторе, а сразу же передаются на выходной порт. Если же он занят, то блокируется входной порт пакета. Он будет заблокирован, пока требуемый выходной порт не освободится. Таким образом, при коммутации «на лету» нет необходимости в буферной памяти в маршрутизаторах, за счет чего увеличивается скорость передачи данных в сети сравнительно с маршрутизацией, опирающейся на коммутацию пакетов с промежуточной буферизацией. При использовании промежуточной буферизации весь пакет сначала принимается в буфер маршрутизатора, затем определяется выходной канал, и только после этого пакет передается дальше. Если же необходимый выходной порт занят, то прием по входному порту приостанавливается.

Червячная коммутация позволяет увеличить производительность сети и понизить стоимость оборудования, так как при отсутствии промежуточной буферизации не нужны дорогие буферы данных в каждом коммутирующем устройстве.

Однако беступиковая маршрутизация накладывает ограничения на возможные маршруты в сети передачи данных. Под «тупиком» понимается ситуация, когда заголовок пакета блокируется из-за того, что следующий канал занят передачей этого же пакета. Также «тупиком» считается состояние, когда несколько пакетов взаимно блокируют друг друга. В таком случае каждый пакет «упирается» в занятый другим пакетом виртуальный канал.

Отсутствие циклических зависимостей является необходимым и достаточным условием детерминированной беступиковой маршрутизации [23]. В сетях с червячной коммутацией особенно важно избегать блокировок типа «тупик» при передаче данных, поскольку пакеты не буферизируются и занимают весь канал на пути своего следования по размеру передаваемого пакета. Такие сети часто используются там, где необходима быстрая и надежная передача данных. Это дорогостоящие системы, где критически важно, чтобы данные не застревали в сети при передаче, приводя к сбоям при работе всей системы.

Современные методы анализа динамики вычислительных сетей

Современные методы анализа динамики не ориентированы на маршруты для ВС с червячной коммутацией [24]. Последние исследования в рамках текущей тематики сосредоточены на изучении динамики блокировки ВС вирусами [25], разработке марковской модели поведения сетевых объектов вычислительных систем [26], разработке алгоритмической основы BeCAUSe для анализа сети [27].

В 2021 г. был предложен новый метод улучшения динамики для телекоммуникационных сетей [28]. Задача исследования динамики информационных потоков формулировалась на основе системного подхода, заключающегося в том, что в качестве системного выбирался многомерный показатель, учитывавший не только связи элементов, но и степень использования этих связей при функционировании сети. Это позволяло распределять информационные потоки путем изменения элементов матрицы маршрутизации в модели открытой сети массового обслуживания в целях оптимизации системы.

В работе [28] представлены не только методика, но и законченный алгоритм улучшения динамики, однако не дана оценка характеристик в явном виде и метод не адаптирован для сети с червячной коммутацией.

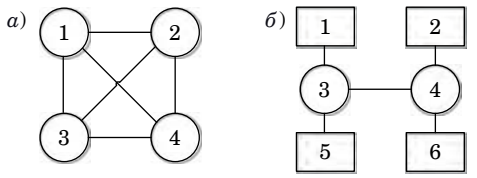
Последние исследования показывают актуальность изучения динамики информационных потоков для сетей с червячной коммутацией, так как для этого класса сетей методов изучения динамических характеристик на данный момент не существует. Для достижения заявленной в статье цели необходимо решить задачу по составлению динамических моделей узлов и беступиковых маршрутов в ВС с червячной коммутацией.

Вычислительная сеть как объект динамического моделирования информационных потоков

Вычислительная сеть представляет собой совокупность устройств, соединенных друг с другом с помощью каналов связи. Математической моделью ВС служит неориентированный граф, вершинами которого являются узлы сети, а ребрами — каналы связи. Примеры подобного описания ВС представлены на рис. 1, *а* и *б*. Каналы передачи данных в самой сети могут быть дуплексными. Дуплексные каналы позволяют с помощью прямой и обратной передачи данных организовывать в коммутаторах разнообразные маршруты информационных потоков, в том числе замкнутые. В замкнутых маршрутах из-за образования положительной обратной связи возможен информационный коллапс, когда в узле-приемнике происходит бесконечное накопление данных. Направление, в котором каждый дуплексный канал будет использоваться для каждого конкретного маршрута, зависит от заданных информационных потоков и схемы ВС.

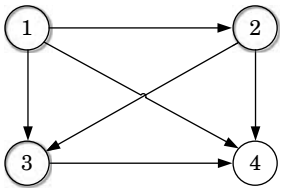
Узлы на рис. 1, *а* представляют устройства, совмещающие в себе функции вычислителя и коммутатора. В то же время могут быть сети, где узлы исполняют роль либо вычислителя, либо коммутатора (см. рис. 1, *б*).

В приведенной на рис. 2 схеме ВС с четырьмя узлами источниками пакетов с данными являются узлы № *1*, *2*, *3*, а приемниками являются узлы № *2*, *3*, *4*. Причем каждый узел представляет собой вычислитель-коммутатор. Узлы № *2* и *3*,



■ ***Рис. 1.*** ВС с дуплексными каналами и двумя видами узлов: *а* — вычислитель и коммутатор вместе; *б* — вычислитель и коммутатор отдельно

■ ***Fig. 1.*** Networks with duplex channels and two types of nodes: *а* — calculator and switch together; *б* — calculator and switch separately



■ ***Рис. 2.*** ВС с четырьмя узлами

■ ***Fig. 2.*** Network with four nodes

получая данные от соседа, выполняют свою коммутационную задачу и передают данные дальше. В таком случае в сети возможны различные червячные маршруты передачи пакетов с данными от узлов-источников к узлам-приемникам. Их можно разделить как по длине, учитывая количество проходимых каналов связи, так и по типу узла-источника:

— короткие беступиковые маршруты с одним каналом связи между вершинами графа сети: $1 \to 2$; $1 \to 3$; $1 \to 4$;

— длинные беступиковые маршруты с двумя и более каналами связи в маршруте между вершинами графа сети: $1 \to 2 \to 4$; $1 \to 3 \to 4$; $1 \to 2 \to 3 \to 4$;

— маршруты, источником информационного потока которых является узел № *2*: $2 \to 3$; $2 \to 4$; $2 \to 3 \to 4$;

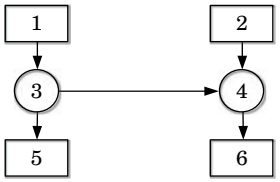
— маршрут, источником информационного потока которого является узел № *3*: $3 \to 4$.

Если ВС строится из узлов различных типов (рис. 3), то в такой сети узлы, показанные прямоугольниками, соответствуют вычислителям, а узлы, показанные кругами, соответствуют коммутаторам. Здесь возможны различные червячные маршруты:

— маршруты, источником информационного потока в которых является узел № *1*: $1 \to 3 \to 5$; $1 \to 3 \to 4 \to 6$;

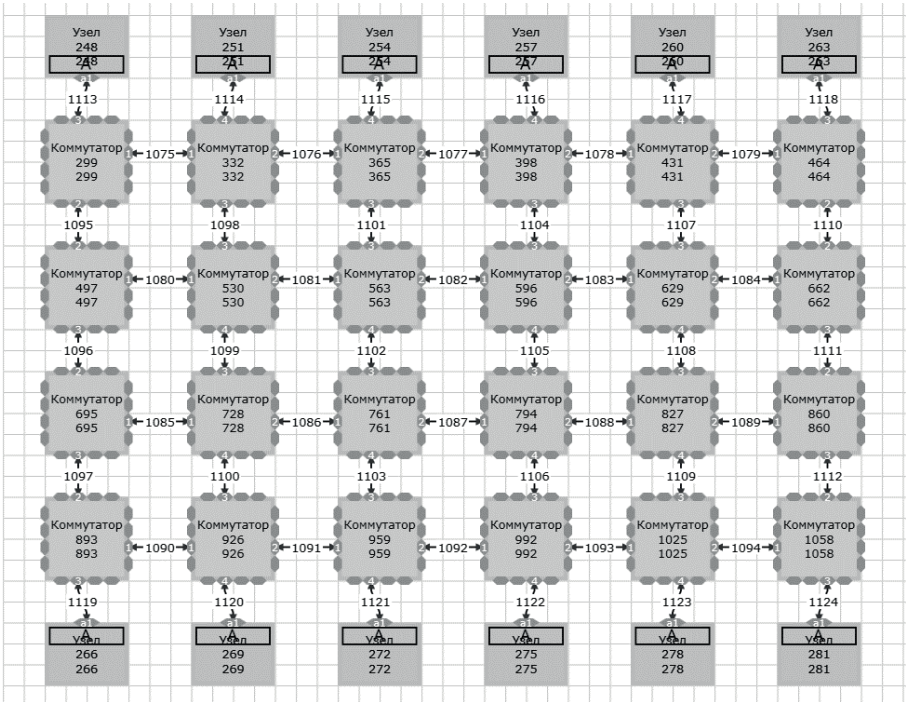
— маршрут, источником информационного потока в котором является узел № *2*: $2 \to 4 \to 6$.

Другим примером ВС может служить схема, представленная на рис. 4. Такая сеть имеет топологию «решетка 4 × 6». В сети 12 узлов, которыми являются вычислители, и 24 узла, которыми являются коммутаторы. Причем каждое вычислительное устройство передает данные



■ ***Рис. 3.*** ВС с двумя видами узлов

■ ***Fig. 3.*** Network with two types of nodes



■ ***Рис. 4.*** ВС с топологией «решетка 4 × 6»
■ ***Fig. 4.*** Network with “4 x 6 grid” topology

каждому другому вычислительному устройству в сети. Благодаря большому количеству коммутаторов и сложной сетевой структуре возможно большое количество различных беступиковых маршрутов между каждой парой узлов, что позволяет рассмотреть новый подход к оценке динамических характеристик беступиковых маршрутов сетей с червячной коммутацией в условиях большой вычислительной нагрузки [29].

Червячным маршрутом, проходящим по сети рис. 4, служит маршрут с узлом-источником № 248 и узлом-приемником № 251. Самый короткий маршрут из возможных беступиковых маршрутов между заданными вычислителями имеет следующий вид: узел 248 → коммутатор 299 → коммутатор 332 → узел 251.

Для указанных маршрутов выполним моделирование процесса накопления данных в конечных узлах. Полученные результаты моделирования используются для анализа беступиковости маршрутов и отсутствия информационного коллапса, возникающего в замкнутых маршрутах.

Динамические модели узлов вычислительной сети

Динамическими моделями узлов ВС, которыми являются вычислители и коммутаторы, могут служить апериодические звенья 1-го порядка. Каналы в сети считаются идеальными.

Известно, что в передаточной функции апериодического звена 1-го порядка имеется только два параметра: $K_{зв}$ — коэффициент передачи звена, $T_{зв}$ — постоянная времени звена. Параметр $K_{зв}$ показывает, как осуществляется обработка данных в вычислителе. Если он равен 1, то объем входных данных сохраняется. Этот объем может уменьшиться или увеличиться. Параметр $T_{зв}$ [мкс] показывает время, которое вычислительное устройство тратит на прием, обработку и передачу данных.

Время переходного процесса принято измерять следующим образом: $T_{п.п} = 3T_{зв}$. Пока не закончится переходный процесс, нельзя считать, что звено готово принимать новый пакет данных. Длина обрабатываемого пакета должна быть не меньше времени переходного процесса, иначе произойдет потеря данных.

Обычно выделяют две временные характеристики звена — весовую и переходную. Каждая из них представляет собой затухающие экспоненты с коэффициентом затухания $K_{зт} = 1/T_{зв}$.

Оценка постоянной времени $T_{зв}$ связана с допуском на получение результата обработки данных Δ и следует из вида весовой функции $h(t)$ в момент времени переходного процесса:

$$h_{зв}(T_{п.п}) = \frac{K_{зв}}{T_{зв}} \cdot e^{-\frac{3T_{зв}}{T_{зв}}} = \frac{K_{зв}}{T_{зв}} \cdot 0{,}0498 = \Delta.$$

Следовательно, постоянная времени вычисляется следующим образом:

$$T_{зв} = \frac{0{,}0498 \cdot K_{зв}}{\Delta}.$$

Постоянная времени $T_{зв}$ определяет инерцию звена сети при накоплении данных.

Например, если $K_{зв} = 1$, $\Delta = 0{,}05$, то $T_{зв} \approx \approx 1$ мкс. Получается простая передаточная функция звена $W_{зв}(s) = 1/(s + 1)$. При $T_{п.п} = 3T_{зв} = = 3$ мкс процесс накопления данных в каждом звене заканчивается с точностью 0,05.

Таким образом, информация о динамических характеристиках звена сети включает время накопления данных ($T_{п.п} = 3T_{зв}$) и относительный объем данных в стационарном режиме ($W_{зв}(0) = K_{зв}$). Для проведения моделирования и достижения цели статьи необходимо построить динамическую модель беступиковых маршрутов, по которым передаются информационные потоки в сети с червячной коммутацией.

Динамическая модель маршрута вычислительной сети

Маршрут ВС может быть представлен цепочкой звеньев согласно последовательному перечислению номеров узлов в схеме сети. Так, маршрут $1 \to 3 \to 4 \to 6$ в вычислительной сети, приведенной на рис. 3, является последовательным соединением узла № *1* (вычислительного устройства *1*), узла № *3* (коммутатора *3*), узла № *4* (коммутатора *4*) и узла № *6* (вычислительного устройства *6*).

Следовательно, динамическая модель маршрута представляет собой последовательное соединение апериодических звеньев 1-го порядка. В таком случае передаточная функция маршрута вычисляется как произведение их передаточных функций. Коэффициент передачи будет произведением коэффициентов передачи вычислительного устройства и коммутатора. Порядок динамической системы будет равен количеству звеньев. Полюсами ее служат обратные значения постоянных времени звеньев, взятые со знаком минус.

В общем случае для построения формулы передаточной функции маршрута в ВС необходимо задать три матрицы: **A**, **B** и **C**. Матрица **A** определяет цепочку связанных узлов, где столбцы определяют направление, откуда передается пакет с данными. Строки определяют направление, куда пакет приходит. Матрицы **B** и **C** определяют номера начальной и конечной точек маршрута. Для рассмотренного маршрута $1 \to 3 \to 4 \to 6$ матрицы будут иметь вид

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix};$$

$$\mathbf{B} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Согласно матрице **A** на вход коммутатора *3* приходит выход вычислительного устройства *1*, и выход коммутатора *3* поступает на вход коммутатора *4*, а его выход поступает на вход вычислительного устройства *6*. Матрицы **B**, **C** определяют, что начальным является вычислитель *1*, а конечным — вычислитель *6*.

После построения указанных матриц должна быть сформирована диагональная матрица передаточных функций. Для рассматриваемого маршрута она будет иметь вид

$$\mathbf{W}(s) = \begin{bmatrix} W_c(s) & 0 & 0 & 0 & 0 & 0 \\ 0 & W_k(s) & 0 & 0 & 0 & 0 \\ 0 & 0 & W_k(s) & 0 & 0 & 0 \\ 0 & 0 & 0 & W_k(s) & 0 & 0 \\ 0 & 0 & 0 & 0 & W_k(s) & 0 \\ 0 & 0 & 0 & 0 & 0 & W_c(s) \end{bmatrix}.$$

Тогда передаточную функцию маршрута можно вычислить следующим образом:

$$W_m(s) = \mathbf{C} \cdot (\mathbf{E} - \mathbf{W}(s) \cdot \mathbf{A}) \cdot \mathbf{W}(s) \cdot \mathbf{B}.$$

Здесь **E** — единичная матрица. С помощью программы-сценария математического программного обеспечения передаточная функция рассматриваемого маршрута будет рассчитана при числовых значениях параметров передаточных функций вычислителей и коммутаторов в следующем разделе.

Моделирование динамики маршрутов вычислительной сети

Моделирование динамики [30, 31] представленных выше маршрутов позволит проверить предлагаемые динамические модели.

Моделирование динамики накопления данных в конечном узле маршрута $1 \to 2 \to 3 \to 4$ для сети рис. 2 выполнено в специально разработанной программе в математическом пакете MatLab. Для этого маршрута матрицы имеют следующий вид:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}; \ \mathbf{B} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \ \mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}.$$

Учитывая передаточную функцию отдельного звена

$$W_{\text{зв}} \quad \frac{1}{s \quad 1},$$

можно получить передаточную функцию для всего маршрута

$$W_m = \frac{1}{s^4 + 4s^3 + 6s^2 + 4s + 1}.$$

В результате проведенного моделирования получены четыре полюса маршрута: –1,0002 + + 0,0000i; –1,0000 + 0,0002i; –1,0000 – 0,0002i; –0,9998 + 0,0000i.

Поведение весовой и переходной характеристик показано на рис. 5.

Из графика переходной характеристики видно, что время переходного процесса $T_{\text{п.п}}$ занимает ≈ 12 мкс, время разгона $T_{\text{р}} \approx 3$ мкс, а время запаздывания $T_{\text{з}} \approx 1$ мкс.

При моделировании реакции системы на входной сигнал подавался единичный импульс длительностью 5 мкс (рис. 6, *а*). Этот показатель почти в 2 раза меньше времени переходного процесса. Выходной сигнал системы представлен на рис. 6, *б*.

Если длина пакета данных меньше 12 мкс, то происходит потеря данных в размере ≈ 15 %, как это видно на рис. 6, *б*. $T_{\text{р}}$ составляет 11 мкс.

Выявлены следующие динамические характеристики рассматриваемого маршрута:

– стационарный коэффициент передачи = 1;
– степень устойчивости ≈ 1.

Полученные данные подтверждают беступиковость рассмотренного маршрута. Если при моделировании маршрута с «тупиком» стационарный коэффициент передачи будет становиться бесконечным или отрицательным, то весовая и переходная характеристики будут стремиться к бесконечности.

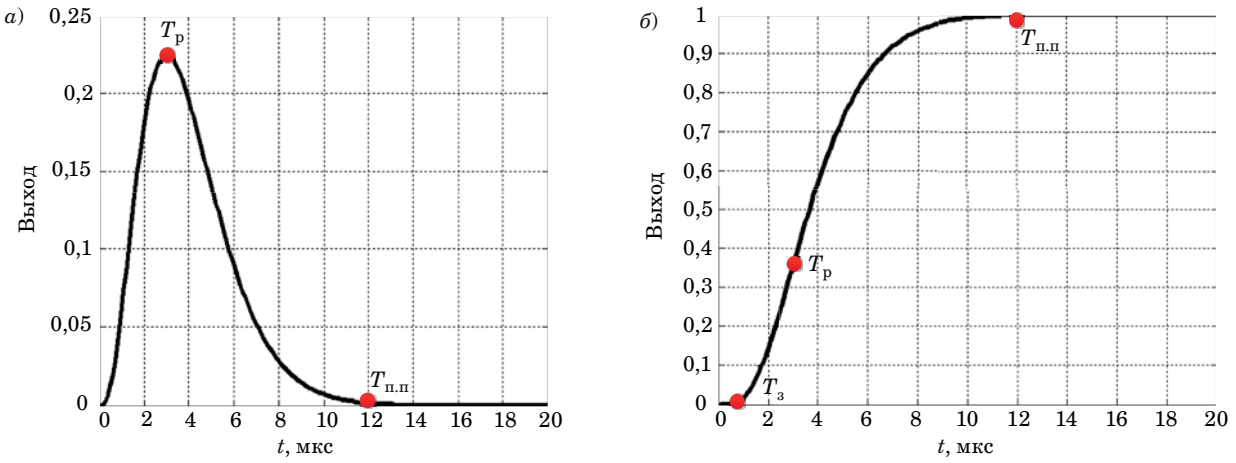
Моделирование динамики накопления данных в конечном узле маршрута $1 \to 3 \to 4 \to 6$ для сети рис. 3 выполнялось при помощи двух раздельных типов устройств: вычислителей и коммутаторов. Матрицы **A**, **B** и **C** для маршрута имеют вид, представленный выше. Передаточные функции вычислительного устройства и коммутатора заданы следующим образом:

$$W_c = \frac{1}{s+1}; \ W_k = \frac{1}{4 \cdot s + 1}.$$

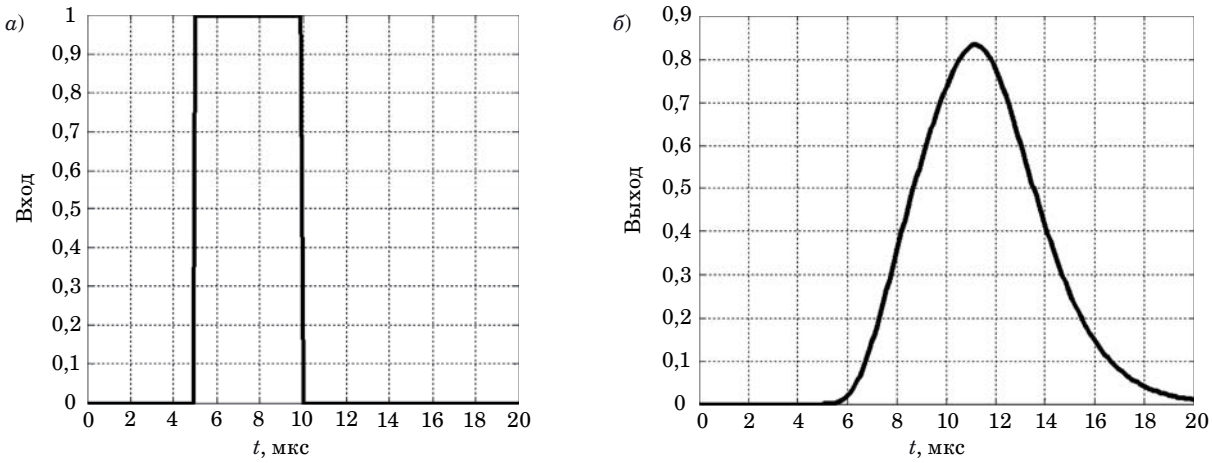
Получена передаточная функция маршрута

$$W_m = \frac{0{,}0625}{s^4 + 2{,}5s^3 + 2{,}062s^2 + 0{,}625s + 0{,}0625}.$$

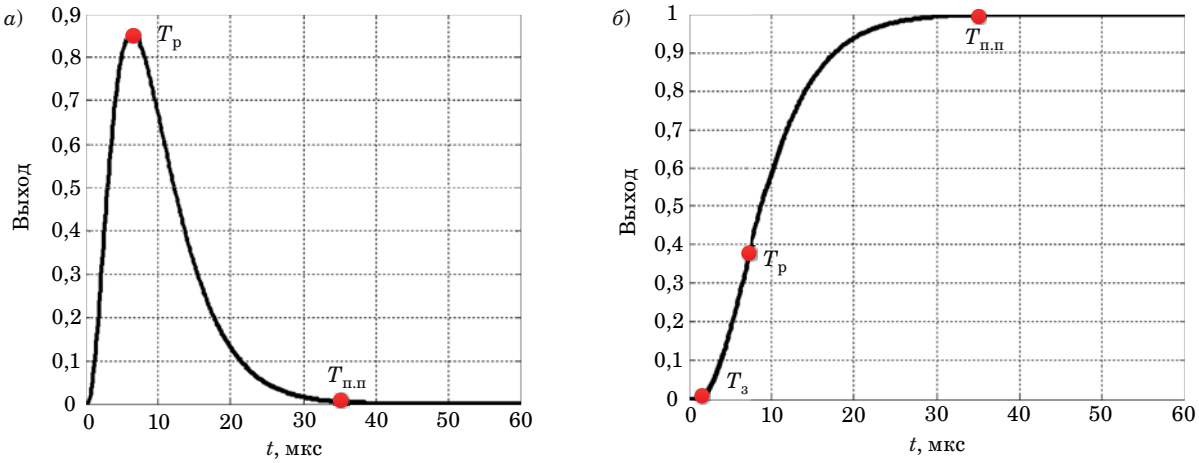
В результате моделирования, проведенного в математическом программном обеспечении, были построены весовая и переходная характеристики (рис. 7, *а* и *б*). Определены четыре полюса рассматриваемого маршрута: –1,0000 + 0,0000i; –1,0000 – 0,0000i; –0,2500 + 0,0000i; –0,2500 –



