

Численное моделирование фотонного тензорного ядра для аппаратного ускорения оптических матрично-векторных вычислений¹⁾

Г. А. Колосов, А. С. Шорохов, А. А. Федянин

Физический факультет, МГУ имени М. В. Ломоносова, 119991 Москва, Россия

Поступила в редакцию 11 октября 2024 г.

После переработки 5 ноября 2024 г.

Принята к публикации 12 ноября 2024 г.

Построена реалистичная численная модель фотонного тензорного ядра на основе архитектуры кросс-бар с поглотительными пленками халькогенидного стекла GeSbTe в качестве весовых элементов фотонной матрицы. Продемонстрирована работоспособность модели для совершения операции матрично-векторного умножения. Показана возможность применения тензорного ядра на основе реализованной архитектуры в сверточных нейронных сетях для задач распознавания изображений. Впервые численным моделированием были получены оценки потенциальной производительности и энергоэффективности фотонного аппаратного ускорителя с учетом современной экспериментальной элементной базы.

DOI: 10.31857/S0370274X24120227, EDN: ELSWWI

1. Введение. Технологические сложности при изготовлении структур нанометрового размера, а также проявление квантовых эффектов, например, туннелирование электронов через очень малый зазор, принципиально ограничивают размеры транзисторов. Традиционная компоновка вычислителей на архитектуре фон Неймана предполагает раздельное размещение процессора и памяти, что приводит к ограничению производительности – так называемое “бутылочное горлышко” фон Неймана [1]. Важность задач машинного обучения и искусственного интеллекта, результаты которых используются в большом количестве приложений, таких как распознавание речи и изображений, медицинская диагностика, предсказание свойств материалов и т.д. [2–4], увеличивает спрос на аппаратные ускорители вычислений в нейронных сетях (НС). Большинство производимых вычислений – матрично-векторное умножение (МВУ), поэтому ускорение данной операции имеет наибольший практический смысл.

В задачах обработки изображений использование полностью связанных НС не оправдано из-за чрезмерного количества свободных параметров и, как следствие, переобучения – алгоритм подстраивается под обучающую выборку и плохо работает на целевых данных. Для решения этой проблемы существует подход сверточной нейронной сети (СНС) [5, 6], в которой вход-

ное изображение преобразовывается системой цифровых фильтров перед поступлением на вход решающей полносвязной НС.

В последние годы было разработано множество электронных цифровых и аналоговых ускорителей [7, 8], однако в них также проявляются проблемы высокого энергопотребления из-за выделения тепла и ограничения скорости вычислений из-за необходимости перезарядки паразитных емкостей.

Фотонные вычислители [9], в том числе в виде фотонного тензорного ядра (ФТЯ) [10] – интегрального ускорителя, нацелены на параллельную обработку данных напрямую в памяти. К преимуществам подобного подхода можно отнести низкую задержку вычислений и высокую параллельность вычислений при использовании мультиплексирования различных длин волн в один канал входного вектора.

Многообещающей физической платформой для реализации весовых элементов матрицы ФТЯ являются фазопеременные материалы, в частности, халькогенидные стекла, такие как GeSbTe [11, 12], Sb₂Se₃ [13] и т.д. Материалы данного класса могут быстро изменять свое фазовое состояние с аморфного на кристаллическое и наоборот под действием оптического или электрического нагрева. У различных фазовых состояний материала отличается показатель преломления, что предоставляет возможность точного контроля оптического излучения в наноструктурах из фазопеременного материала.

Ключевыми преимуществами структур из фазопеременных материалов являются энергонезависи-

¹⁾См. дополнительный материал к данной статье на сайте нашего журнала www.jetpletters.ac.ru

²⁾e-mail: fedyanin@nanolab.phys.msu.ru

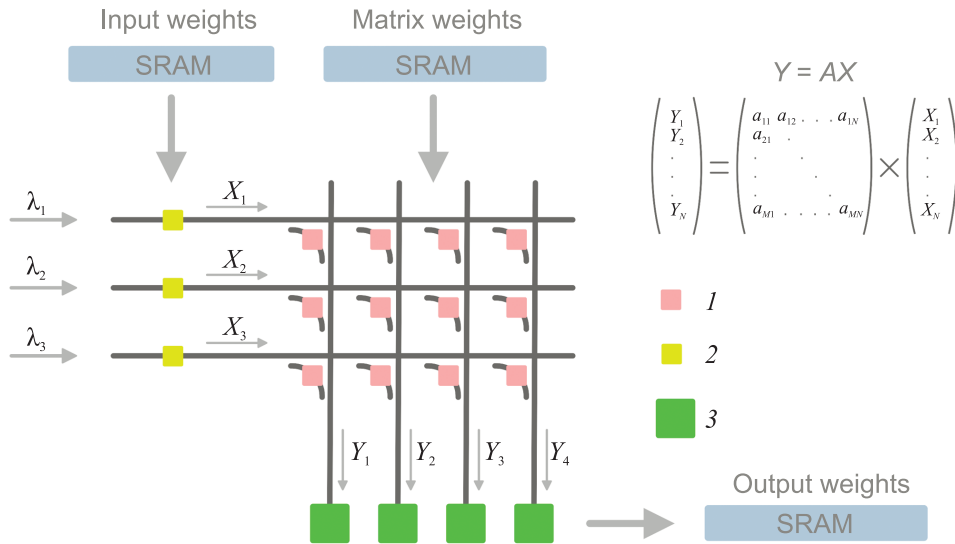


Рис. 1. (Цветной онлайн) Схема архитектуры некогерентного кроссбар-массива. Используемые входные и матричные веса загружаются из оперативной памяти (SRAM). Выходной вектор после считывания массивом детекторов загружается обратно в SRAM. На рисунке обозначения: 1 – идеальный модулятор или GST-поглотитель; 2 – микрокольцевой модулятор или модулятор Маха–Цендера; 3 – PIN-диод с усилителем и фильтром низких частот

мость фазового состояния – энергия потребляется лишь при переключении задаваемого структурой веса, что особенно актуально для вычислений в СНС, специфика которых заключается в умножении множества различных входных векторов на одинаковую весовую матрицу в процессе обработки изображения цифровым фильтром.

Для доказательства конкурентоспособности подхода ФТЯ необходимо знать параметры энергоэффективности и производительности ФТЯ. В предыдущих работах данные характеристики оценивались аналитически [14], что могло привести к переоценке этих значений. Целью данной работы является получение ключевых характеристик ФТЯ с помощью численного моделирования.

2. Методы. В данной работе с помощью метода переходных процессов в S-матрицах была построена модель ФТЯ на архитектуре кроссбар (см. рис. 1) [11]. Такую архитектуру возможно реализовать двумя принципиально различными способами: некогерентным и когерентным. В некогерентной архитектуре каждому каналу входного вектора соответствует своя длина волны излучения для возможности демultipлексирования выходного сигнала по длинам волн, а в когерентной архитектуре ФТЯ работает на одной длине волны, что приводит к труднорешаемой проблеме фазовой подстройки системы для обеспечения конструктивной интерференции [15]. Нами был сделан выбор в пользу некогерентной реализации ФТЯ