■ ***Рис. 5.*** Весовая (*а*) и переходная (*б*) характеристики маршрута $1 \to 2 \to 3 \to 4$
■ ***Fig. 5.*** Weight (*а*) and transient (*б*) characteristics of route $1 \to 2 \to 3 \to 4$



■ ***Рис. 6.*** Входной импульсный сигнал (*а*) и выходной сигнал (*б*) ВС
■ ***Fig. 6.*** Input pulse signal (*а*) and output signal (*б*) network



■ ***Рис. 7.*** Весовая (*а*) и переходная (*б*) характеристики маршрута $1 \to 3 \to 4 \to 6$
■ ***Fig. 7.*** Weight (*а*) and transient (*б*) characteristics of route $1 \to 3 \to 4 \to 6$

– 0,0000i, — которые определяют степень устойчивости маршрута 0,25.

Согласно графикам, $T_{п.п} \approx 50$ мкс, $T_р \approx 7$ мкс, $T_з \approx 5$ мкс.

Выявлены следующие динамические характеристики рассматриваемого маршрута:

— стационарный коэффициент передачи = 1;
— степень устойчивости ≈ 0,25.

При моделировании реакции системы на входной сигнал подавался единичный импульс длительностью 5 мкс (рис. 8, *а*). Длительность входного сигнала почти в 10 раз меньше времени переходного процесса. Выходной сигнал системы показан на рис. 8, *б*. Длина пакета данных должна быть не меньше 50 мкс. Потеря данных ≈ 60 %, как это видно на рис. 8, *б*.

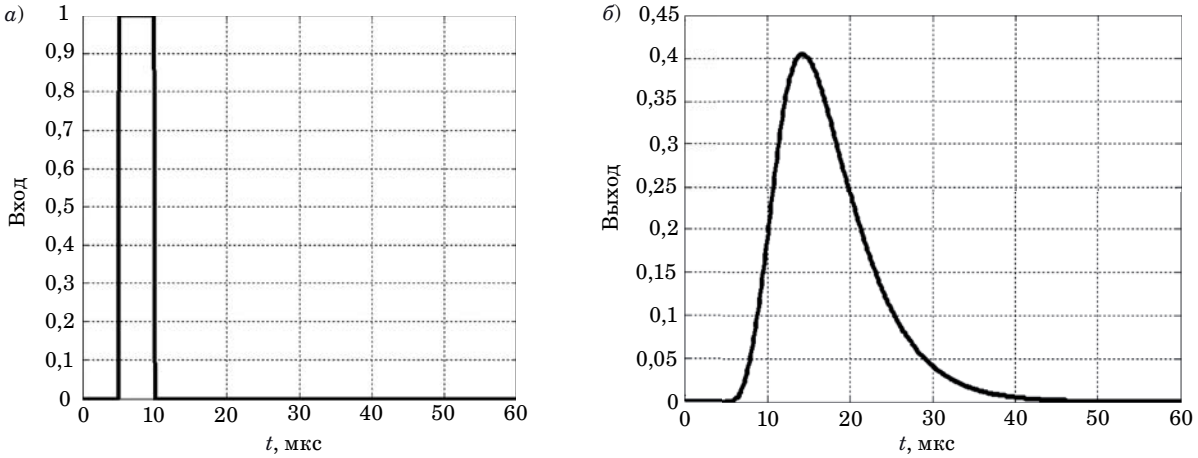
Моделирование динамики накопления данных в конечном узле маршрута узел 248 → коммутатор 299 → коммутатор 332 → узел 251 сетевой структуры рис. 4 выполнялось при помощи матриц и передаточных функций звеньев маршрута, построенных аналогично ранее рассмотренным случаям. При этом передаточная функция самого маршрута имеет вид

$$W_m = \frac{0,01563}{s^4 + 1,75s^3 + 0,9375s^2 + 0,2031s + 0,01563}.$$

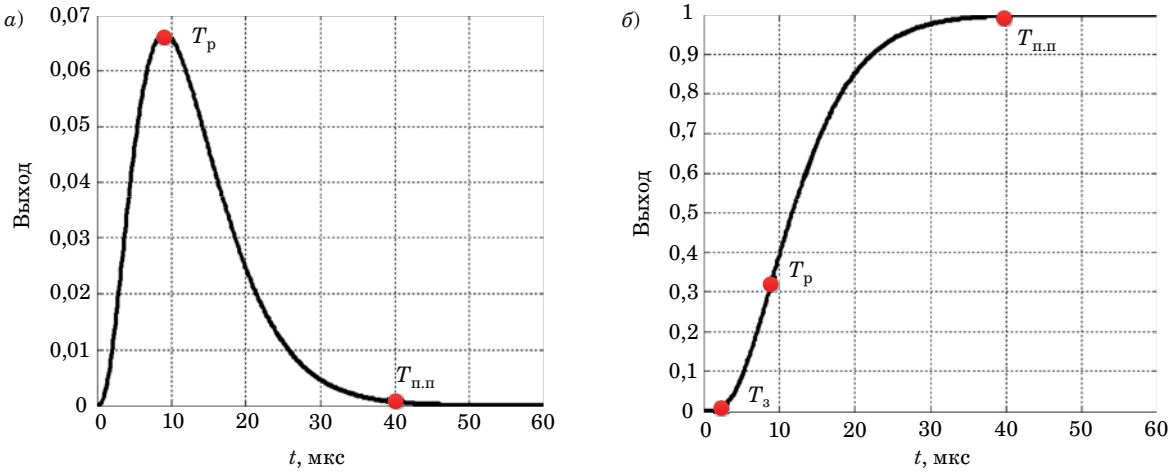
Весовая и переходная характеристики маршрута узел 248 → коммутатор 299 → коммутатор 332 → узел 251 показаны на рис. 9, *а*, *б*.

Динамические характеристики для данного маршрута имеют следующие параметры:

— стационарный коэффициент передачи = 1;
— степень устойчивости ≈ 0,2500;
— время переходного процесса 40 мкс;



■ ***Рис. 8.*** Оценка потери данных при помощи входного импульсного сигнала (*а*) и выходного сигнала (*б*) ВС
■ ***Fig. 8.*** Input pulse signal (*а*) and output signal (*б*) network



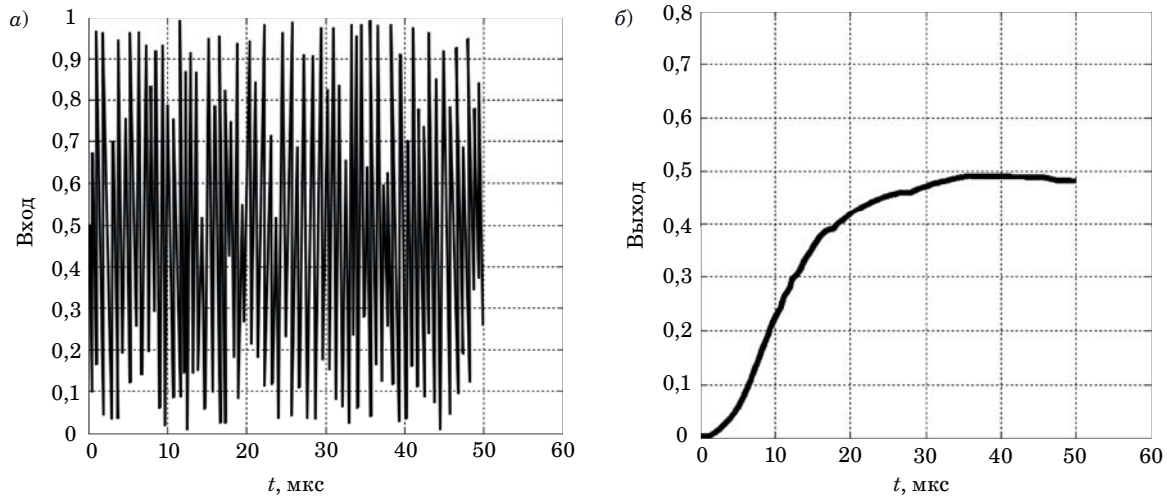
■ ***Рис. 9.*** Весовая (*а*) и переходная (*б*) характеристики маршрута узел 248 → коммутатор 299 → коммутатор 332 → узел 251
■ ***Fig. 9.*** Weight (*а*) and transient (*б*) characteristics of route node 248 → switch 299 → switch 332 → node 251

– время разгона 9 мкс;
– время запаздывания 2 мкс.

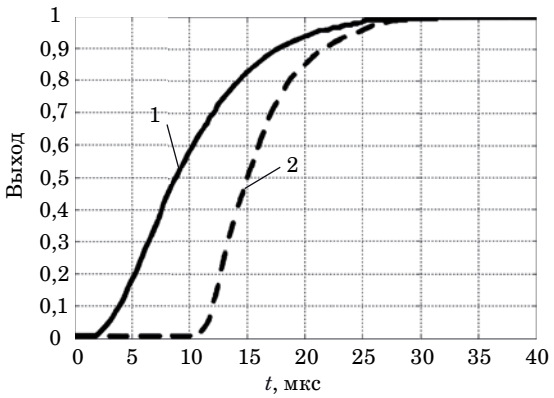
Для исследования данного маршрута в условиях высокозагруженной сети на вход системы подавался случайный сигнал, отражающий различные события в сети. Реакция системы на сложный сигнал показана на рис. 10, *б*. Как видно из графика, характер поведения выходного сигнала является стабильным и маршрут проложен верно.

Результаты моделирования двух видов ВС показывают справедливость принятых подходов к построению динамических моделей сетей. Следует отметить, что все стационарные коэффициенты передачи равны 1. Все модели сетей устойчивы, их степени устойчивости равны 1 и 0,25. В замкнутых маршрутах из-за образования положительной обратной связи возможен информационный коллапс, когда в узле-источнике происходит бесконечное накопление данных. При этом стационарный коэффициент передачи становится бесконечным, а весовая и переходная характеристики стремятся к бесконечности.

Предложенная модель динамики накопления данных в информационных потоках позволяет получить результаты совместного прохождения червячных маршрутов через один коммутатор. Например, рис. 3 показывает, что через коммутатор *3* проходят два маршрута $1 \to 3 \to 5$; $1 \to 3 \to 4 \to 6$. На рис. 11 показаны результаты моделирования процесса совместного использования рассматриваемых маршрутов. Оба маршрута не мешают друг другу.



■ ***Рис. 10.*** Входной случайный сигнал (*а*) и выходной сигнал (*б*) ВС
■ ***Fig. 10.*** Input random signal (*а*) and output signal (*б*) network



■ ***Рис. 11.*** Сравнение переходных характеристик маршрутов $1 \to 3 \to 4 \to 6$ (линия *1*) и $1 \to 3 \to 5$ (линия *2*)
■ ***Fig. 11.*** Comparison of transient responses of routes $1 \to 3 \to 4 \to 6$ (line *1*) and $1 \to 3 \to 5$ (line *2*)

Выполненное моделирование относится к сетям не очень высокого порядка. Однако следует отметить, что построенные динамические модели являются линейными системами. Поэтому увеличение порядка сетей не приведет к получению любых новых результатов.

Заключение

В рамках проведенного исследования обоснована актуальность получения динамических характеристик, описывающих накопление данных информационных потоков в маршрутах ВС с беступиковой маршрутизацией и червячной коммутацией. Выделены особенности таких сетей, представлены динамические модели узлов и беступиковых маршрутов. Впервые разработаны динамические модели узлов и беступиковых маршрутов в ВС с червячной коммутацией, и также впервые получен набор динамических характеристик процессов накопления данных от узла-источника к узлу-приемнику сети. Важным результатом работы являются данные моделирования процессов накопления данных в информационных потоках в беступиковых маршрутах ВС с червячной коммутацией. В частности, достигнута поставленная цель по получению динамических характеристик червячных маршрутов.

По результатам моделирования определено, что динамические характеристики соответствуют ожидаемым результатам. Оценены такие параметры, как стационарный коэффициент передачи, степень устойчивости, время переходного процесса, время разгона и время запаздывания.

Динамические модели маршрутов оказались устойчивыми, что обеспечивает стабильную работу всей системы. Для определения устойчивости использовались полюса передаточной функции маршрута. Согласно полученным результатам устойчивая динамика наблюдалась для всех исследуемых беступиковых маршрутов, что говорит об успешно пройденной проверке и корректности проложенных маршрутов рассматриваемых сетей. Все полученные параметры удовлетворяют ограничениям, принятым в сетях с червячной коммутацией.

Результаты исследования могут быть использованы при проектировании ВС в специализированном программном обеспечении.

Литература

1. **Gouveia P., Neves J., Segarra C., Liechti L., Issa Sh., Schiavoni V., Matos M.** Kollaps: decentralized and dynamic topology emulation. *Proc. of the Fifteenth European Conf. on Computer Systems (EuroSys'20)*, April 2020, New York, NY, USA, 2020, Article 23, pp. 1–16. doi:10.1145/3342195.3387540
2. **Mukul K.** Evolution of routers. *International Journal of Advanced Research in Science, Communication and Technology*, 2021, vol. 10, no. 10, pp. 510–518. doi:10.48175/IJARSCT-1284
3. **Camarero C., Carmen M., Enrique V., Ramón B.** Projective networks: Topologies for large parallel computer systems. *IEEE Transactions on Parallel and Distributed Systems*, 2017, vol. 28, no. 7, pp. 1–14. doi:10.48550/arXiv.1512.07574
4. **Ebert A., Wu P., Mengersen K., Ruggeri F.** Computationally efficient simulation of queues: The R Package queuecomputer. *Journal of Statistical Software*, 2020, vol. 95, no. 5, pp. 1–29. doi:10.18637/jss.v095.i05
5. **Wong R. T.** Telecommunications network design: Technology impacts and future directions. *Networks*, 2021, vol. 76, no. 3 pp. 205–224. doi:10.1002/net.21997
6. **Fanou R., Huffaker B., Mok R. K., Claffy K.** Unintended consequences: effects of submarine cable deployment on internet routing. *Passive and Active Measurement*. PAM, Lecture Notes in Computer Science, 2020, vol. 12048, no. 1, pp. 211–227. doi:10.1007/978-3-030-44081-7_13
7. **Al-Ani D. R., Al-Ani A. R.** The performance of IPv4 and IPv6 in terms of routing protocols using GNS 3 simulator. *Procedia Computer Science*, 2018, vol. 130, no. 1, pp. 1051–1056. doi:10.1016/j.procs.2018.04.147
8. **Bolanowski M., Byczek T.** Measure and compare the convergence time of network routing protocols. *ITM Web of Conferences 21(1)*, 2018, vol. 21, no. 13, pp. 1–9. doi:10.1051/itmconf/20182100013
9. **Sochor T., Sochorova H.** Dynamic routing protocol convergence in simulated and real IPv4 and IPv6 networks. *Cybernetics and Automation Control Theory Methods in Intelligent Algorithms, Proc. of 8th Computer Science On-line Conf.*, 2019, vol. 986, no. 3, pp. 117–126. doi:10.1007/978-3-030-19813-8_13
10. **Kala S. M., Sathya V., Seah W. K. G., Yamaguchi H., Higashino T.** Evaluation of theoretical interference estimation metrics for dense Wi-Fi networks. *Intern. Conf. on COMmunication Systems & NETworkS (COMSNETS),* Bengaluru, India, 6–8 Jan 2021, pp. 351–359. doi:10.1109/COMSNETS51098.2021.9352925
11. **Osukhivska H., Tysh I., Lobur T., Shylinska I., Lupenko S.** Method for estimating the convergence parameters of dynamic routing protocols in computer networks. *IEEE 16th Intern. Conf. on Computer Sciences and Information Technologies (CSIT)*, 2021, vol. 1, no. 1, pp. 228–231. doi:10.1109/CSIT52700.2021.9648792
12. **Xu L., Liang Y., Duan Z., Zhou G.** Route-based dynamics modeling and tracking with application to air traffic surveillance. *IEEE Transactions on Intelligent Transportation Systems*, 2020, vol. 21, no. 1, pp. 209–221. doi:10.1109/TITS.2018.2890570
13. **Семенов В. Ю.** Исследование и анализ средств и методов мониторинга вычислительных сетей. *Решетневские чтения*, 2014, т. 2, № 18, с. 328–330.
14. **Cordeiro M., Sarmento R., Brazdil P., Gama J.** Evolving networks and social network analysis methods and techniques. *Social Media and Journalism — Trends, Connections, Implications*, 2018, vol. 7, no. 1, pp. 101–134. doi:10.5772/intechopen.79041
15. **Camacho D., Panizo-Lledot Á., Bello Orgaz G., Gonzalez-Pardo A., Cambria E.** The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion*, 2020, vol. 63, no. 1, pp. 88–120. doi:10.1016/j.inffus.2020.05.009
16. **Mangat V.** Modeling information flow in social networks: Metrics and evaluation. *Journal of Cognitive Human-Computer Interaction*, 2022, vol. 2, no. 1, pp. 8–14. doi:10.54216/JCHCI.020101
17. **Biyun C., Hongbin L., Bin Z.** Real-time identification of false data injection attacks: A novel dynamic-static parallel state estimation based mechanism. *IEEE Access*, 2019, vol. 7, no. 1, pp. 1–9. doi:10.1109/ACCESS.2019.2929785
18. **Kwon S., Yoo H., Shon T.** IEEE 1815.1 — Based power system security with bidirectional RNN-based network anomalous attack detection for cyber-physical system. *IEEE Access*, 2020, vol. 8, no. 1, pp. 77572–77586, doi:10.1109/ACCESS.2020.2989770
19. **Han S.-Y., Sun Q.-W., Zhao Q., Han R.-Z., Chen Y.-H.** Traffic forecasting based on integration of adaptive subgraph reformulation and spatio-temporal deep learning model. *Electronics*, 2022, vol. 11, no. 6, pp. 1–13. doi:10.3390/electronics11060861
20. **Bonniot L.** *Computer network modeling and root cause analysis with statistical learning.* Université Rennes 1, France, RISA (UMR 6074), 2021. 173 p.
21. **Guideng X.** Analysis on the influence of artificial intelligence in computer network technology. *2nd Intern. Conf. on Intelligent Systems Research and Mechatronics Engineering (ISRME 2019)*, 2019, vol. 1, no. 1, pp. 437–440. doi:10.25236/isrme.2019.085
22. **Guo K., Hu Y., Qian Zh., Liu H., Zhang K., Sun Y., Gao J., Yin B.** Optimized graph convolution recurrent neural network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2021, vol. 22, no. 2, pp. 1138–1149. doi:10.1109/TITS.2019.2963722
23. **Duato J.** A necessary and sufficient condition for deadlock-free adaptive routing in wormhole networks. *IEEE Transactions on Parallel and Distributed Systems*, 1995, vol. 6, no. 10, pp. 1055–1067. doi:10.1109/71.473515
24. **Ma F., Zhu Y., Yuen K. F., Sun Q., He H., Xu X., Shang Z., Xu Y.** Exploring the spatiotemporal evolu-

tion and sustainable driving factors of information flow network: A public search attention perspective. *Int. J. Environ. Res. Public Health 2022*, 2022, vol. 19, no. 1, pp. 1–25. doi:10.3390/ijerph19010489

25. **Лесько С. А., Жуков Д. О., Истратов Л. А.** Модели описания динамики блокировки узлов вычислительных сетей вирусами на основе использования перколяционных кинетических и стохастических методов. *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*, 2020, № 52, с. 22–32. doi:10.17223/19988605/52/3
26. **Tereikovskyi I., Tereykovska L., Mussiraliyeva Sh., Tsiutsiura M., Achkoski Ju.** Markov model of unsteady profile of normal behavior of network objects of computer systems. *Proc. of the Intern. Workshop on Conflict Management in Global Information Networks (CMiGIN 2019)*, 2019, vol. 1, no. 1, pp. 1–13.
27. **Gray C., Mosig C., Bush R., Pelsser C., Roughan M., Schmidt T. C., Wahlisch M.** BGP beacons, network tomography, and Bayesian computation to locate route flap damping. *Proc. of the ACM Internet Measurement Conf. (IMC '20)*, Association for Computing Machinery, New York, NY, USA, 2020, vol. 20, no. 1, pp. 492–505. doi:10.1145/3419394.3423624
28. **Evstafiev V., Rudenko N.** Improving the dynamics of information flows for optimizing telecommunication systems. *IOP Conf. Series: Materials Science and Engineering*, 2021, vol. 1029, no. 1, pp. 1–11. doi:10.1088/1757-899X/1029/1/012131
29. **Karandashev A. A., Olenev V. L.** Selection methods for deadlock-free routes' optimal configuration in onboard SpaceWire networks. *Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, IEEE, 2021, vol. 21, no. 1, pp. 1–10. doi:10.1109/WECONF51603.2021.9470619
30. **Бритов Г. С.** Верификация, валидация и тестирование компьютерных моделей линейных динамических систем. *Информационно-управляющие системы*, 2013, № 2, с. 75–82.
31. **Бритов Г. С.** Бизнес-процесс тестового диагностирования линейных динамических систем на основе передаточных функций. *Системный анализ и логистика*, 2016, т. 12, № 1, с. 4–11.

UDC 004.021
doi:10.31799/1684-8853-2023-3-39-50
EDN: AUBKNL

Modeling the dynamics of information flows in the routes of computer networks

A. A. Karandashev[a], Post-Graduate Student, orcid.org/0000-0002-0358-0165
V. L. Olenev[a], PhD, Tech., Associate Professor, orcid.org/0000-0002-1817-2754, Valentin.Olenev@guap.ru
G. S. Britov[a], PhD, Tech., Associate Professor, orcrid.org/0000-0002-0452-523X
[a]Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation

Introduction: For the development of modern highly specialized software systems for designing wormhole routing networks, it is necessary to compile routing tables. This requires obtaining efficient configurations of deadlock-free routes laid at the initial stages of network design. **Purpose:** To make the simulation of the dynamics of inertial processes of data accumulation in the receiving nodes of computer networks during the transmission of packets from transmitter nodes. **Results:** Three examples of network schemes are obtained, which allows us to consider classical network structures with different degrees of computational load. For each network, we construct matrices that describe its structure. As a result of the simulation, using specially developed scenarios in the mathematical software, we have demonstrated the resulting graphs of weight and transient characteristics and provide a number of characteristics of each route under study. The simulation carried out makes it possible to make sure that there is no information collapse of the routes laid in deadlock-free routes. A separate study of two simultaneously functioning data transmission paths in a network with "worm" switching also shows stable transient characteristics, which thus indicates the correctness of the proposed dynamic models and the deadlock-free wormhole routes themselves. **Practical relevance:** The results of the study can be applied for the design of computer networks with the use of specialized software. Dynamic performance analysis is intended to help in making decisions about the choice of deadlock-free routes offered by a designer of a wormhole switching network.

Keywords — data transmission dynamics, information flows, deadlock-free routes, wormhole routing techniques, weight characteristics, transient characteristics, data accumulation process, route models.

For citation: Karandashev A. A., Olenev V. L., Britov G. S. Modeling the dynamics of information flows in the routes of computer networks. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 39–50 (In Russian). doi:10.31799/1684-8853-2023-3-39-50, EDN: AUBKNL

References

1. Gouveia P., Neves J., Segarra C., Liechti L., Issa Sh., Schiavoni V., Matos M. Kollaps: decentralized and dynamic topology emulation. *Proc. of the Fifteenth European Conf. on Computer Systems (EuroSys'20)*, April 2020, New York, NY, USA, 2020, Article 23, pp. 1–16. doi:10.1145/3342195.3387540
2. Mukul K. Evolution of routers. *International Journal of Advanced Research in Science, Communication and Technology*, 2021, vol. 10, no. 10, pp. 510–518. doi:10.48175/IJARSCT-1284
3. Camarero C., Carmen M., Enrique V., Ramón B. Projective networks: Topologies for large parallel computer systems. *IEEE Transactions on Parallel and Distributed Systems*, 2017, vol. 28, no. 7, pp. 1–14. doi:10.48550/arXiv.1512.07574

4. Ebert A., Wu P., Mengersen K., Ruggeri F. Computationally efficient simulation of queues: The R Package queuecomputer. *Journal of Statistical Software*, 2020, vol. 95, no. 5, pp. 1–29. doi:10.18637/jss.v095.i05
5. Wong R. T. Telecommunications network design: Technology impacts and future directions. *Networks*, 2021, vol. 76, no. 3 pp. 205–224. doi:10.1002/net.21997
6. Fanou R., Huffaker B., Mok R. K., Claffy K. Unintended consequences: effects of submarine cable deployment on internet routing. *Passive and Active Measurement*. PAM, Lecture Notes in Computer Science, 2020, vol. 12048, no. 1, pp. 211–227. doi:10.1007/978-3-030-44081-7_13
7. Al-Ani D. R., Al-Ani A. R. The performance of IPv4 and IPv6 in terms of routing protocols using GNS 3 simulator. *Procedia Computer Science*, 2018, vol. 130, no. 1, pp. 1051–1056. doi:10.1016/j.procs.2018.04.147
8. Bolanowski M., Byczek T. Measure and compare the convergence time of network routing protocols. *ITM Web of Conferences 21(1)*, 2018, vol. 21, no. 13, pp. 1–9. doi:10.1051/itmconf/20182100013
9. Sochor T., Sochorova H. Dynamic routing protocol convergence in simulated and real IPv4 and IPv6 networks. *Cybernetics and Automation Control Theory Methods in Intelligent Algorithms, Proc. of 8th Computer Science On-line Conf.*, 2019, vol. 986, no. 3, pp. 117–126. doi:10.1007/978-3-030-19813-8_13
10. Kala S. M., Sathya V., Seah W. K. G., Yamaguchi H., Higashino T. Evaluation of theoretical interference estimation metrics for dense Wi-Fi networks. *Intern. Conf. on COMmunication Systems & NETworkS (COMSNETS)*, Bengaluru, India, 6–8 Jan 2021, pp. 351–359. doi:10.1109/COMSNETS51098.2021.9352925
11. Osukhivska H., Tysh I., Lobur T., Shylinska I., Lupenko S. Method for estimating the convergence parameters of dynamic routing protocols in computer networks. *IEEE 16th Intern. Conf. on Computer Sciences and Information Technologies (CSIT)*, 2021, vol. 1, no. 1, pp. 228–231. doi:10.1109/CSIT52700.2021.9648792
12. Xu L., Liang Y., Duan Z., Zhou G. Route-based dynamics modeling and tracking with application to air traffic surveillance. *IEEE Transactions on Intelligent Transportation Systems*, 2020, vol. 21, no. 1, pp. 209–221. doi:10.1109/TITS.2018.2890570
13. Semenov V. Yu. Research and analysis of tools and methods for monitoring computer networks. *Reshetnevskie chteniya*, 2014, vol. 2, no. 18, pp. 328–330 (In Russian)
14. Cordeiro M., Sarmento R., Brazdil P., Gama J. Evolving networks and social network analysis methods and techniques. *Social Media and Journalism — Trends, Connections, Implications*, 2018, vol. 7, no. 1, pp. 101–134. doi:10.5772/intechopen.79041
15. Camacho D., Panizo-Lledot Á., Bello Orgaz G., Gonzalez-Pardo A., Cambria E. The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion*, 2020, vol. 63, no. 1, pp. 88–120. doi:10.1016/j.inffus.2020.05.009
16. Mangat V. Modeling information flow in social networks: Metrics and evaluation. *Journal of Cognitive Human-Computer Interaction*, 2022, vol. 2, no. 1, pp. 8–14. doi:10.54216/JCHCI.020101
17. Biyun C., Hongbin L., Bin Z. Real-time identification of false data injection attacks: A novel dynamic-static parallel state estimation based mechanism. *IEEE Access*, 2019, vol. 7, no. 1, pp. 1–9. doi:10.1109/ACCESS.2019.2929785
18. Kwon S., Yoo H., Shon T. IEEE 1815.1 — Based power system security with bidirectional RNN-based network anomalous attack detection for cyber-physical system. *IEEE Access*, 2020, vol. 8, no. 1, pp. 77572–77586, doi:10.1109/ACCESS.2020.2989770
19. Han S.-Y., Sun Q.-W., Zhao Q., Han R.-Z., Chen Y.-H. Traffic forecasting based on integration of adaptive subgraph reformulation and spatio-temporal deep learning model. *Electronics*, 2022, vol. 11, no. 6, pp. 1–13. doi:10.3390/electronics11060861
20. Bonniot L. *Computer network modeling and root cause analysis with statistical learning*. Université Rennes 1, France, RISA (UMR 6074), 2021. 173 p.
21. Guideng X. Analysis on the influence of artificial intelligence in computer network technology. *2nd Intern. Conf. on Intelligent Systems Research and Mechatronics Engineering (ISRME 2019)*, 2019, vol. 1, no. 1, pp. 437–440. doi:10.25236/isrme.2019.085
22. Guo K., Hu Y., Qian Zh., Liu H., Zhang K., Sun Y., Gao J., Yin B. Optimized graph convolution recurrent neural network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2021, vol. 22, no. 2, pp. 1138–1149. doi:10.1109/TITS.2019.2963722
23. Duato J. A necessary and sufficient condition for deadlock-free adaptive routing in wormhole networks. *IEEE Transactions on Parallel and Distributed Systems*, 1995, vol. 6, no. 10, pp. 1055–1067. doi:10.1109/71.473515
24. Ma F., Zhu Y., Yuen K. F., Sun Q., He H., Xu X., Shang Z., Xu Y. Exploring the spatiotemporal evolution and sustainable driving factors of information flow network: A public search attention perspective. *Int. J. Environ. Res. Public Health 2022*, 2022, vol. 19, no. 1, pp. 1–25. doi:10.3390/ijerph19010489
25. Lesko S. A., Zhukov D. O., Istratov L. A. Models of describing the dynamics of blocking nodes of computer networks by viruses based on the use of percolation, kinetic and stochastic methods. *Tomsk State University Journal of Control and Computer Science*, 2020, no. 52, pp. 22–32 (In Russian). doi:10.17223/19988605/52/3
26. Tereikovskyi I., Tereykovska L., Mussiraliyeva Sh., Tsiutsiura M., Achkoski Ju. Markov model of unsteady profile of normal behavior of network objects of computer systems. *Proc. of the Intern. Workshop on Conflict Management in Global Information Networks (CMiGIN 2019)*, 2019, vol. 1, no. 1, pp. 1–13.
27. Gray C., Mosig C., Bush R., Pelsser C., Roughan M., Schmidt T. C., Wahlisch M. BGP beacons, network tomography, and Bayesian computation to locate route flap damping. *Proc. of the ACM Internet Measurement Conf. (IMC '20)*, Association for Computing Machinery, New York, NY, USA, 2020, vol. 20, no. 1, pp. 492–505. doi:10.1145/3419394.3423624
28. Evstafiev V., Rudenko N. Improving the dynamics of information flows for optimizing telecommunication systems. *IOP Conf. Series: Materials Science and Engineering*, 2021, vol. 1029, no. 1, pp. 1–11. doi:10.1088/1757-899X/1029/1/012131
29. Karandashev A. A., Olenev V. L. Selection methods for deadlock-free routes' optimal configuration in on-board SpaceWire networks. *Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, IEEE, 2021, vol. 21, no. 1, pp. 1–10. doi:10.1109/WECONF51603.2021.9470619
30. Britov G. S. Verification, validation and testing of computer models of linear dynamic systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2013, no. 2, pp. 75–82 (In Russian).
31. Britov G. S. Tests of business processes for diagnosing linear comprehensive systems based on transfer functions. *The System Analysis and Logistics*, 2016, vol. 12, no. 1, pp. 4–11 (In Russian).

УДК 621.396
doi:10.31799/1684-8853-2023-3-51-58
EDN: DACXFO

Научные статьи
Articles

Применение нечеткого регулятора для устранения флуктуаций переходного процесса возмущенной системы автоматического управления углом тангажа летательного аппарата

М. Е. Иванов[a], *аспирант, orcid.org/0000-0002-7584-6946*
В. В. Езерский[a], *доктор техн. наук, профессор, orcid.org/0000-0002-1223-6841, ezerskij2010@yandex.ru*
[a]*Санкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ*

Введение: *к системе автоматического управления угла тангажа летательного аппарата приложены многочисленные возмущающие воздействия, вследствие которых возникают значительные флуктуации переходного процесса по тангажу данной системы.* ***Цель:*** *преобразовать схему управления угла тангажа и разработать для нее нечеткий регулятор, который позволит уменьшить флуктуации и улучшить качество переходного процесса; провести моделирование исходной и модернизированной систем управления.* ***Результаты:*** *проведено моделирование с использованием Matlab Simulink работы классической и интеллектуальной систем управления угла тангажа. По результатам моделирования достигнуто улучшение качества переходного процесса системы управления за счет значительного уменьшения флуктуаций. Для этого в схему системы управления добавлен пропорционально-интегрирующий регулятор, использующий один из методов искусственного интеллекта — нечеткую логику. Настройка нечеткого регулятора осуществлена преобразованием типового алгоритма: для входных лингвистических переменных в два раза уменьшен интервал универсального множества, на котором определены функции принадлежности. Моделирование нечеткой системы произведено в Matlab Fuzzy Logic Toolbox.* ***Практическая значимость:*** *оптимальная настройка нечеткого регулятора позволяет значительно улучшить качество переходного процесса системы управления, на которую действует большое число возмущающих воздействий.*

Ключевые слова *— флуктуация, возмущающее воздействие, искусственный интеллект, система автоматического управления, нечеткий регулятор, переходный процесс.*

Для цитирования: Иванов М. Е., Езерский В. В. Применение нечеткого регулятора для устранения флуктуаций переходного процесса возмущенной системы автоматического управления углом тангажа летательного аппарата. *Информационно-управляющие системы*, 2023, № 3, с. 51–58. doi:10.31799/1684-8853-2023-3-51-58, EDN: DACXFO
For citation: Ivanov M. E., Yezerskiy V. V. The use of the intelligent fuzzy controller for the elimination of fluctuations in the transient process of the perturbed automatic pitch angle control system of an aircraft. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 51–58 (In Russian). doi:10.31799/1684-8853-2023-3-51-58, EDN: DACXFO

Введение

Основная теоретическая база и математический аппарат систем автоматического управления (САУ; Automatic Control Systems, ACS) были разработаны еще в середине XX в. Однако и сегодня данные системы являются объектом регулярных модернизаций, преобразований [1–4]. Происходит интеграция систем в единые измерительные комплексы [5–8], разрабатываются новые инструменты для улучшения качества переходных процессов (уменьшения флуктуаций, уменьшения времени переходного процесса) и т. п. Для этих целей разрабатываются перспективные регуляторы, синтезированные с применением методов искусственного интеллекта — нечеткой логики и искусственной нейронной сети. Так, в статье [9] А. С. Коноваловым и П. Е. Шумиловым исследуется возможность применять нечеткую логику в авиационных системах антиюзовой автоматики. В более современной работе [10] Н. В. Гридина, И. А. Евдокимов и В. И. Солодовников рассматривают использование нечеткой логики для синтеза гибридных искусственных нейронных сетей. Д. О. Пушкарев, Д. Ю. Киселев, Ю. В. Киселев в статье [11] составили математическую модель нейронной сети применительно к расчету авиационных рычажно-поплавковых клапанов. А. В. Шилоносов, Д. В. Молдован, авторы труда [12], провели анализ и синтез сложного нейросетевого регулятора для динамической коррекции управляющих воздействий системы управления воздушно-космического самолета.

Таким образом, создание интеллектуального регулятора САУ является типовой научно-технической задачей, решаемой широким кругом ученых применительно к различным областям в самых разных технических агрегатах. В данной статье рассмотрено использование нечеткого регулятора для

компенсирования возмущающих воздействий на САУ углом тангажа летательного аппарата с целью улучшить качество переходного процесса (по каналу тангажа) системы. Выбор темы для исследования связан с целеполаганием снижения колебательности переходного процесса системы в условиях большого количества возмущающих воздействий.

Краткие теоретические сведения

Рассматриваемая система управления угла тангажа является непрерывной нелинейной системой с незначительной степенью нелинейности. Данную САУ целесообразно заменить линейной моделью: аналитическая линеаризация может быть достигнута разложением функции $\delta_в(t)$ в ряд Тейлора

$$\delta_в(t) = K_\vartheta \cdot \vartheta + K_{\dot\vartheta} \cdot \dot\vartheta + K_{\ddot\vartheta} \cdot \ddot\vartheta + ... + K_{\vartheta^{(n)}} \cdot \vartheta^{(n)}$$

с последующим исключением из ряда слагаемых второй степени и выше. Члены ряда второго и высшего порядков исключаются в предположении о достаточной малости изменения параметра ϑ (отклонения от установившегося значения). Таким образом, для линеаризованной САУ угла тангажа будут справедливы линейные алгоритмы управления, которые используются далее.

В идеальном случае, когда исполнительные устройства являются безынерционными, закон отклонения рулей высоты для управления углом тангажа имеет вид [13–15]

$$\delta_в = K_{\dot\vartheta} \cdot \dot\vartheta + K_\vartheta \cdot (\vartheta - \vartheta_з), \quad (1)$$

где $K_{\dot\vartheta}$ – коэффициент обратной связи по угловой скорости; K_ϑ – коэффициент усиления ошибки по углу тангажа; $\vartheta - \vartheta_з$ – разница между реальным значением угла тангажа и заданным значением этого угла с учетом запаздывания системы.

Синтезируемая САУ угла тангажа требует наложения сразу нескольких условий:

1) появления инерционности – необходимо учитывать дисперсию воздействующего на систему возмущающего воздействия M, состоящего из значений дисперсий локальных возмущающих воздействий: угловой скорости тангажа $M_{\dot\vartheta}$, угловой скорости атаки $M_{\dot\alpha}$, угла атаки M_α и дисперсии случайных погрешностей M_δ:

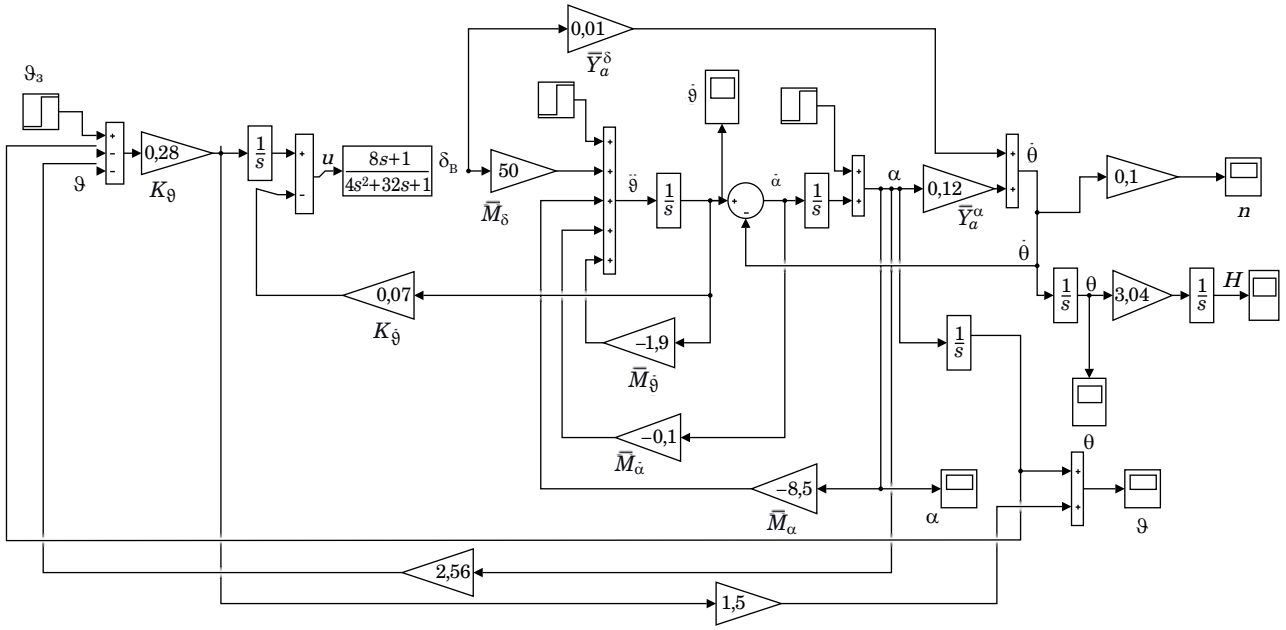
$$\{M\} = \{M_{\dot\vartheta}, M_{\dot\alpha}, M_\alpha, M_\delta\}; \quad (2)$$

2) добавления в САУ звеньев астатизма для сведения к нулю накапливающихся ошибок регулирования. В качестве звена астатизма используется интегратор $1/s$, в таком случае закон управления преобразуется к виду

$$\delta_в = K_{\dot\vartheta} \cdot \dot\vartheta + K_\vartheta \cdot \int (\vartheta - \vartheta_з) dt; \quad (3)$$

3) составления оптимальной передаточной функции рулевого привода руля высоты $W_{р.п} \neq 1$.

Полная структурная схема САУ параметров полета в плоскости симметрии летательного аппарата (рис. 1) позволяет осуществлять управле-



■ ***Рис. 1.*** САУ параметров полета в вертикальной плоскости симметрии

■ ***Fig. 1.*** Automatic control system for flight parameters in the vertical plane of symmetry

ние не только углом тангажа ϑ, но и остальными параметрами полета, измеряемыми в плоскости симметрии: углом атаки α, углом наклона траектории θ, текущей высотой полета H и нормальной (вертикальной) перегрузкой n. Все коэффициенты усилительных звеньев выбраны таким образом, чтобы получить наиболее приемлемые зависимости и обеспечить максимальное быстродействие системы с наименьшим временем переходного процесса по углу тангажа ϑ и углу атаки α.

Переходный процесс возмущенной системы управления

Заданы следующие параметры системы, необходимые для моделирования:

– на вход системы подается детерминированное входное воздействие – единичная ступенчатая функция Хевисайда с параметрами: $\vartheta_з(t) = 1$ (final value), временной шаг 1 (step time), начальное значение 0 (initial value);

– коэффициенты, отвечающие за быстродействие системы, т. е. за качество переходного процесса: $K_\vartheta = 0{,}07$, $K_{\dot\vartheta} = 0{,}28$;

– передаточная функция руля высоты

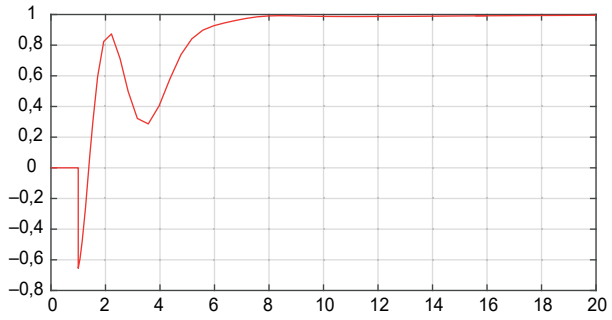
$$W_{р.п}(s) = \frac{8s+1}{4s^2+32s+1}; \qquad (4)$$

– дисперсия случайных погрешностей $\delta = 50$, дисперсии возмущающих воздействий по угловой скорости $M_{\dot\vartheta}$, угловой скорости по углу атаки $M_{\dot\alpha}$ и самому углу атаки M_α соответственно равны –1,9, –0,1 и –8,5;

– коэффициент усиления всей системы, связанный с подъемной силой, Y_a равен 0,01, а этот же коэффициент усиления по углу атаки $Y_a^\alpha = 0{,}12$. Коэффициенты усиления, связанные с вертикальной перегрузкой и с высотой полета, приняты равными 0,1 и 2,34 соответственно. Дополнительные звенья обратной связи и усиления для угла тангажа ϑ соответственно имеют значения 2,56 и 1,5.

В ходе моделирования, время t которого составило 20 с, был получен следующий результат. С помощью заранее заданных коэффициентов усиления при данной передаточной функции рулей высоты удалось максимально улучшить качество переходного процесса от включения системы к установившемуся режиму работы по каналу угла тангажа ϑ (рис. 2).

Увеличение коэффициента усиления ошибки по углу тангажа K_ϑ приводит к потере устойчивости системы, а уменьшение – к значительному ухудшению качества переходного процесса без потери устойчивости. Увеличение коэффициента обратной связи аналогично ухудшает качество переходного процесса и увеличивает время установления рабочего режима, а уменьшение этого коэффициента принципиально не влияет на переходный процесс и не изменяет его.



■ ***Рис. 2.*** Переходный процесс САУ по каналу угла тангажа ϑ

■ ***Fig. 2.*** The ACS transient process along the pitch angle ϑ channel

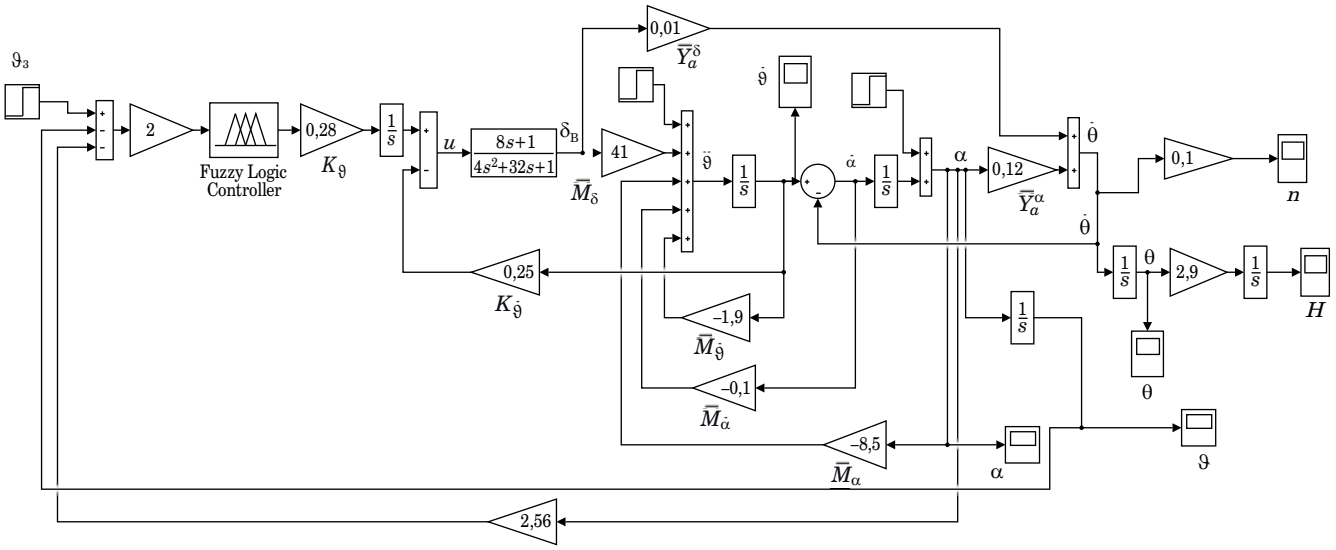
Модернизация системы управления звеном с искусственным интеллектом

Применение интеллектуального нечеткого регулятора позволяет получить более качественный переходный процесс по каналу угла тангажа ϑ даже при наличии большого количества дестабилизирующих возмущений M_i. Нечеткий регулятор может не только значительно снизить флуктуации переходного процесса, но и исключает увеличение амплитудного значения недорегулирования $Y_{\min}$ переходного процесса, возникающего вследствие действия на систему возмущающих воздействий по угловой скорости $M_{\dot\vartheta}$, угловой скорости атаки $M_{\dot\alpha}$ и углу атаки M_α.

Схема измерителей параметров в вертикальной плоскости симметрии (см. рис. 2) с учетом добавления интеллектуального нечеткого регулятора Fuzzy Logic Controller примет вид, представленный на рис. 3. При сохранении той же передаточной функции (4) системы увеличился коэффициент $K_{\dot\vartheta} = 0{,}25$, а также уменьшилась дисперсия $M_\delta = 41$. Кроме того, наличие нечеткого регулятора исключает необходимость в отдельном звене усиления по каналу тангажа K_U 1,5, которое было добавлено для стабилизации исходной САУ. В контур САУ включено усилительное звено перед регулятором $K_{FUZZY} = 2$.

Параметрический синтез нечеткой системы регулятора

Задание параметров нечеткой системы осуществляется по стандартному алгоритму [16–21].



■ ***Рис. 3.*** Интеллектуальная САУ параметров вертикальной плоскости симметрии
■ ***Fig. 3.*** Intelligent ACS parameters of the vertical plane of symmetry

1. Задаются входные лингвистические переменные $\tilde{\alpha}$ (term), для каждой из которых составляются функции принадлежности $\mu_N(x)$ треугольного типа (trimf) в количестве N штук на универсальном множестве [–1; 1].

2. Задаются выходные лингвистические переменные $\tilde{\beta}$ с M треугольными (trimf) функциями принадлежности $\mu_M(y)$.

3. Перед дефаззификацией входных лингвистических переменных на выходе в центральном блоке нечеткой системы записываются правила (Rules) вида “If < input1 is i > then < output1 is j >”.

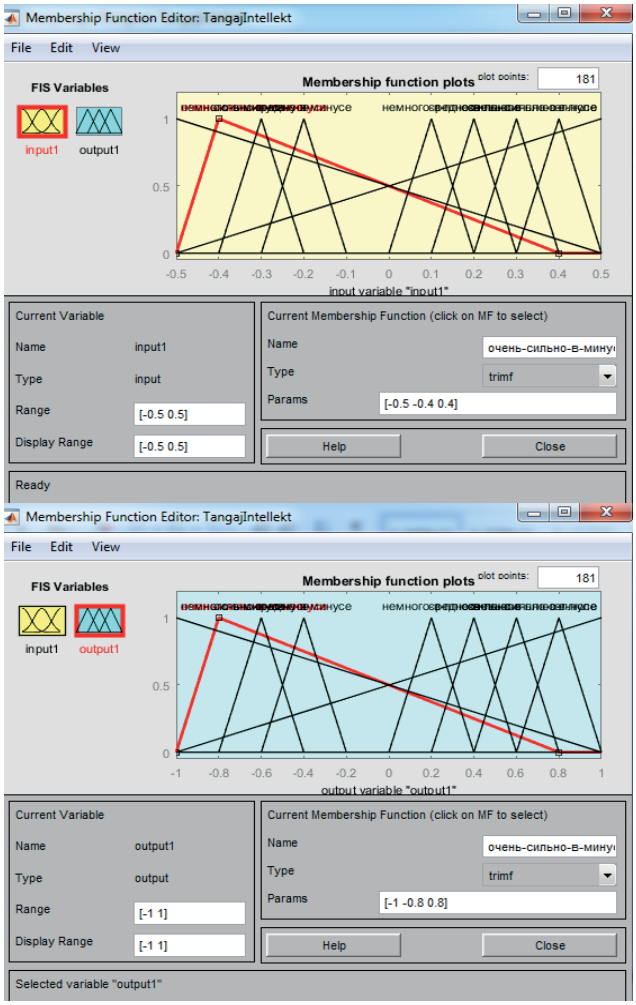
4. Анализируется полученная статическая характеристика output1 (input1) синтезированной нечеткой системы.

Представленный типовой алгоритм был преобразован в целях улучшения качества переходного процесса следующим образом.

1. Для input и output были изменены три симметричные треугольные функции принадлежности: «очень-сильно-в-минусе», «немного-в-минусе» и «в-нуле». Модифицированные функции принадлежности имеют интервалы [–1 –0,8 0,8], [–1,2 –1 1] и [–1 1 1,2] соответственно.

2. Для input был уменьшен в два раза интервал универсального множества, на котором определены $\mu_N(x)$, до [–0,5 0,5]. Таким образом, изменили свои значения как стандартные функции принадлежности, так и модифицированные «очень-сильно-в-минусе», «немного-в-минусе» и «в-нуле», получившие новые интервалы [–0,5 –0,4 0,4], [–0,6 –0,5 0,5] и [–0,5 0,5 0,6] соответственно (рис. 4).

3. В блоке Rule Editor были прописаны решающие правила для функций принадлежности. Всего записано девять неповторяющихся пра-



■ ***Рис. 4.*** Функции принадлежности блоков input и output нечеткой системы
■ ***Fig. 4.*** Membership functions of the input and output blocks of a fuzzy system

вил вида “If input is $\tilde{\alpha}$ then output is $\tilde{\beta}$”, связывающих входные лингвистические переменные с выходными.

Значения интервалов, которые принимают функции принадлежности $\mu_M(y)$ и $\mu_N(x)$ для β и $\tilde{\alpha}$, представлены в табл. 1.

Девять решающих правил (табл. 2) связывают некоторые входные и выходные термы таким образом, чтобы регулятор сформировал целевой показатель устойчивости и качества переходного процесса.

С полученной нечеткой системы была снята следующая статическая характеристика (рис. 5), которая возрастает на всем участке от min input1 до max input1. Она показывает характер нелинейности нечеткого регулятора как отдельного нелинейного звена в составе нелинейной САУ угла тангажа.

Сравнительный анализ классической и интеллектуальной систем управления

С учетом интеллектуализации САУ был получен следующий переходный процесс по каналу измерения угла тангажа ϑ (рис. 6). Данный график позволяет провести сравнительный анализ переходных процессов канала измерения угла тангажа ϑ в случае классической и интеллектуальной САУ (рис. 7).

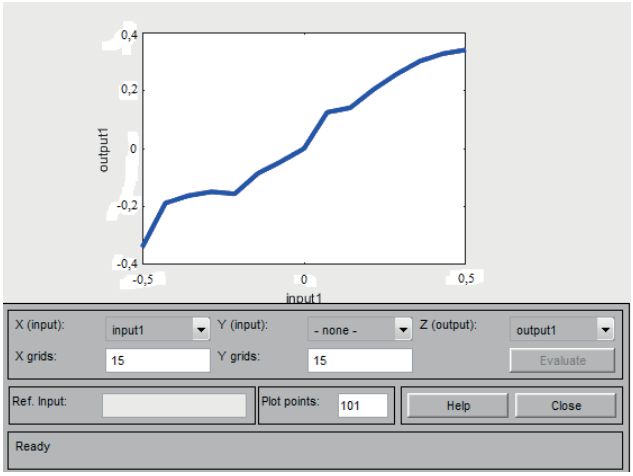
Исходя из данных моделирования, можно сделать следующие выводы.

1. Применение интеллектуального нечеткого регулятора позволило существенно снизить колебательность переходного процесса, в том числе полностью исключить необходимость в недорегулировании системы, что было обязательным условием достижения устойчивости классической САУ.

■ ***Таблица 1.*** Значения интервалов функций принадлежности

■ ***Table 1.*** Values of intervals of membership functions

Термы лингвистических переменных	Интервалы параметров и координаты вершин функций принадлежности	
	для output	для input
Очень-сильно-в-минусе	[–1 –0,8 0,8]	[–0,5 –0,4 0,4]
Сильно-в-минусе	[–0,8 –0,6 –0,4]	[–0,4 –0,3 –0,2]
Средне-в-минусе	[–0,6 –0,4 –0,2]	[–0,3 –0,2 –0,1]
Немного-в-минусе	[–1,2 –1 1]	[–0,6 –0,5 0,5]
В-нуле	[–1 1 1,2]	[–0,5 0,5 0,6]
Немного-в-плюсе	[0 0,2 0,4]	[0 0,1 0,2]
Средне-в-плюсе	[0,2 0,4 0,6]	[0,1 0,2 0,3]
Сильно-в-плюсе	[0,4 0,6 0,8]	[0,2 0,3 0,4]
Очень-сильно-в-плюсе	[0,6 0,8 1]	[0,3 0,4 0,5]



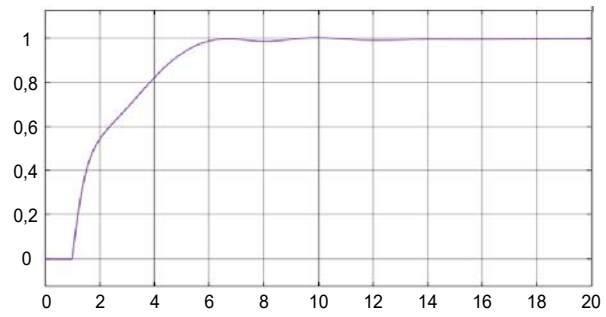
■ ***Рис. 5.*** Статическая характеристика нечеткого регулятора САУ параметров вертикальной плоскости симметрии

■ ***Fig. 5.*** Static characteristic of the fuzzy controller of the ACS parameters of the vertical plane of symmetry

■ ***Таблица 2.*** Правила нечеткой системы для нечеткого регулятора исследуемой САУ

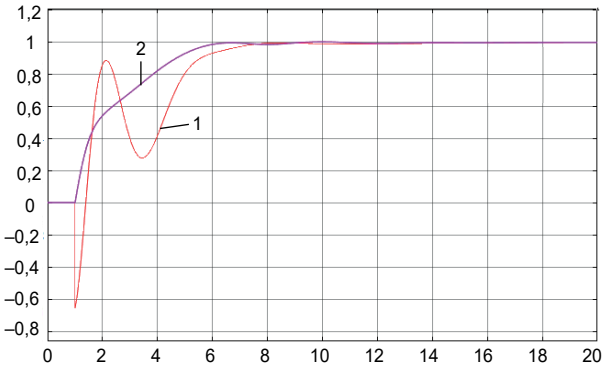
■ ***Table 2.*** Rules of the fuzzy system for the fuzzy controller of the researched ACS

№	Правило							
1	If	input1	is	очень-сильно-в-минусе	then	output1	is	очень-сильно-в-минусе
2	If	input1	is	сильно-в-минусе	then	output1	is	средне-в-минусе
3	If	input1	is	средне-в-минусе	then	output1	is	немного-в-минусе
4	If	input1	is	немного-в-минусе	then	output1	is	немного-в-минусе
5	If	input1	is	в-нуле	then	output1	is	в-нуле
6	If	input1	is	немного-в-плюсе	then	output1	is	в-нуле
7	If	input1	is	средне-в-плюсе	then	output1	is	немного-в-плюсе
8	If	input1	is	сильно-в-плюсе	then	output1	is	немного-в-плюсе
9	If	input1	is	очень-сильно-в-плюсе	then	output1	is	немного-в-плюсе



■ ***Рис. 6.*** Переходный процесс угла тангажа ϑ интеллектуальной САУ

■ ***Fig. 6.*** Transient process of pitch angle ϑ measurement intelligent ACS



■ ***Рис. 7.*** Переходные процессы классической *1* и интеллектуальной *2* САУ по каналу измерения угла тангажа ϑ

■ ***Fig. 7.*** Transient processes of classical *1* and intelligent *2* ACS through the pitch angle ϑ measurement channel

2. Интеллектуализация САУ позволила не только компенсировать флуктуации, но и снизить время установления рабочего режима. Для классической САУ оно составило $t_{\text{п}} \approx 7$ с, для интеллектуальной САУ $t_{\text{п}} \approx 5{,}2$ с.

3. Применение нечеткого регулятора хоть и позволило существенно снизить колебания переходного процесса, однако не исключило их полностью. Ввиду большого количества возмущающих воздействий в исследуемой САУ небольшие флуктуации наблюдаются до момента времени $t \approx 15$ с, после чего система окончательно стабилизируется.

4. Улучшение качества переходного процесса по каналу угла тангажа ϑ немного ухудшает качество переходных процессов по каналам угла атаки α и нормальной перегрузки n (увеличивает их время $t_{\text{п}}$, однако не нарушает устойчивость САУ по другим летным параметрам). Это связано с подключением нечеткого регулятора к общему каналу системы управления, формирующему качество переходных процессов всех параметров, измеряемых системой. Таким образом, оптимизация регулятора в целях улучшения качества переходного процесса по тангажу незначительно влияет на аналогичные показатели других летных параметров в вертикальной плоскости симметрии летательного аппарата.

Заключение

В ходе данного исследования было произведено моделирование работы системы управления угла тангажа в классическом исполнении и с использованием звена нечеткой регуляции. На основании проведенного сравнительного анализа двух систем был сделан вывод о целесообразности применения интеллектуального нечеткого регулятора в САУ параметров вертикальной плоскости симметрии для улучшения качества переходного процесса канала измерения угла тангажа ϑ.

Несмотря на основополагающие преимущества, которыми обладает нечеткий регулятор (лучшее качество переходного процесса, быстродействие), он не имеет универсального алгоритма настройки функций принадлежности. Эти функции формируются эмпирически: в текущем исследовании $\mu_N(x)$ и $\mu_M(y)$ были подобраны за 15–20 итераций. Этот показатель может быть выше для нечетких регуляторов, применяемых в более сложных САУ с еще большим числом возмущающих воздействий, что на практике может затруднить процесс адаптивной настройки системы управления. Поэтому целесообразно провести дополнительный анализ, связанный с применимостью нечетких регуляторов для локальных задач управления, имеющих конкретные начальные условия.

В рамках дальнейших исследований и усовершенствования рассмотренной в данной статье модели интеллектуальной САУ по углу тангажа предполагается:

– реализовать имитационное моделирование САУ по тангажу в случае стохастического входного воздействия и провести дополнительную оптимизацию звеньев системы для улучшения качества переходного процесса;

– для повышения качества характеристик измерения угла атаки α и нормальной перегрузки n_y провести дополнительную настройку системы по каналам других параметров;

– рассмотреть нелинейный режим работы системы управления, линеаризованной в ходе исследования, проведенного в данной статье;

– осуществить синтез САУ по тангажу с применением другого интеллектуального звена – нейросетевого регулятора, провести сравнительный анализ качества переходного процесса полученного и переходного процесса САУ с нечетким регулятором.

Литература

1. **Zhang X., Qian L.** Heading control of the air cushion vehicle based on fuzzy PID. *2022 Intern. Conf. on Cyber-Physical Social Intelligence (ICCSI)*, 2022. doi:10.1109/ICCSI55536.2022.9970669
2. **Титов Ю. К.** *Адаптивные нечеткие устройства систем управления с гарантированной устойчивостью*: дис. … канд. техн. наук. Пермь, ПНИПУ, 2019. 204 с.
3. **Княжский А. Ю.** *Разработка и исследование алгоритма управления движением низколетящего аппарата над неровной поверхностью, минимизирующего его среднюю высоту*: дис. … канд. техн. наук. СПб., СПбГУТ им. проф. Бонч-Бруевича, 2018. 135 с.
4. **Ахрамович С. А., Баринов А. В., Малышев В. В., Старков А. В.** Синтез системы управления беспилотного летательного аппарата по высоте методом бэкстеппинга. *Вестник Самарского университета. Аэрокосмическая техника, технологии и машиностроение*, 2018, т. 17, № 2, с. 7–22. doi:10.18287/2541-7533-2018-17-2-7-22
5. **Иванов М. Е., Езерский В. В.** Применение методов искусственного интеллекта в системе управления воздушным движением тяжелого многоцелевого экраноплана. *Волновая электроника и инфокоммуникационные системы: материалы 14-й Междунар. науч. конф.* В 3 ч. Ч. 2. СПб., 2021, с. 323–330. EDN: GXGBAX
6. **Kulik A.** Artificial intelligence-based aircraft accident threat parrying method. *Proc. of Telecommunication Universities*, 2021, vol. 7, no. 4, pp. 110–117. doi:10.31854/1813-324X2021-7-4-110-117, EDN: UWGRXJ
7. **Иванов М. Е., Езерский В. В.** Моделирование системы управления низколетящим аппаратом по каналу рысканья в Matlab Simulink. *3-я Междунар. науч. конф. «Аэрокосмическое приборостроение и эксплуатационные технологии»: сб. докл.*, Санкт-Петербург, 14–22 апреля 2022 г. СПб., 2022, с. 162–165. EDN: YINSAZ
8. **Кулида Е. Л., Лебедев В. Г.** Перспективы использования методов искусственного интеллекта в авиации. *Управление развитием крупномасштабных систем*: *тр. 13-й Междунар. конф.,* Москва, 28–30 сентября 2020 г. М., Институт проблем управления им. В. А. Трапезникова РАН, 2020, с. 1535–1541. EDN: OIJOYG
9. **Коновалов А. С., Шумилов П. Е.** Применение нечеткой логики в авиационных системах антиюзовой автоматики. *Информационно-управляющие системы*, 2003, № 5, с. 12–17. EDN: KXZXCB
10. **Гридина Н. В., Евдокимов И. А., Солодовников В. И.** Построение гибридных нейронных сетей с использованием элементов нечеткой логики. *Искусственный интеллект и принятие решений,* 2019, № 2, с. 91–97. doi:10.14357/20718594190209, EDN: ZUOMUP
11. **Пушкарев Д. О., Киселев Д. Ю., Киселев Ю. В.** Расчет авиационных рычажно-поплавковых клапанов с помощью нейронных сетей. *Вестник Самарского университета. Аэрокосмическая техника, технологии и машиностроение*, 2022, т. 21, № 4, с. 44–51. doi:10.18287/2541-7533-2022-21-4-44-51
12. **Шилоносов А. В., Молдован Д. В.** Динамическая коррекция управляющих воздействий системы управления сложного технического объекта с использованием нейросетевых технологий. *Альманах Пермского военного института войск национальной гвардии*, 2021, № 4 (4), с. 415–422. EDN: SYZWBS
13. **Петунин В. И.** Синтез законов управления канала тангажа автопилота. *Вестник Уфимского государственного авиационного технического университета*, 2007, т. 9, № 2 (20), с. 25–31. EDN: IBJMNB
14. **Абдуллина Э. Ю., Ефанов В. Н.** Синтез системы управления углом тангажа с каналом ограничения угла атаки. *Известия высших учебных заведений. Авиационная техника*, 2020, № 1, с. 25–32. EDN: IKOPOO
15. *Управление углом тангажа ЛА*. https://helpiks.org/5-70621.html (дата обращения: 13.10.2015).
16. **Бураков М. В., Коновалов А. С.** Синтез нечетких логических регуляторов. *Информационно-управляющие системы*, 2011, № 1, с. 22–27.
17. **Бураков М. В., Коновалов А. С.** Нечеткий супервизор ПИД-регулятора. *Информационно-управляющие системы*, 2018, № 5, с. 13–21. doi:10.31799/1684-8853-2018-5-13-21
18. **Althubaiti M., Bernard M., Musilek P.** Fuzzy logic controller for hybrid renewable energy system with multiple types of storage. *2017 IEEE 30th Canadian Conf. on Electrical and Computer Engineering (CCECE)*, 2017. doi:10.1109/CCECE.2017.7946738
19. **Чертилин К. Э., Ивченко В. Д.** Настройка адаптивных пропорционально-интегрально-дифференциальных регуляторов системы автоматического регулирования частоты вращения газотурбинного двигателя. *Российский технологический журнал*, 2020, т. 8, № 6 (38), с. 143–156. doi:10.32362/2500-316X-2020-8-6-143-156. https://rtj.mirea.ru/ (дата обращения: 13.01.2023).
20. **Демидова Г. Л.** *Разработка и исследование регуляторов с нечеткой логикой для следящих электроприводов оптико-механических комплексов*: дис. … канд. техн. наук. СПб., НИУ ИТМО, 2016. 210 с.
21. **Андриевская Н. В., Билоус О. А., Семенов С. В.** Методика проектирования нечеткого регулятора на базе ПИ-регулятора в среде MATLAB. *Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления*, 2012, № 6, с. 282–287. EDN: PWNZBX

UDC 621.396
doi:10.31799/1684-8853-2023-3-51-58
EDN: DACXFO

The use of the intelligent fuzzy controller for the elimination of fluctuations in the transient process of the perturbed automatic pitch angle control system of an aircraft

M. E. Ivanov[a], Post-Graduate Student, orcid.org/0000-0002-7584-6946
V. V. Yezerskiy[a], Dr. Sc., Phys.-Math., Professor, orcid.org/0000-0002-1223-6841, ezerskij2010@yandex.ru
[a]Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation

Introduction: The system of automatic control of the pitch angle of an aircraft is affected by numerous disturbing effects, which results in significant fluctuations in the transient process along the pitch of this system. **Purpose:** To transform the pitch angle control scheme and develop a fuzzy controller for it which will reduce the fluctuations of the transient process and improve the quality of the transient process. **Results:** The simulation using Matlab Simulink of the operation of the classical and modernized pitch angle control systems was carried out. An improvement in the quality of the transient process of the pitch angle control system has been achieved due to a significant reduction in fluctuations. To do this, a proportional–integrating controller has been added to the control system scheme using one of the methods of artificial intelligence — fuzzy logic. The adjustment of the fuzzy controller is carried out by converting a typical algorithm: for input linguistic variables we halve the interval of the universal set on which the membership functions are defined. **Practical relevance:** The correct adjustment of the fuzzy controller can significantly improve the quality of the transient process of the control system, which is affected by a large number of disturbing effects.

Keywords — fluctuation, disturbing effect, artificial intelligence, automatic control system, fuzzy controller, transient process.

For citation: Ivanov M. E., Yezerskiy V. V. The use of the intelligent fuzzy controller for the elimination of fluctuations in the transient process of the perturbed automatic pitch angle control system of an aircraft. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 51–58 (In Russian). doi:10.31799/1684-8853-2023-3-51-58, EDN: DACXFO

References

1. Zhang X., Qian L. Heading control of the air cushion vehicle based on fuzzy PID. *2022 Intern. Conf. on Cyber-Physical Social Intelligence (ICCSI)*, 2022. doi:10.1109/ICCSI55536.2022.9970669
2. Titov Yu. K. *Adaptivnye nechetkie ustrojstva sistem upravleniya s garantirovannoj ustojchivost'yu*. Dis. kand. tech. nauk [Adaptive fuzzy control system evices with guaranteed stability. Cand. tech. sci. diss.]. Perm, PNIPU Publ., 2019. 204 p. (In Russian).
3. Knyazhskij A. Yu. *Razrabotka i issledovanie algoritma upravleniya dvizheniem nizkoletyashhego apparata nad nerovnoj poverxnost`yu, minimiziruyushhego ego srednyuyu vysotu*. Dis. kand. tech. nauk [Development and research of an algorithm for controlling the movement of a low-flying vehicle over an uneven surface, minimizing its average height. Cand. tech. sci. diss.] Saint-Petersburg, SPbGUT im. prof. Bonch-Bruevicha Publ., 2018. 135 p. (In Russian).
4. Akhramovich S. A., Barinov A. V., Malyshev V. V., Starkov A. V. Backstepping synthesis of the height control system of an unmanned aerial vehicle. *Vestnik of Samara University. Aerospace and Mechanical Engineering*, 2018. vol. 17, no. 2, pp. 7–22 (In Russian). doi:10.18287/2541-7533-2018-17-2-7-22
5. Ivanov M. E., Yezersky V. V. Application of artificial intelligence methods in the air traffic control system of a heavy multipurpose WIG-craft. *Materialy 14-j Mezhdunarodnoj nauchnoj konferencii "Volnovaya elektronika i infokommunikacionnye sistemy". Chast' 2* [Proc. of the 14th Intern. Scientific Conf. "Wave electronics and infocommunication systems". Part 2]. Saint-Petersburg, 2021, pp. 323–330 (In Russian). EDN: GXGBAX
6. Kulik A. Artificial intelligence-based aircraft accident threat parrying method. *Proc. of Telecommunication Universities*, 2021, vol. 7, no. 4, pp. 110–117. doi:10.31854/1813-324X2021-7-4-110-117, EDN: UWGRXJ
7. Ivanov M. E., Yezersky V. V. Modeling the operation of an automatic control system for a low-flying vehicle along the yaw channel in Matlab Simulink. *Sbornik dokladov tret'ej Mezhdunarodnoj nauchnoj konferencii "Aerokosmicheskoe priborostroenie i ekspluatacionnye tekhnologii"* [Proc. of the 3rd Intern. Scient. Conf. "Aerospace instrumentation and operational technologies"]. Saint-Petersburg, 2022, pp. 162–165 (In Russian). EDN: YINSAZ
8. Kulida E. L., Lebedev V. G. Prospects for the use of artificial intelligence methods in aviation. *Trudy 13-j Mezhdunarodnoj konferencii "Upravlenie razvitiem krupnomasshtabnyh system"* [Proc. 13th Intern. Conf. "Managing the development of large-scale systems"]. Moscow, 2020, pp. 1535–1541 (In Russian). EDN: OIJOYG
9. Konovalov A. S., Shumilov P. E. Application of fuzzy logic in aircraft antilock braking systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2003, no. 5, pp. 12–17 (In Russian). EDN: KXZXCB
10. Gridina N. V., Evdokimov I. A., Solodovnikov V. I. Construction of hybrid neural networks using fuzzy logic elements. *Scientific and Technical Information Processing*, 2019, no. 2, pp. 91–97 (In Russian). doi:10.14357/20718594190209, EDN: ZUOMUP
11. Pushkarev D. O., Kiselev D. Yu., Kiselev Yu. V. Calculation of aircraft lever-float valves using neural networks. *Vestnik of Samara University. Aerospace and Mechanical Engineering*, 2022, vol. 21, no. 4, pp. 44–51 (In Russian). doi:10.18287/2541-7533-2022-21-4-44-51
12. Shilonosov A. V., Moldovan D. V. Dynamic correction of control actions of the control system of a complex technical object using neural network technologies. *Al'manah Permskogo voennogo instituta vojsk nacional'noj gvardii*, 2021, no. 4, pp. 415–422 (In Russian). EDN: SYZWBS
13. Petunin V. I. Synthesis of control laws of a pitch channel of the autopilot. *Vestnik Ufimskogo gosudarstvennogo aviacionnogo tekhnicheskogo universiteta*, 2007, vol. 9, no. 2, pp. 25–31 (In Russian). EDN: IBJMNB
14. Abdullina E. Y., Efanov V. N. Synthesis of pitch angle control system with angle of attack limiting channel. *Russian Aeronautics*, 2020, vol. 63, no. 1, pp. 25–32 (In Russian).
15. *Control of the pitch angle of the aircraft*. Available at: https://helpiks.org/5-70621.html (accessed 13 October 2015).
16. Burakov M. V., Konovalov A. S. Fuzzy controllers design. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2011, no. 1, pp. 22–27 (In Russian).
17. Burakov M. V., Konovalov A. S. Fuzzy supervisor for PID controller. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2018, no. 5, pp. 13–21 (In Russian). doi:10.31799/1684-8853-2018-5-13-21
18. Althubaiti M., Bernard M., Musilek P. Fuzzy logic controller for hybrid renewable energy system with multiple types of storage. *2017 IEEE 30th Canadian Conf. on Electrical and Computer Engineering (CCECE)*, 2017. doi:10.1109/CCECE.2017.7946738
19. Chertilin K. E., Ivchenko V. D. Configuring adaptive PID-controllers of the automatic speed control system of the GTE. *Russian Technological Journal*, 2020, vol. 8, no. 6, pp. 143–156 (In Russian). doi:10.32362/2500316X-2020-8-6-143-156. Available at: https://rtj.mirea.ru/ (accessed 13 January 2023).
20. Demidova G. L. *Razrabotka i issledovanie regulyatorov s nechetkoj logikoj dlya sledyashhix elektroprivodov optiko-mexanicheskix kompleksov*. Dis. kand. tech. nauk [Development and research of fuzzy logic controllers for optical-mexanicheskix kompleksov tracking electric drives. Cand. tech. sci. diss.] Saint-Petersburg, ITMO Publ., 2016. 210 p. (In Russian).
21. Andrievskaya N. V., Bilous O. A., Semenov S. V. The methodology of designing a fuzzy controller based on a PI-controller in the Matlab environment. *Bulletin of Perm National Research Polytechnic University. Electrotechnics, Informational Technologies, Control Systems*, 2012, no. 6, pp. 282–287 (In Russian). EDN: PWNZBX

UDC 003.26
doi:10.31799/1684-8853-2023-3-59-69
EDN: GXPTKZ

Articles

Post-quantum signature algorithms with a hidden group and doubled verification equation

A. A. Moldovyan[a], *Dr. Sc., Tech., Professor, Chief Researcher, orcid.org/0000-0001-5480-6016*
N. A. Moldovyan[a], *Dr. Sc., Tech., Professor, Chief Researcher, orcid.org/0000-0002-4483-5048, nmold@mai.ru*
[a]*St. Petersburg Federal Research Center of the RAS, 39, 14th Line, 199178, Saint-Petersburg, Russian Federation*

Introduction: *One of the current topical problems of cryptography is the development of post-quantum digital signature algorithms with relatively small sizes of the public key and signature.* ***Purpose:*** *To develop a new method for designing post-quantum algebraic signature algorithms with a hidden group, based on the computational complexity of solving large systems of quadratic multivariate equations, which allows to reduce the size of the public key and signature as compared to the known analogues.* ***Results:*** *We propose a new method for designing digital signature algorithms with a signature of the form (e, S), where e is a natural number (randomization parameter) and S is a vector (fitting parameter. The method makes it possible to reduce the dimension of finite non-commutative associative algebras used as an algebraic support. The method is distinguished by the use of the technique of doubling the verification equation for fixing the hidden group, which allows one to set the formation of the vector S depending on the random reversible vector and thereby eliminates the influence of the number of signed documents on the security, which is typical of the known analogous algorithms. The method has been tested by the development of a specific post-quantum signature algorithm, various modifications of which use algebras of different dimensions. A preliminary security assessment of the proposed algorithm has been performed.* ***Practical relevance:*** *Due to comparatively small sizes of signature and public key, the introduced signature algorithm represents significant practical interest as a prototype of a post-quantum signature standard.*

Keywords *— post-quantum crypto schemes, computer security, digital signature, discrete logarithm problem, cryptography, finite non-commutative algebras, associative algebras, commutative groups.*

For citation: Moldovyan A. A., Moldovyan N. A. Post-quantum signature algorithms with a hidden group and doubled verification equation. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 3, pp. 59–69. doi:10.31799/1684-8853-2023-3-59-69, EDN: GXPTKZ

Introduction

Development of post-quantum standards on public-key cryptographic algorithms is one of current challenges faced by the global cryptographic community [1, 2]. This challenge is due to the fact that modern standards for public key cryptographic algorithms are based on the computational difficulty of the discrete logarithm problem and of the factoring problem, each of which can be solved on a quantum computer in polynomial time [3, 4].

Post-quantum public-key cryptalgorithms are to be based on computationally difficult problems different from the said two ones. One can mention post-quantum algorithms on codes [5, 6], and on hash functions [7]. Regarding post-quantum digital signature algorithms, their main disadvantage is the large total size of the public key and signature. To overcome this shortcoming, signature schemes with a hidden group were proposed in [8–11], using computational complexity of so called hidden discrete logarithm problem (HLP). However, some of such signature schemes (for example, introduced in [8, 9]) are vulnerable against algebraic attacks [12].

A new paradigm for the development of algorithms with a hidden group has been proposed in [13, 14]. That paradigm exploits the computational difficulty of finding a solution to a large system of quadratic multivariate equations with many unknowns. The latter problem is considered as an attractive post-quantum primitive [15, 16]. It had been put into the base of security of multivariate cryptographic algorithms, like EFLASH [17], MQDSS [18], Rainbow [19, 20], GeSMM [21], and UOV [22]. However, the multivariate cryptographic algorithms have a very significant drawback, which consists in the extremely large size of the public key (up to several hundred kilobytes (several megabytes) for the case of 128-bit (256-bit) security [20, 21]).

The algebraic algorithms with a hidden group exploiting computational complexity of solving a large system of quadratic multivariate equations provides possibility to develop signature algorithms with small size of both the public key and the signature. In algorithms of this type, a digital signature is a pair of values (e, $\mathbf{S}$), the number e and the vector $\mathbf{S}$. For example, the signature algorithm [13] uses a collision-resistant hash function $f(\cdot)$ and the next verification equation with two entries of the vector $\mathbf{S}$:

$$\mathbf{R}^* = \left(\mathbf{Y}_1\mathbf{S}\mathbf{Z}_1\right)^e \left(\mathbf{Y}_2\mathbf{S}\mathbf{Z}_2\right)^{e^2}.$$

In the latter equation the vectors $\mathbf{Y}_1$, $\mathbf{Z}_1$, $\mathbf{Y}_2$, $\mathbf{Z}_2$ are elements of the public key, which are calculat-

ed as follows: $\mathbf{Y}_1 = \mathbf{A}\mathbf{G}^u\mathbf{B}$; $\mathbf{Z}_1 = \mathbf{C}\mathbf{J}\mathbf{A}^{-1}$; $\mathbf{Y}_2 = \mathbf{A}\mathbf{J}^w\mathbf{B}$; $\mathbf{Z}_2 = \mathbf{C}\mathbf{G}\mathbf{A}^{-1}$ (when performing a direct attack against the algorithm, these four formulas define four vector quadratic equations with the unknown vectors $\mathbf{G}^u$, $\mathbf{J}$, $\mathbf{J}^w$, and $\mathbf{G}$), where u, w are private integers; $\mathbf{G}$, $\mathbf{J}$, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are private reversible vectors such that $\mathbf{GJ} = \mathbf{JG}$ and $\mathbf{G}$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are pairwise non-permutable vectors. The signature (e, $\mathbf{S}$) is valid, if $f(M||\mathbf{R}^*) = e$, where M is a signed document and || denotes the concatenation operation. Note the verification equation connects the signature with the public key and the latter equality connects both the signature and the public key with the document M. The use of verification equations with two [13] and multiple [14, 23] entries of the vector $\mathbf{S}$ in the verification equation is focused on increasing resistance to attacks using $\mathbf{S}$ as a fitting parameter.

Unfortunately, for a fixed value of the public key, the number of different vectors that can potentially be a signature element is limited by the order of the hidden commutative group, since the vector $\mathbf{S}$ is computed by the formula $\mathbf{S} = \mathbf{B}^{-1}\mathbf{G}^n\mathbf{J}^r\mathbf{C}^{-1}\mathbf{X}$ [13], where integers n and r are computed depending on the document to be signed; vectors $\mathbf{G}$ and $\mathbf{J}$ compose a minimum generator system of the commutative hidden group; vectors $\mathbf{B}$ and $\mathbf{C}$ are elements of the private key. It can be shown that, given five (or more) different signatures, it is possible to compose a system of five vector quadratic equations, which can be solved independently of other secret values. The latter provides a significant reduction in the complexity of solving a system of equations that connects the elements of the public key with the elements of the secret key.

Therefore, to ensure the required security level of the algorithm from [13], it is required to use finite non-commutative associative algebras (FNAAs) of a sufficiently large dimension m ($m \geq 10$ for 128-bit security and $m \geq 20$ for 256-bit security), which limits the possibility of reducing the size of the public key and signature and increasing performance.

This article proposes a new method for developing algebraic signature algorithms with a hidden group, which, for a given security level, provides possibility to use FNAAs of comparatively small dimension as algebraic support. A new post-quantum signature algorithm is introduced as implementation of the method. The signature has the form (e, $\mathbf{S}$) and the indicated restriction is eliminated due to the fact that the signature element $\mathbf{S}$ can take any value in the FNAA used as an algebraic support. To insure such possibility, the technique of doubling the verification equation is used for setting the hidden group. Previously [24], this technique was used as a way to define a hidden group in signature algorithms on finite commutative algebras, which are based on computational difficulty of the HLP. Earlier, such technique was used by authors to implement the post-quantum resistance criterion, when designing the signature algorithms on FNAAs, based on the HLP [9, 24]. The introduced method is illustrated by a signature algorithm in which two different hash functions are used as an auxiliary technique for providing security against attacks using the signature element $\mathbf{S}$ as a fitting element.

Preliminaries

The technique of doubling the verification equation (see, for example, [9]) consists in specifying two different equations defining computation of the next two vectors $\mathbf{R}_1$ and $\mathbf{R}_2$, depending on the signature ($\mathbf{S}$, e) and the public key ($\mathbf{P}_{11}$, $\mathbf{P}_{12}$, …, $\mathbf{P}_{1b}$, $\mathbf{P}_{21}$, $\mathbf{P}_{22}$, …, $\mathbf{P}_{2d}$):

$$\mathbf{R}_1 = f_1(\mathbf{P}_{11}, \mathbf{P}_{12}, \ldots, \mathbf{P}_{1b}, \mathbf{S}, e);$$
$$\mathbf{R}_2 = f_1(\mathbf{P}_{21}, \mathbf{P}_{22}, \ldots, \mathbf{P}_{2d}, \mathbf{S}, e),$$

where b and d are some small natural numbers; the vectors $\mathbf{P}_{11}$, $\mathbf{P}_{12}$, …, $\mathbf{P}_{1b}$, $\mathbf{P}_{21}$, $\mathbf{P}_{22}$, …, $\mathbf{P}_{2d}$ are elements of the public key.

And then, using a collision-resistant hash function $f(\cdot)$, connection of the signature and public key with an electronic document M is confirmed by checking the validity of the equality $f(M||\mathbf{R}_1||\mathbf{R}_2) = e$.

In the introduced algorithm with a hidden group, FNAAs are used as algebraic support. An m-dimensional FNAAs is defined as a finite vector space (defined over a field $GF(p)$) with the non-commutative associative multiplication operation of the vectors $\mathbf{A} = a_0\mathbf{e}_0 + a_1\mathbf{e}_1 + \ldots + a_{m-1}\mathbf{e}_{m-1}$ and $\mathbf{B} = b_0\mathbf{e}_0 + b_1\mathbf{e}_1 + \ldots + b_{m-1}\mathbf{e}_{m-1}$ (where $\mathbf{e}_i$, $i = 0, 1, \ldots, m-1$, are basis vectors), defined by the next formula:

$$\mathbf{AB} = \sum_{i=0}^{m-1}\sum_{j=0}^{m-1} a_i b_j \mathbf{e}_i \mathbf{e}_j, \qquad (1)$$

in which the values a_i and b_i are multiplied as the field elements. Using a basis vector multiplication table (BVMT), every product $\mathbf{e}_i\mathbf{e}_j$ is replaced by a one-component vector (see more details in [13, 24]).

Table 1 sets a 4-dimensional FNAA with the two-sided global unit $\mathbf{E} = (1, 1, 0, 0)$, for which the decomposition into commutative subalgebras is studied well in [10]:

1) the said algebra contains $p^2 + p + 1$ commutative subalgebras having order p^2; every non-scalar vector is contained in a unique subalgebra; every scalar vector is included in all subalgebras, i. e., the latter intersect exactly in the set of scalar vectors $\{\mathbf{L}: \mathbf{L} = h\mathbf{E}, h = 0, 1, \ldots, p-1\}$;

2) the multiplicative group Γ of the FNAA has order equal to

$$\Omega = p(p^2 - 1)(p - 1); \qquad (2)$$

3) the group Γ includes $p(p + 1)/2$ commutative subgroups Γ_1 possessing 2-dimensional cyclicity (i. e., a minimum generator system of the subgroup Γ_1 contains two vectors of the same order) and having order equal to

$$\Omega_1 = (p - 1)^2; \quad (3)$$

4) the group Γ includes $p(p - 1)/2$ commutative cyclic subgroups Γ_2 of the order

$$\Omega_2 = p^2 - 1 = (p - 1)(p + 1); \quad (4)$$

5) the group Γ includes $p + 1$ commutative cyclic subgroups Γ_3 of the order

$$\Omega_3 = p(p - 1). \quad (5)$$

The vector $\mathbf{G} = g_0\mathbf{e}_0 + g_1\mathbf{e}_1 + g_2\mathbf{e}_2 + g_3\mathbf{e}_3$, such that $g_2 \neq 0$, and $g_3 \neq 0$, determines the commutative subalgebra described as the next set of vectors $\mathbf{X}$ [10], which includes $\mathbf{G}$:

$$\mathbf{X} = \left(x_0, x_2, x_3, x_4\right) = \left(d,\ d + k\frac{g_1 - g_0}{g_2},\ k,\ k\frac{g_3}{g_2}\right), \quad (6)$$

where $d, k = 0, 1, \ldots, p - 1$. Type of the commutative group including all reversible vectors of the set (6) depends on the value of

$$\Delta = (g_0 - g_1)^2 + 4\lambda\, g_2 g_3.$$

If $\Delta \neq 0$ is a quadratic residue in $GF(p)$, then the multiplicative group of the latter subalgebra relates to the Γ_1-type groups and has order $(p - 1)^2$ [10]. Note that the probability of a vector $\mathbf{G}$ selected at random from the set (6) is equal to ≈ 0.5 (i. e., to the probability that the value $\Delta \neq 0$ is a quadratic residue).

It is assumed that, depending on the required level of security, FNAAs of different dimensions m will be used as algebraic support of the developed algorithm. For the case $m \geq 6$, the FNAAs are set with using the next formula for generating the BVMTs for arbitrary even dimensions [25]:

■ ***Table 1.*** Defining non-commutative associative multiplication of 4-dimensional vectors ($\lambda \neq 0$) [10]

∘	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e}_0$	$\mathbf{e}_0$	0	0	$\mathbf{e}_3$
$\mathbf{e}_1$	0	$\mathbf{e}_1$	$\mathbf{e}_2$	0
$\mathbf{e}_2$	$\mathbf{e}_2$	0	0	$\lambda\mathbf{e}_1$
$\mathbf{e}_3$	0	$\mathbf{e}_3$	$\lambda\mathbf{e}_0$	0

$$\mathbf{e}_i\mathbf{e}_j = \begin{cases} \mathbf{e}_{i+j \bmod m}, & \text{if } i \bmod 2 = 0; \\ \mathbf{e}_{i-j \bmod m}, & \text{if } i \bmod 2 = 1,\ j \bmod 2 = 0; \\ \lambda\mathbf{e}_{i-j \bmod m}, & \text{if } i \bmod 2 = 1,\ j \bmod 2 = 1, \end{cases}$$

where $i, j \in \{0, 1, \ldots, m - 1\}$. The latter formula allows to construct BVMTs setting FNAAs with global two-sided unit of the form $\mathbf{E} = (1, 0, 0, \ldots, 0)$. Tables 2 and 3 shows the BVMTs for the cases $m = 6$ and $m = 8$. In framework of this article, we consider the FNAAs set over the ground finite field $GF(p)$ with odd characteristic p (such that $p = 2q + 1$, where p and q are prime numbers). However, the method by [25] can be also used for setting the FNAAs over the finite fields of characteristic two.

The condition for the vector $\mathbf{A} = (a_0, a_1, a_2, a_3, a_4, a_5)$ to be reversible is the following non-equality [25]:

$$\frac{1}{4}\left(\left(a_0 + a_2 + a_4\right)^2 - \lambda\left(a_1 + a_3 + a_5\right)^2\right) \times$$
$$\times \left(\left(a_0 - a_2\right)^2 + \left(a_0 - a_4\right)^2 + \left(a_2 - a_4\right)^2 - \right.$$
$$\left. -\lambda\left(a_1 - a_3\right)^2 - \lambda\left(a_1 - a_5\right)^2 - \lambda\left(a_3 - a_5\right)^2\right)^2 \neq 0. \quad (7)$$

Finding a reversible 6-dimensional vector $\mathbf{A}$ can be done by generating at random six of its coordinates and checking the validity of inequality (7). For the cases $m \geq 8$ there are no formulas for the condition of vector reversibility and the following method for generation of reversible vectors can be applied: generate random vectors $\mathbf{B}$ until $\mathbf{B}^{p^2-1} = \mathbf{E}$ or $\mathbf{B}^{(p-1)^2} = \mathbf{E}$.

In the proposed signature algorithm, we use the primes p having the size $|p| = 80$ and $|p| = 128$ bits (the corresponding values of $|q|$ are equal to 79 and 127 bits). We also use two collision-resistant hash functions, the usual type $f(\cdot)$ and the vector type $\mathbf{H}(\cdot)$. The hash function of the latter type take on the values in the FNAA used as algebraic support.

■ ***Table 2.*** Defining non-commutative associative multiplication of the 6-dimensional vectors ($\lambda \neq 0$) [25]

∘	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$
$\mathbf{e}_0$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$
$\mathbf{e}_1$	$\mathbf{e}_1$	$\lambda\mathbf{e}_0$	$\mathbf{e}_5$	$\lambda\mathbf{e}_4$	$\mathbf{e}_3$	$\lambda\mathbf{e}_2$
$\mathbf{e}_2$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_0$	$\mathbf{e}_1$
$\mathbf{e}_3$	$\mathbf{e}_3$	$\lambda\mathbf{e}_2$	$\mathbf{e}_1$	$\lambda\mathbf{e}_0$	$\mathbf{e}_5$	$\lambda\mathbf{e}_4$
$\mathbf{e}_4$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e}_5$	$\mathbf{e}_5$	$\lambda\mathbf{e}_4$	$\mathbf{e}_3$	$\lambda\mathbf{e}_2$	$\mathbf{e}_1$	$\lambda\mathbf{e}_0$

■ ***Table 3.*** Defining the 8-dimensional FNAA ($\lambda \neq 0$)

$\circ$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$	$\mathbf{e}_7$
$\mathbf{e}_0$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$	$\mathbf{e}_7$
$\mathbf{e}_1$	$\mathbf{e}_1$	$\lambda\mathbf{e}_0$	$\mathbf{e}_7$	$\lambda\mathbf{e}_6$	$\mathbf{e}_5$	$\lambda\mathbf{e}_4$	$\mathbf{e}_3$	$\lambda\mathbf{e}_2$
$\mathbf{e}_2$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$	$\mathbf{e}_7$	$\mathbf{e}_0$	$\mathbf{e}_1$
$\mathbf{e}_3$	$\mathbf{e}_3$	$\lambda\mathbf{e}_2$	$\mathbf{e}_1$	$\lambda\mathbf{e}_0$	$\mathbf{e}_7$	$\lambda\mathbf{e}_6$	$\mathbf{e}_5$	$\lambda\mathbf{e}_4$
$\mathbf{e}_4$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$	$\mathbf{e}_7$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e}_5$	$\mathbf{e}_5$	$\lambda\mathbf{e}_4$	$\mathbf{e}_3$	$\lambda\mathbf{e}_2$	$\mathbf{e}_1$	$\lambda\mathbf{e}_0$	$\mathbf{e}_7$	$\lambda\mathbf{e}_6$
$\mathbf{e}_6$	$\mathbf{e}_6$	$\mathbf{e}_7$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$
$\mathbf{e}_7$	$\mathbf{e}_7$	$\lambda\mathbf{e}_6$	$\mathbf{e}_5$	$\lambda\mathbf{e}_4$	$\mathbf{e}_3$	$\lambda\mathbf{e}_2$	$\mathbf{e}_1$	$\lambda\mathbf{e}_0$

To compute $\mathbf{H}(M)$ from a document M, other hash function of the usual type $f'(\cdot)$ is used to calculate coordinates of the vector $\mathbf{H} = (h_0, h_1, \ldots, h_{m-1})$. The size of the values of the usual type hash functions is equal to $2|p|$. Calculation of the value of $\mathbf{H}(M)$ is performed as follows:

1. Compute the value $h'_{(0)} = h'_{(0)1}||h'_{(0)2} = f'(M)$ represented as concatenation of two $|p|$-bit values.

2. Calculate the values $h'_{(j)} = h'_{(j)1}||h'_{(j)2} = f'(h'_{(j-1)})$ for $j = 1, \ldots, (m-2)/2$.

3. Set values of the coordinates $h_{2i} = h'_{(i)1}$ and $h_{2i+1} = h'_{(i)2}$, where $i = 0, 1, \ldots, (m-2)/2$.

Setting the hidden group and formation of the public key

A commutative group contained in the FNAA used as algebraic support can serve as a hidden group in the proposed method and algorithm. However, we will consider the case of setting the hidden group with two-dimensional cyclicity, i. e., the group containing the minimum generator system $<\mathbf{G}, \mathbf{J}>$ including two vectors of the order $p-1$ (in the case $m = 4$, this is the Γ_1-type group).

Algorithm 1: Setting a hidden group possessing two-dimensional cyclicity in the case $m \geq 4$.

1. Generate a random reversible vector $\mathbf{G}$ of order $p-1$.

2. If $\mathbf{G}$ is a scalar vector, then go to step 1.

3. Generate at random an integer k ($k < p-1$) and a primitive element β in $GF(p)$.

4. Compute the vector $\mathbf{J} = \beta\mathbf{G}^k$.

5. Output the pair $\mathbf{J}$ and $\mathbf{G}$ as a basis $<\mathbf{G}, \mathbf{J}>$ of the hidden group $\Gamma_{<\mathbf{G},\mathbf{J}>}$.

In the case $m = 4$ the efficiency of Algorithm 1 can be estimated as follows. Taking into account that the number of the Γ_1-type groups in the 4-dimensional FNAA used as algebraic support is equal to $p(p+1)/2$ [10], one can estimate that the number of the vectors contained in all groups of the Γ_1-type is equal to $\approx p^4/2$. Therefore, a random vector $\mathbf{G}$ is a non-scalar vector and is contained in a group of the Γ_1-type with probability ≈ 0.5. For the case of prime $p = 2q + 1$ (where q is also a prime), one can easily show that a fixed Γ_1 group contains $\approx q^2$ of vectors of the order q and $\approx 3q^2$ of vectors of the order $p-1$, i. e., a random vector from the fixed Γ_1 group has order $p-1$ with the probability $\approx 3/4$. Thus, a random vector $\mathbf{G}$ passes the steps 1 and 2 with the probability $\approx 3/8$ and generation of the required vector $\mathbf{G}$ requires on the average performing steps 1 and 2 about 8/3 times.

Taking into account that step 3 is performed on the average two times, one can conclude that generation of a random basis $<\mathbf{G}, \mathbf{J}>$ of the hidden group $\Gamma_{<\mathbf{G},\mathbf{J}>}$ requires on the average performing Algorithm 1 less than 3 times.

Estimate of the efficiency of Algorithm 1 for the cases $m > 4$ requires using the detailed information about decomposition of the corresponding FNAAs into commutative subalgebras. Because of the lack of such information we have experimentally found the average number ψ of performing Algorithm 1 for generating the base $<\mathbf{G}, \mathbf{J}>$ of the hidden group with two-dimensional cyclicity. For every of the cases $m = 6$, $m = 8$, and $m = 10$, we have get $\psi < 5$ for different values of structural constant λ and of prime p. Thus, Algorithm 1 has acceptable efficiency, since it is intended to be used only at the stage of forming the public key.

Algorithm 2: Alternative procedure for setting a hidden group of the Γ_1 type in the case $m = 4$.

1. Generate a random reversible vector $\mathbf{G} = (g_0, g_1, g_2, g_3)$ such that $g_2 \neq 0$ and $g_3 \neq 0$ and compute the value of $\Delta = (g_0 - g_1)^2 + 4\lambda g_2 g_3$.

2. If $\Delta = 0$ or Δ is a quadratic non-residue in $GF(p)$, then go to step 1.

3. Generate two random integers $d \neq 0$ and $k \neq 0$ and, using formula (6), compute the vector $\mathbf{T}$.

4. If the order of $\mathbf{T}$ is not equal to $p-1$, then go to step 3.

5. Generate a random integer k ($0 < k < p-1$) and a random primitive element β in $GF(p)$.

6. Compute the vector $\mathbf{J} = \beta\mathbf{T}^k$.

7. Output the pair $\mathbf{J}$ and $\mathbf{G} = \mathbf{V}$ as a basis $<\mathbf{G}, \mathbf{J}>$ of the hidden group $\Gamma_{<\mathbf{G},\mathbf{J}>}$.

The probability that a random vector $\mathbf{G}$ defines the value of $\Delta \neq 0$ that is a quadratic residue is equal to ≈ 0.5, therefore the steps 1 and 2 are performed on the average two times. Like in the case of Algorithm 1, one can easily show the probability that a random vector $\mathbf{T}$ has order $p-1$ is equal to $\approx 3/4$. Therefore the steps 3 and 4 are performed on the average $\approx 4/3$ times. Like in the case of Algorithm 1, step 5 is performed on the average two times. Thus, genera-

tion of a random basis <**G**, **J**> requires performing Algorithm 2 approximately two times.

The vectors from the hidden group are used to calculate two parts of the public key. The elements of the first part are included in the first verification equation, and the elements of the second part are included in the second verification equation. The vector hash function and two similar verification equations, performed on the same signature value (e, **S**), are used to prevent signature-fitting attacks.

The next procedure for generating the public key has been developed:

1. Generate at random a hidden group $\Gamma_{<\mathbf{G},\mathbf{J}>}$.

2. Generate at random reversible vectors **A**, **B**, **F**, and **P** such that $\mathbf{AG} \neq \mathbf{GA}$, $\mathbf{AB} \neq \mathbf{BA}$, $\mathbf{AF} \neq \mathbf{FA}$, $\mathbf{AP} \neq \mathbf{PA}$, $\mathbf{BG} \neq \mathbf{GB}$, $\mathbf{BF} \neq \mathbf{FB}$, $\mathbf{BP} \neq \mathbf{PB}$, $\mathbf{FG} \neq \mathbf{GF}$, $\mathbf{FP} \neq \mathbf{PF}$, and $\mathbf{PG} \neq \mathbf{GP}$ (for random vectors **A**, **B**, **F**, and **P** these ten inequalities holds true with a high probability).

3. Calculate the public key elements $\mathbf{Y}_1$, $\mathbf{U}_1$, $\mathbf{Y}_2$, and $\mathbf{U}_2$:

$$\mathbf{Y}_1 = \mathbf{AGA}^{-1}, \mathbf{U}_1 = \mathbf{BJB}^{-1}, \mathbf{Y}_2 = \mathbf{FGF}^{-1}, \text{ and } \mathbf{U}_2 = \mathbf{PJP}^{-1}. \quad (8)$$

4. Generate at random a reversible vector **D** and hidden group elements $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{J}_1$, and $\mathbf{J}_2$ such that $\mathbf{DG} \neq \mathbf{GD}$, $\mathbf{DA} \neq \mathbf{AD}$, $\mathbf{DB} \neq \mathbf{BD}$, $\mathbf{DF} \neq \mathbf{FD}$, $\mathbf{DP} \neq \mathbf{PD}$.

5. Calculate the public key elements $\mathbf{Z}_1$, $\mathbf{W}_1$, $\mathbf{Z}_2$, and $\mathbf{W}_2$:

$$\mathbf{Z}_1 = \mathbf{AG}_1\mathbf{B}^{-1}, \mathbf{W}_1 = \mathbf{BJ}_1\mathbf{D}^{-1}, \mathbf{Z}_2 = \mathbf{FG}_2\mathbf{P}^{-1}, \text{ and } \mathbf{W}_2 = \mathbf{PJ}_2\mathbf{D}^{-1}. \quad (9)$$

Thus, the secret key (private key) represents the next set of vectors {**G**, **J**, $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{J}_1$, $\mathbf{J}_2$, **A**, **B**, **F**, **P**, **D**} with the total size of $110m$ ($176m$) bytes for 80-bit (128-bit) prime p. The public key represents the set of vectors {$\mathbf{Y}_1$, $\mathbf{Z}_1$, $\mathbf{U}_1$, $\mathbf{W}_1$, $\mathbf{Y}_2$, $\mathbf{Z}_2$, $\mathbf{U}_2$, $\mathbf{W}_2$} with the total size of $80m$ ($128m$) bytes for 80-bit (128-bit) prime p.

Note all elements of the public key are calculated as masked elements **G**, **J**, $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{J}_1$, $\mathbf{J}_2$ of the hidden group, besides, elements $\mathbf{Y}_1$ and $\mathbf{Y}_2$ ($\mathbf{U}_1$ and $\mathbf{U}_2$) are calculated from the same hidden group element **G** (**J**) using different masking factors **A** and **F** (**B** and **P**). Such connection of the public key elements with the hidden group underlies the correctness of the developed signature algorithm.

Signature generation procedure

A digital signature (e, **S**) to an electronic document M is calculated using the next randomized procedure:

1. Using the vector hash function $\mathbf{H}(\cdot)$ and concatenating the public key elements $\mathbf{Y}_1$ and $\mathbf{Y}_2$ to the document M, calculate the hash values $\mathbf{H}_1$ and $\mathbf{H}_2$:

$$\mathbf{H}_1 = \mathbf{H}(\mathbf{Y}_1||M) \text{ and } \mathbf{H}_2 = \mathbf{H}(M||\mathbf{Y}_2). \quad (10)$$

2. Generate at random a reversible vector **V** and two natural numbers k ($k < p - 1$) and t ($t < p - 1$) and calculate the vectors $\mathbf{R}_1$ and $\mathbf{R}_2$:

$$\mathbf{R}_1 = \mathbf{AG}^k\mathbf{J}^t\mathbf{G}_1\mathbf{J}_1\mathbf{VH}_1; \quad (11)$$

$$\mathbf{R}_2 = \mathbf{FG}^k\mathbf{J}^t\mathbf{G}_2\mathbf{J}_2\mathbf{VH}_2. \quad (12)$$

3. Using a specified $2|p|$-bit hash function f, generate the first signature element e as a hash-function value calculated from the document M to which the vectors $\mathbf{R}_1$ and $\mathbf{R}_2$ are concatenated: $e = e_1||e_2 = f(M||\mathbf{R}_1||\mathbf{R}_2)$, where the hash-value e is considered as concatenation of two $|p|$-bit natural numbers e_1 and e_2.

4. Calculate the integers s_1 and s_2:

$$s_1 = k - e_1 \bmod p - 1; \quad (13)$$

$$s_2 = t - e_2 \bmod p - 1. \quad (14)$$

5. Calculate the vector **S** as the second signature element:

$$\mathbf{S} = \mathbf{DG}^{s_1}\mathbf{J}^{s_2}\mathbf{V}. \quad (15)$$

6. Output the pair (e, **S**) as the signature to the document M.

Note that the accumulation of many unique values of the signature element **S** does not make it possible to form a system of power vector equations from which the attacker would be able to calculate the secret vectors **D**, **G**, and **J**. This is due to the fact that each unique signature (e, **S**) defines an equation of the form (15) with a unique unknown vector **V** that has a random value in the FNAA used as an algebraic support (see step 2 of the signature generation procedure).

Without taking into account the computational difficulty of finding the hash values e, $\mathbf{H}_1$, and $\mathbf{H}_2$ (which depend on the size of document M), the computational difficulty of the rest of the signature generation procedure can be approximately evaluated as four exponentiation operations: i) two exponentiations are performed to calculate both of the vectors $\mathbf{R}_1$ and $\mathbf{R}_2$ [note that in (11) and (12) the same two exponentiations are performed] and ii) two exponentiations are performed to calculate the vector **S**. Four exponentiations in the FNAA used as algebraic support take about $6m^2|p|$ multiplications in $GF(p)$ for $m \geq 6$ and $48|p| \approx 6150$ multiplications in $GF(p)$ for $m = 4$.

The signature size is equal to $(m + 2)|p|$ bits, for example, to $10(m + 2)$ bytes for the case of 80-bit prime p and $16(m + 2)$ bytes for 128-bit prime p.

Thus, the proposed signature algorithm has significantly lower signature size than many known post-quantum signature algorithms [15, 20].

Signature verification procedure

To verify a signature $(e, \mathbf{S})$ assigned to the document M, one is to use the public key of the signer $\{\mathbf{Y}_1, \mathbf{Z}_1, \mathbf{U}_1, \mathbf{W}_1, \mathbf{Y}_2, \mathbf{Z}_2, \mathbf{U}_2, \mathbf{W}_2\}$ and the following procedure:

1. Using the vector hash function $\mathbf{H}(\cdot)$ and formulas (10), calculate the values $\mathbf{H}_1$ and $\mathbf{H}_2$ from the document M to which the public key elements $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are attached.

2. Calculate the vectors $\mathbf{R}'_1$ and $\mathbf{R}'_2$ by the next two formulas:

$$\mathbf{R}'_1 = \mathbf{Y}_1^{e_1}\mathbf{Z}_1\mathbf{U}_1^{e_2}\mathbf{W}_1\mathbf{S}\mathbf{H}_1; \quad (16)$$

$$\mathbf{R}'_2 = \mathbf{Y}_2^{e_1}\mathbf{Z}_2\mathbf{U}_2^{e_2}\mathbf{W}_2\mathbf{S}\mathbf{H}_2. \quad (17)$$

3. Calculate the hash-value e' from the document to which the vectors $\mathbf{R}'_1$ and $\mathbf{R}'_2$ are attached: $e' = f(M||\mathbf{R}'_1||\mathbf{R}'_2)$.

4. If $e' = e$, then the signature is genuine, else the signature is false.

The computational difficulty of the signature verification procedure is approximately equal to that of the signature generation algorithm.

Correctness of the signature scheme is proven as follows.

Signature scheme correctness proof.

Compute the vectors $\mathbf{R}'_1$ and $\mathbf{R}'_2$:

$$\mathbf{R}'_1 = \mathbf{Y}_1^{e_1}\mathbf{Z}_1\mathbf{U}_1^{e_2}\mathbf{W}_1\mathbf{S}\mathbf{H}_1 =$$
$$= \mathbf{A}\mathbf{G}^{e_1}\mathbf{A}^{-1}\mathbf{A}\mathbf{G}_1\mathbf{B}^{-1}\mathbf{B}\mathbf{J}^{e_2}\mathbf{B}^{-1}\mathbf{B}\mathbf{J}_1\mathbf{D}^{-1}\mathbf{S}\mathbf{H}_1 =$$
$$= \mathbf{A}\mathbf{G}^{e_1}\mathbf{G}_1\mathbf{J}^{e_2}\mathbf{J}_1\mathbf{D}^{-1}\mathbf{D}\mathbf{G}^{s_1}\mathbf{J}^{s_2}\mathbf{V}\mathbf{H}_1 =$$
$$= \mathbf{A}\mathbf{G}^{e_1}\mathbf{G}_1\mathbf{J}^{e_2}\mathbf{J}_1\mathbf{G}^{k-e_1}\mathbf{J}^{t-e_2}\mathbf{V}\mathbf{H}_1 =$$
$$= \mathbf{A}\mathbf{G}^{k}\mathbf{J}^{t}\mathbf{G}_1\mathbf{J}_1\mathbf{V}\mathbf{H}_1 = \mathbf{R}_1;$$

$$\mathbf{R}'_2 = \mathbf{Y}_2^{e_1}\mathbf{Z}_2\mathbf{U}_2^{e_2}\mathbf{W}_2\mathbf{S}\mathbf{H}_2 =$$
$$= \mathbf{F}\mathbf{G}^{e_1}\mathbf{F}^{-1}\mathbf{F}\mathbf{G}_2\mathbf{P}^{-1}\mathbf{P}\mathbf{J}^{e_2}\mathbf{P}^{-1}\mathbf{P}\mathbf{J}_2\mathbf{D}^{-1}\mathbf{S}\mathbf{H}_2 =$$
$$= \mathbf{F}\mathbf{G}^{e_1}\mathbf{G}_2\mathbf{J}^{e_2}\mathbf{J}_2\mathbf{D}^{-1}\mathbf{D}\mathbf{G}^{s_1}\mathbf{J}^{s_2}\mathbf{V}\mathbf{H}_2 =$$
$$= \mathbf{F}\mathbf{G}^{e_1}\mathbf{G}_2\mathbf{J}^{e_2}\mathbf{J}_2\mathbf{G}^{k-e_1}\mathbf{J}^{t-e_2}\mathbf{V}\mathbf{H}_2 =$$
$$= \mathbf{F}\mathbf{G}^{k}\mathbf{J}^{t}\mathbf{G}_2\mathbf{J}_2\mathbf{V}\mathbf{H}_2 = \mathbf{R}_2.$$

Then compute the hash value $e' = f(M||\mathbf{R}'_1||\mathbf{R}'_2)$:

$$\{\mathbf{R}'_1 = \mathbf{R}_1; \mathbf{R}'_2 = \mathbf{R}_2\} \Rightarrow$$
$$\Rightarrow f(M\|\mathbf{R}'_1\|\mathbf{R}'_2) = f(M\|\mathbf{R}_1\|\mathbf{R}_2) \Rightarrow$$
$$\Rightarrow e' = e.$$

The latter equality determines validity of the verified signature.

Forging-signature attacks

Let's consider two attacks related to the formation of a genuine signature without knowing the secret key. In the first attack the forger selects arbitrary values of $\mathbf{S}$ and of e_1 and e_2 and, using the signature verification equations (16) and (17) computes the vectors $\mathbf{R}'_1$ and $\mathbf{R}'_2$ (note the attacker can also fix the values of e_1 and e_2 and modify only the value of $\mathbf{S}$). Then he computes the hash value $e' = e'_1||e'_2 = f(M||\mathbf{R}'_1||\mathbf{R}'_2)$ until $e'_1 = e_1$ and $e'_2 = e_2$. Probability that both of the latter two equalities hold true is equal to $\approx p^{-2}$, therefore, the computational difficulty of such attack can be estimated as $O(p^2)$ or as $O(2^{2|p|})$ multiplications in FNAA used as algebraic support. For the cases of 80-bit and 128-bit prime numbers p such attack is computationally infeasible.

One can propose some versions of the first attack, in which the vectors $\mathbf{R}'_1$ and $\mathbf{R}'_2$ are selected at random. Then the value $e' = f(M||\mathbf{R}'_1||\mathbf{R}'_2)$ and signatures $\mathbf{S}_1$ and $\mathbf{S}_2$ are computed from (16) and (17), correspondingly, until the equality $\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}$ holds true. The latter equality take place with probability $\approx p^{-m}$, causing the computational difficulty of the attack equal to $O(p^m)$, where $m \geq 4$.

In the second attack the forger uses some known genuine signature $(e, \mathbf{S})$ assigned to the document M (thus, he can compute $\mathbf{H}_1 = \mathbf{H}(\mathbf{Y}_1||M)$ and $\mathbf{H}_2 = \mathbf{H}(M||\mathbf{Y}_2)$) and attempts to calculate a valid signature $(e'', \mathbf{S}'')$ assigned to the document M''. He calculates the vectors $\mathbf{H}''_1 = \mathbf{H}(\mathbf{Y}_1||M'')$ and $\mathbf{H}''_2 = \mathbf{H}(M''||\mathbf{Y}_2)$. Then, using the formulas (16) and (17) and value of $e = e_1||e_2$ he calculates the vectors $\mathbf{R}'_1$ and $\mathbf{R}'_2$ and the value $e'' = e''_1||e''_2 = f(M, \mathbf{R}'_1||\mathbf{R}'_2)$. At the next step of the attack the forger tries to compute the vector $\mathbf{S}''$ satisfying the both verification equations. From (16) he has

$$\mathbf{R}'_1 = \mathbf{Y}_1^{e_1}\mathbf{Z}_1\mathbf{U}_1^{e_2}\mathbf{W}_1\mathbf{S}\mathbf{H}_1 = \mathbf{Y}_1^{e''_1}\mathbf{Z}_1\mathbf{U}_1^{e''_2}\mathbf{W}_1\mathbf{S}''\mathbf{H}''_1. \quad (18)$$

The equality (18) gives

$$\mathbf{S}'' = \mathbf{D}\mathbf{G}^{e_1-e''_1}\mathbf{J}^{e_2-e''_2}\mathbf{D}^{-1}\mathbf{S}\mathbf{H}_1\mathbf{H}''^{-1}_1. \quad (19)$$

In a similar way from (17) the forger gets

$$\mathbf{S}'' = \mathbf{D}\mathbf{G}^{e_1-e''_1}\mathbf{J}^{e_2-e''_2}\mathbf{D}^{-1}\mathbf{S}\mathbf{H}_2\mathbf{H}''^{-1}_2. \quad (20)$$

The second attack is successful, if the values of $\mathbf{S}''$ calculated from (19) and (20) are equal, i. e., if the following equality is true:

$$\mathbf{H}''^{-1}_2\mathbf{H}''_1 = \mathbf{H}_2^{-1}\mathbf{H}_1.$$

The latter takes place with a probability p^{-m} and determines the computational complexity of the attack equals to p^m, when a limited number (for example, less than 2^{40}) of valid signatures are available for the attacker, or to $p^{m/2}$, when an extremely large number ($\approx p^{m/2}$) of valid signatures is available for the attacker (consider, for example, a model of the oracle that signs documents generated by the attacker). In the latter case, the attacker selects $p^{m/2}$ different documents $M'^{(i)}$ ($i = 1, 2, ..., p^{m/2}$) and computes for every document $M'^{(i)}$ the value of $(\mathbf{H}_2'^{(i)})^{-1}\mathbf{H}_1'^{(i)}$. He also computes the values of $(\mathbf{H}_2^{(j)})^{-1}\mathbf{H}_1^{(j)}$($j = 1, 2, ..., p^{m/2}$) connected with the corresponding documents $M^{(j)}$ signed with the signatures $(e^{(j)}, \mathbf{S}^{(j)})$.

In correspondence with the birthday paradox, with probability equal to ≈ 0.5 the first set contains a value $(\mathbf{H}_2'^{(k)})^{-1}\mathbf{H}_1'^{(k)}$ such that $(\mathbf{H}_2'^{(k)})^{-1}\mathbf{H}_1'^{(k)} = (\mathbf{H}_2^{(t)})^{-1}\mathbf{H}_1^{(t)}$ for some natural number $t \le p^{m/2}$. The attacker orders all values from the first and second sets (performing $\approx m|p|p^{m/2}/2$ operations of comparison) and finds the values $(\mathbf{H}_2'^{(k)})^{-1}\mathbf{H}_1'^{(k)}$ and $(\mathbf{H}_2^{(t)})^{-1}\mathbf{H}_1^{(t)}$. Then, using formula (19) or (20) and the signature $\mathbf{S}^{(t)}$, he calculates the valid signature $\mathbf{S}'^{(k)}$ to the document $M'^{(k)}$ connected with the product $(\mathbf{H}_2'^{(k)})^{-1}\mathbf{H}_1'^{(k)}$. Calculation of the said two sets takes $4p^{m/2}$ operations of computing the vector hash function and $2p^{m/2}$ multiplications in the FNAA used as algebraic support. Computational complexity of the birthday-paradox-based attack can be roughly estimated as $O(p^{m/2})$.

Thus, the second forging attack is also impractical for the cases of 80-bit and 128-bit prime numbers p.

Attacks connected with calculation of the private key

The formulas (8) and (9) define connection between elements of the private and public keys. Namely, from (8) and (9) we have the following system of eight quadratic vector equations with eleven unknown vectors $\mathbf{G}$, $\mathbf{J}$, $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{J}_1$, $\mathbf{J}_2$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{F}$, $\mathbf{P}$, and $\mathbf{D}$:

$$\begin{cases} \mathbf{Y}_1\mathbf{A} = \mathbf{AG}; & \mathbf{Z}_1\mathbf{B} = \mathbf{AG}_1; \\ \mathbf{U}_1\mathbf{B} = \mathbf{BJ}; & \mathbf{W}_1\mathbf{D} = \mathbf{BJ}_1; \\ \mathbf{Y}_2\mathbf{F} = \mathbf{FG}; & \mathbf{Z}_2\mathbf{P} = \mathbf{FG}_2; \\ \mathbf{U}_2\mathbf{P} = \mathbf{PJ}; & \mathbf{W}_2\mathbf{D} = \mathbf{PJ}_2. \end{cases} \quad (21)$$

In this system, the vectors $\mathbf{G}$, $\mathbf{J}$, $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{J}_1$, and $\mathbf{J}_2$ are elements of the hidden commutative group. This fact is to be taken into account when representing the system (21) in the form of scalar equations in $GF(p)$. Such a reduction of the system (21) is performed using the BVMT setting the FNAA used as algebraic support.

Consider the case of 4-dimensional FNAA for which detailed decomposition into the set of commutative subalgebras is known [10]. One can consider the coordinates of one vector (for example, $\mathbf{G}$) from the hidden group as four scalar unknowns and then represent the coordinates of every of the other unknown vectors from the hidden group (for example, $\mathbf{J}$, $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{J}_1$, and $\mathbf{J}_2$) as four scalar values depending on two unique scalar values d and k in accordance with the formula (6). Such representation leads to transformation of quadratic equations into cubic ones, however, this does not increase the computational difficulty of solving the system of power equations, since the best known methods (based on the F4 [26] and F5 [27] algorithms) for finding a solution of such system have about the same efficiency for quadratic and cubic equations

The system (21) reduces to the system of $8m$ power equations (quadratic and cubic). In the case $m = 4$ we have the system of 32 equations with 34 unknowns in $GF(p)$. For the FNAAs of the dimensions $m \ge 6$ their decomposition into commutative subalgebras has not yet been investigated in detail, however, we have get preliminary results for the cases $m = 6$ and $m = 8$, which show these FNAAs contain commutative subalgebras every one of which can be described as a vector subspace (that is set by coordinates of a vector contained in the subalgebra) of the dimension $m/2$.

The latter means that, like in the case $m = 4$, in the cases $m = 6$ and $m = 8$ the system (21) reduces to the system of $\mu' = 8m$ scalar equations with $\eta = 6m + 5m/2$ scalar unknowns. Since the number of unknowns is larger than the number of equations, one can suppose the considered system of scalar power equations has many solutions. The latter fact is easily confirmed by considering system (21) at the level of the FNAA. Indeed, let system (21) has a solution $\mathbf{G}_0$, $\mathbf{J}_0$, $\mathbf{G}_{10}$, $\mathbf{G}_{20}$, $\mathbf{J}_{10}$, $\mathbf{J}_{20}$, $\mathbf{A}_0$, $\mathbf{B}_0$, $\mathbf{F}_0$, $\mathbf{P}_0$, and $\mathbf{D}_0$. Then for every reversible vector $\mathbf{X}$ (the number of reversible vectors is approximately equal to p^m) one has the following unique solution

$$\mathbf{G}_\mathbf{X} = \mathbf{X}^{-1}\mathbf{G}_0\mathbf{X};\ \mathbf{J}_\mathbf{X} = \mathbf{X}^{-1}\mathbf{J}_0\mathbf{X};$$
$$\mathbf{G}_{1\mathbf{X}} = \mathbf{X}^{-1}\mathbf{G}_{10}\mathbf{X};\ \mathbf{G}_{2\mathbf{X}} = \mathbf{X}^{-1}\mathbf{G}_{20}\mathbf{X};$$
$$\mathbf{J}_{1\mathbf{X}} = \mathbf{X}^{-1}\mathbf{J}_{10}\mathbf{X};\ \mathbf{J}_{2\mathbf{X}} = \mathbf{X}^{-1}\mathbf{J}_{20}\mathbf{X};$$
$$\mathbf{A}_\mathbf{X} = \mathbf{A}_0\mathbf{X};\ \mathbf{B}_\mathbf{X} = \mathbf{B}_0\mathbf{X};$$
$$\mathbf{F}_\mathbf{X} = \mathbf{F}_0\mathbf{X};\ \mathbf{P}_\mathbf{X} = \mathbf{P}_0\mathbf{X};\ \mathbf{D}_\mathbf{X} = \mathbf{D}_0\mathbf{X}.$$

At the level of a system of scalar equations, the presence of $\approx p^m$ solutions means the presence of dependent scalar equations. Namely, we can assume that the number (μ) of independent scalar equations is equal to the number of unknowns minus m. Thus, we have $\mu = \eta - m = 6m + 5m/2 - m$, i. e.

$$\mu = 6m + 3m/2. \quad (22)$$

Like in the case of multivariate public-key algorithms, the attack based on solving a system of many power equations with many unknowns can be called direct attack. Computational difficulty of solving such systems depends mainly on the number of equations (μ) and number of unknowns (η). In our case we have $\mu < \eta$, therefore, to evaluate security W of the introduced signature scheme to the direct attack, one can take the number of equations μ equal to the number of the unknowns (for example, the values of $\eta - \mu$ unknowns are predetermined) and use the minimum number of equations μ_{min} to get a given level of security, which has been calculated in [28], taking into account the best known algorithms for solving a system of power equations in the field $GF(n)$.The results of [28] are presented in Table 4. For the developed algorithm we have system of power equations in $GF(p)$, where $p \gg 2^8$, therefore, for a rough estimate of the required dimension of the used FNAAs one can take the values of μ_{min} from Table 4, which relate to the case $n = 2^8$.

Taking into account formula (22), for the case $m = 4$, $m = 6$, and $m = 10$ one gets $\mu = 30$, $\mu = 45$, and $\mu = 75$, correspondingly. Thus, for the said dimensions we have 80-bit, 128-bit, and 192-bit security of the introduced algorithm against the direct attack. To provide 256-bit security the introduced signature algorithms should be implemented on the FNAA of the dimension $m \geq 14$.

In the developed signature algorithm the signature element $\mathbf{S}$ is computed by formula (15), where the vector $\mathbf{V}$ is selected at random from the multiplicative group of the FNAA used as algebraic support. One can represent (15) in the form $\mathbf{V} = \mathbf{D}^{-1}\mathbf{G}^{-s_1}\mathbf{J}^{-s_2}\mathbf{S}$. The value of $\mathbf{V}$ is unique for every valid signature (e, $\mathbf{S}$), therefore, arbitrary fixed triple of values of the private-key elements $\mathbf{D}$, $\mathbf{G}$, and $\mathbf{J}$ can be connected with every known valid signature. This means that no information about the values of $\mathbf{D}$, $\mathbf{G}$, and $\mathbf{J}$ can be obtained from a large set of valid signatures, until two different signatures are calculated using the same value of $\mathbf{V}$. However, probability of the latter event is negligibly small ($\leq p^{-2}$, if the number of available signatures is equal or less $p^{m/2-1}$). In addition, it is not obvious how to establish signatures connected with the same value of $\mathbf{V}$. This is due to the fact that the value of $\mathbf{S}$ depends not only on $\mathbf{V}$, but also on the values of s_1 and s_2 that change from one signature to another.

The fundamental role of using a random value of $\mathbf{V}$ is easily seen when compared with the case of calculating the signature element $\mathbf{S}$ by the formula $\mathbf{S} = \mathbf{B}^{-1}\mathbf{G}^n\mathbf{J}^r\mathbf{C}^{-1}$ in [13], where natural numbers n and r are unique for every signature. The vector $\mathbf{G}^n\mathbf{J}^r$ is contained in the hidden group. For the case of using the 4-dimentional FNAA used as algebraic support, the unknown coordinates of the vector $\mathbf{G}' = (g'_0, g'_1, g'_2, g'_3) = \mathbf{G}^{n'}\mathbf{J}^{r'}$, where n' and r' relates to some given valid signature (e', $\mathbf{S}'$), can be used to describe the vector $\mathbf{G}^n\mathbf{J}^r$ with two unknown scalar values d and k [see formula (6)]. Suppose we have five different valid signatures (e', $\mathbf{S}'$), (e_1, $\mathbf{S}_1$), (e_2, $\mathbf{S}_2$), (e_3, $\mathbf{S}_3$), and (e_4, $\mathbf{S}_4$). Then the next system of five vector equations can be written:

$$\begin{cases} \mathbf{BS}'\mathbf{C} = \mathbf{G}'; \\ \mathbf{BS}_1\mathbf{C} = \mathbf{G}^{n_1}\mathbf{J}^{r_1}; \\ \mathbf{BS}_2\mathbf{C} = \mathbf{G}^{n_2}\mathbf{J}^{r_2}; \\ \mathbf{BS}_3\mathbf{C} = \mathbf{G}^{n_3}\mathbf{J}^{r_3}; \\ \mathbf{BS}_4\mathbf{C} = \mathbf{G}^{n_4}\mathbf{J}^{r_4}. \end{cases}$$

This system of the vector equations can be reduced to the system of 20 scalar power (quadratic and cubic) equations with 20 scalar unknowns (coordinates of the unknown vectors $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G}'$ and four pairs of unknows (d_1, k_1), (d_2, k_2), (d_3, k_3), and (d_4, k_4)). Solving the latter system one gets the values of the private-key elements. Taking into account the required minimum number of equations (see Table 4) one can recommend to implement algorithm from [13] on FNAAs of the dimensions $m = 10$, $m = 16$, and $m \geq 20$ for the cases of 2^{128}, 2^{192}, and 2^{256} security levels, correspondingly.

■ ***Table 4.*** The value of μ_{min} providing a given level of security against the direct attack [28]

$\log_2 W$	μ_{min} at	
	$n = 2^4$	$n = 2^8$
80	30	26
100	39	33
128	51	43
192	80	110
256	68	93

Discussion and conclusion

The article proposes a new method for developing signature algorithms with a hidden group, which are based on computational difficulty of solving large systems of quadratic multivariate equations, and introduces a new post-quantum signature algorithm. In the latter, the fitting parameter $\mathbf{S}$ of the signature (e, $\mathbf{S}$) is calculated by the formula (15) in which a random vector $\mathbf{V}$ is used as a multiplier. Due to the presence of a random multiplier, the accumulation of a large number of different signatures cannot be used to obtain additional equations, which made it possible to reduce the computational complexity of calculating the elements of the secret key, when using the known elements of the public key

and formulas (8) and (9). Thus, the proposed method for developing algebraic signature algorithms with a hidden group eliminates the disadvantage (mentioned in the Introduction) of the algebraic algorithms [13, 14] that use a verification equation with multiple occurrences of the signature element **S**.

The introduced method and the post-quantum signature algorithm represent an attractive alternative for using it as a prototype of a post-quantum signature standard, due to comparatively small sizes of signature, public and private keys.

Table 5 presents a rough comparison of the introduced algorithm with the multivariate signature algorithms selected as finalists of the NIST world competition on the development of the post-quantum public-key algorithms [20, 21].

In the proposed signature scheme we have specified using the FNAAs set over the ground finite field $GF(p)$ with characteristic $p = 2q + 1$, where q is a prime. However, one can use the primes p of an arbitrary form (for example, $p = 2^{80} + c_1$ and $p = 2^{128} + c_2$, where c_1 and c_2 are some specified natural numbers having a small size), since the latter does not influence the security level.

From a practical point of view, it seems very interesting to implement the proposed algorithm on FNAAs defined over finite fields of characteristic two, for example, over $GF(2^{80})$ and $GF(2^{128})$. This will reduce the hardware implementation cost and improve performance.

The performed evaluation of the security of the proposed algorithm is rather preliminary. A more detailed analysis of the features of the emerging system of power equations is required, which could potentially lead to finding special solution methods and to refinement of the security evaluation. It should also be noted that in the framework of future research, attention should be also paid to studying the detailed decomposition of the FNAAs into a set of commutative subalgebras for the cases of dimension $m \geq 6$.

■ ***Table 5.*** Comparison with two finalists of the NIST competition

Signature algorithm	Signature size, bytes	Key size, bytes		W
		Public	Private	
3 versions of Rainbow [20]	66	158 000	101 000	2^{128}
	164	861 000	611 000	2^{192}
	204	1 885 000	1 375 000	2^{256}
3 versions of GeMSS [21]	29	358 000	16	2^{128}
	47	1 294 000	24	2^{192}
	64	3 223 000	32	2^{256}
Proposed $m = 8$ $\|p\| = 80$	100	640	880	2^{128}
Proposed $m = 10$ $\|p\| = 128$	192	1280	1760	2^{192}
Proposed $m = 14$ $\|p\| = 128$	256	1792	2464	2^{256}

Acknowledgement

The authors sincerely thank anonymous Referee for valuable remarks and comments.

References

1. *Call for Additional Digital Signature Schemes for the Post-Quantum Cryptography Standardization Process*. September 6, 2022. 99 pp. Available at: https://csrc.nist.gov/csrc/media/Projects/pqc-dig-sig/documents/call-for-proposals-dig-sig-sept-2022.pdf (accessed 01 March 2023).
2. Ikematsu Y., Nakamura S., Takagi T. Recent progress in the security evaluation of multivariate public-key cryptography. *IET Information Security*, 2022, pp. 1–17. doi:10.1049/ise2.12092
3. Shor P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer. *SIAM Journal of Computing*, 1997, vol. 26, pp. 1484–1509.
4. Ekert A., Jozsa R. Quantum computation and Shor's factoring algorithm. *Reviews of Modern Physic*, 1996, vol. 68, pp. 733–752.
5. Alamelou Q., Blazy O., Cauchie S., Gaborit Ph. A code-based group signature scheme. *Designs, Codes and Cryptography*, 2017, vol. 82, no. 1–2, pp. 469–493. doi:10.1007/s10623-016-0276-6
6. Kosolapov Y. V., Turchenko O. Y. On the construction of a semantically secure modification of the McEliece cryptosystem. *Prikl. Diskr. Mat.*, 2019, no. 45, pp. 33–43. doi:10.17223/20710410/45/4
7. Dahmen E., Okeya K., Takagi T., Vuillaume C. *Digital signatures out of Second-Preimage Resistant Hash Functions*. In: *Post-Quantum Cryptography*. Johannes Buchmann J. and J. Ding ed. Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2008, vol. 5299, pp. 109–123. Available at: http://dblp.uni-trier.de/db/conf/pqcrypto/pqcrypto2008.html#DahmenOTV08 (accessed 01 March 2023).
8. Moldovyan A. A., Moldovyan N. A. Post-quantum signature algorithms based on the hidden discrete logarithm problem. *Computer Science Journal of Moldova*, 2018, vol. 26, no. 3(78), pp. 301–313.
9. Moldovyan D. N., Moldovyan A. A., Moldovyan N. A. Digital signature scheme with doubled verification

equation. *Computer Science Journal of Moldova*, 2020, vol. 28, no.1(82), pp. 80–103.
10. Moldovyan D. N. A practical digital signature scheme based on the hidden logarithm problem. *Computer Science Journal of Moldova*, 2021, vol. 29, no. 2(86), pp. 206–226.
11. Moldovyan D. N. New form of the hidden logarithm problem and its algebraic support. *Bulletin of Academy of Sciences of Moldova. Mathematics*, 2020, no. 2 (93), pp. 3–10.
12. Roman'kov V., Ushakov A., Shpilrain V. Algebraic and quantum attacks on two digital signature schemes. *Journal of Mathematical Cryptology*, 2023, vol. 17, no. 1, pp. 20220023. doi:10.1515/jmc-2022-0023
13. Moldovyan A. A., Moldovyan D. N., Moldovyan N. A. A novel method for developing post-quantum digital signature algorithms on non-commutative associative algebras. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 1, pp. 44–53. doi:10.31799/1684-8853-2022-1-44-53
14. Moldovyan D. N. A new type of digital signature algorithms with a hidden group. *Computer Science Journal of Moldova*, 2023, vol. 31, no. 1(91), pp. 111–124.
15. Ding J., Petzoldt A., Schmidt D. S. *Multivariate Cryptography*. In: *Multivariate Public Key Cryptosystems. Advances in Information Security*. Springer, New York, 2020, vol. 80, pp. 7–23. doi:10.1007/978-1-0716-0987-3_2
16. Ding J., Petzoldt A., Schmidt D. S. *Solving Polynomial Systems*. In: *Multivariate Public Key Cryptosystems. Advances in Information Security*. Springer, New York, 2020, vol. 80, pp. 185–248. doi:10.1007/978-1-0716-0987-3_8
17. Øygarden M., Felke P., Raddum H., Cid C. Cryptanalysis of the multivariate encryption scheme EFLASH. *Topics in Cryptology — CT-RSA 2020*. Lecture Notes in Computer Science, 2020, vol. 12006, pp. 85–105.
18. Ding J., Petzoldt A., Schmidt D. S. *MQDSS*. In: *Multivariate Public Key Cryptosystems. Advances in Information Security*. Springer, New York, 2020, vol. 80, pp. 153–168.
19. Ding J., Schmidt D. Rainbow, a new multivariable polynomial signature scheme. *Conference on Applied Cryptography and Network Security – ACNS 2005*. Springer Lecture Notes in Computer Science, 2005, vol. 3531, pp. 164–175.
20. *Rainbow Signature. One of Three NIST Post-quantum Signature Finalists*. 2021. Available at: https://www.pqcrainbow.org/ (accessed 01 March 2023).
21. *GeMSS: A Great Multivariate Short Signature*. Available at: https://www-polsys.lip6.fr/Links/NIST/GeMSS.html (accessed 01 March 2023).
22. Park A., Shim K.-A., Koo N., and Han D.-G. Side-channel attacks on post-quantum signature schemes based on multivariate quadratic equations. *Rainbow and UOV. IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2018, vol. 2018, no. 3, pp. 500–523. doi:10.46586/tches.v2018.i3.500-523
23. Moldovyan A. A., Moldovyan N. A. Post-quantum algebraic signature algorithms with a hidden group. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 1, pp. 29–40. doi:10.31799/1684-8853-2023-1-29-40
24. Moldovyan D. N., Moldovyan A. A., Moldovyan N. A. A novel method for development of post-quantum digital signature schemes. *Informatsionno-upravliaiushchie systemy* [Information and Control Systems], 2020, no. 6, pp. 21–29. doi:10.31799/1684-8853-2020-6-21-29
25. Moldovyan N. A. Unified method for defining finite associative algebras of arbitrary even dimensions. *Quasigroups and Related Systems*, 2018, vol. 26, no. 2, pp. 263–270.
26. Faugére J.-C. A new efficient algorithm for computing Gröbner basis (*F*4). *J. Pure Appl. Algebra*, 1999, vol. 139, no. 1–3, pp. 61–88.
27. Faugére J.-C. A new efficient algorithm for computing Gröbner basis without reduction to zero (*F*5). *Proc. of the Intern. Symp. on Symbolic and Algebraic Computation*, 2002, pp. 75–83. doi:10.1145/780506.780516
28. Ding J., Petzoldt A. Current state of multivariate cryptography. *IEEE Security and Privacy Magazine*, 2017, vol. 15, no. 4, pp. 28–36.

УДК 003.26
doi:10.31799/1684-8853-2023-3-59-69
EDN: GXPTKZ

Постквантовые алгоритмы цифровой подписи со скрытой группой и удвоенным проверочным уравнением

А. А. Молдовян[a], доктор техн. наук, главный научный сотрудник, orcid.org/0000-0001-5480-6016
Н. А. Молдовян[a], доктор техн. наук, главный научный сотрудник, orcid.org/0000-0002-4483-5048, nmold@mail.ru
[a]Санкт-Петербургский Федеральный исследовательский центр РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ

Введение: одной из текущих актуальных проблем криптографии является разработка постквантовых алгоритмов электронной цифровой подписи со сравнительно малыми размерами открытого ключа и подписи. **Цель:** разработать новый способ построения постквантовых алгебраических алгоритмов цифровой подписи со скрытой группой, основанных на вычислительной сложности решения больших систем квадратных уравнений с многими неизвестными, обеспечивающий уменьшение размеров открытого ключа и подписи по сравнению с известными аналогами. **Результаты:** предложен новый способ построения алгоритмов цифровой подписи

с подписью вида (e, $\mathbf{S}$), где e — натуральное число (параметр рандомизации) и $\mathbf{S}$ — вектор (подгоночный параметр), позволяющий уменьшить размерность конечных некоммутативных ассоциативных алгебр, используемых в качестве алгебраического носителя. Способ отличается использованием приема удвоения проверочного уравнения для фиксирования скрытой группы, в котором формирование вектора $\mathbf{S}$ выполняется в зависимости от случайного обратимого вектора и тем самым устраняется влияние числа подписанных документов на стойкость в известных алгоритмах-аналогах. Способ апробирован разработкой конкретного постквантового алгоритма цифровой подписи, использующего алгебры различных размерностей в зависимости от требуемого уровня стойкости. Выполнена предварительная оценка безопасности предложенного алгоритма. **Практическая значимость:** благодаря сравнительно небольшим размерам подписи и открытого ключа рассмотренный алгоритм подписи представляет значительный практический интерес как прототип постквантового стандарта подписи.

Ключевые слова — постквантовые криптосхемы, компьютерная безопасность, электронная цифровая подпись, криптография, задача дискретного логарифмирования, конечные некоммутативные алгебры, ассоциативные алгебры, коммутативные группы.

Для цитирования: Moldovyan A. A., Moldovyan N. A. Post-quantum signature algorithms with a hidden group and doubled verification equation. *Информационно-управляющие системы*, 2023, № 3, с. 59–69. doi:10.31799/1684-8853-2023-3-59-69, EDN: GXPTKZ

СВЕДЕНИЯ ОБ АВТОРАХ

БАЛОНИН Николай Алексеевич

Профессор кафедры вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 1982 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Автоматика и телемеханика».
В 2008 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 100 научных публикаций, в том числе трех монографий.
Область научных интересов — теория динамических систем, теория идентификации, теория операторов, теория матриц, вычислительные методы, интернет-робототехника, интернет-книги с исполняемыми алгоритмами, научные социальные сети.
Эл. адрес: korbendfs@mail.ru

ЕЗЕРСКИЙ Владимир Васильевич

Профессор кафедры эксплуатации и управления в аэрокосмических системах Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 1980 году окончил Военный инженерно-космический институт имени А. Ф. Можайского по специальности «Летательные аппараты».
В 2004 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 120 научных публикаций и восьми изобретений.
Область научных интересов — космические средства и воздействия космотехногенной среды на космические аппараты, системы ориентации и управления космическими аппаратами.
Эл. адрес: ezerskij2010@yandex.ru

КАРАНДАШЕВ Александр Александрович

Аспирант, ассистент кафедры аэрокосмических компьютерных и программных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 2021 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Встроенные системы обработки информации и управления Embedded Systems».
Является автором десяти научных публикаций.
Область научных интересов — сетевые протоколы, маршрутизация, алгоритмы на графах, бортовые сети, изучение динамики объектов, моделирование.
Эл. адрес:
aleksandr.karandashev@guap.ru

БРИТОВ Георгий Семенович

Доцент кафедры информационных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 1963 году окончил Ленинградский институт авиационного приборостроения по специальности «Авиационное приборостроение».
В 1968 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором более 100 научных публикаций и 13 патентов на изобретения.
Область научных интересов — теория надежности и техническая диагностика информационных систем.
Эл. адрес:
bgs@ibi.metrocom.ru

ИВАНОВ Михаил Евгеньевич

Аспирант кафедры эксплуатации и управления в аэрокосмических системах Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 2020 году окончил магистратуру Санкт-Петербургского государственного университета аэрокосмического приборостроения по специальности «Приборостроение».
Является автором пяти научных публикаций и одного свидетельства о регистрации ЭВМ.
Область научных интересов — космонавтика, моделирование аэрокосмических систем управления, методы искусственного интеллекта.
Эл. адрес: sindbad1995@list.ru

КАГИРОВ Ильдар Амирович

Научный сотрудник лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского Федерального исследовательского центра РАН.
В 2008 году окончил филологический факультет Санкт-Петербургского государственного университета по специальности «Лингвистика».
Является автором более 30 научных публикаций и одного патента на изобретения.
Область научных интересов — лингвистика жестовых языков, человеко-машинное взаимодействие, синтаксис и грамматическая семантика естественных языков, корпусная лингвистика.
Эл. адрес: kagirov@iias.spb.su

КИПЯТКОВА Ирина Сергеевна

Старший научный сотрудник лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского Федерального исследовательского центра РАН.
В 2008 году окончила Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Роботы и робототехнические системы».
В 2011 году защитила диссертацию на соискание ученой степени кандидата технических наук.
Является автором 100 научных публикаций и одного патента на изобретение.
Область научных интересов – автоматическое распознавание речи, нейронные сети.
Эл. адрес: kipyatkova@iias.spb.su

МОЛДОВЯН Николай Андреевич

Профессор, заведующий научно-исследовательским отделом проблем информационной безопасности Санкт-Петербургского института информатики и автоматизации РАН, заслуженный изобретатель РФ.
В 1975 году окончил Кишиневский политехнический институт по специальности «Полупроводниковые приборы».
В 2001 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 250 научных публикаций и 60 патентов на изобретения.
Область научных интересов – информационная безопасность, криптография, электронная цифровая подпись, блочные шифры.
Эл. адрес: nmold@mail.ru

ОБУХОВА Наталия Александровна

Доцент, декан факультета радиотехники и телекоммуникаций, заведующая кафедрой телевидения и видеотехники Санкт-Петербургского государственного электротехнического университета «ЛЭТИ».
В 1991 году окончила Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Системы автоматизированного проектирования».
В 2009 году защитила диссертацию на соискание ученой степени доктора технических наук.
Является автором 130 научных публикаций.
Область научных интересов – цифровая обработка изображений, машинное обучение, колориметрия, видеоаналитика.
Эл. адрес: naobukhova@etu.ru

МОЛДОВЯН Александр Андреевич

Профессор, главный научный сотрудник лаборатории кибербезопасности и постквантовых криптосистем Санкт-Петербургского института информатики и автоматизации РАН.
В 1974 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Автоматизированные системы управления».
В 2005 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 200 научных публикаций и 60 патентов на изобретения.
Область научных интересов – компьютерная безопасность, защита информации, криптография, протоколы электронной цифровой подписи.
Эл. адрес: maa1305@yandex.ru

МОТЫКО Александр Александрович

Доцент кафедры телевидения и видеотехники Санкт-Петербургского государственного электротехнического университета «ЛЭТИ».
В 2008 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Аудиовизуальная техника».
В 2012 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 80 научных публикаций.
Область научных интересов – цифровая обработка изображений, машинное обучение, колориметрия, видеоаналитика.
Эл. адрес: aamotyko@etu.ru

ОЛЕНЕВ Валентин Леонидович

Директор Института высокопроизводительных компьютерных и сетевых технологий Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 2007 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Информатика и вычислительная техника».
В 2012 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором более 100 научных публикаций и трех патентов на изобретения.
Область научных интересов – встроенные системы, бортовые космические и авиационные сети и др.
Эл. адрес: Valentin.Olenev@guap.ru

СВЕДЕНИЯ ОБ АВТОРАХ

СЕРГЕЕВ Михаил Борисович

Профессор, заведующий кафедрой вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения, почетный работник высшего профессионального образования РФ.
В 1980 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Электронные вычислительные машины».
В 2001 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 100 научных публикаций и 14 патентов на изобретения.
Область научных интересов — теория разрядных вычислений, методы проектирования спецпроцессоров для систем контроля и управления и др.
Эл. адрес: mbse@mail.ru

ПАМЯТКА ДЛЯ АВТОРОВ

Поступающие в редакцию статьи проходят обязательное рецензирование.

При наличии положительной рецензии статья рассматривается редакционной коллегией. Принятая в печать статья направляется автору для согласования редакторских правок. После согласования автор представляет в редакцию окончательный вариант текста статьи.

Процедуры согласования текста статьи могут осуществляться как непосредственно в редакции, так и по e-mail (ius.spb@gmail.com).

При отклонении статьи редакция представляет автору мотивированное заключение и рецензию, при необходимости доработать статью — рецензию.

Редакция журнала напоминает, что ответственность за достоверность и точность рекламных материалов несут рекламодатели.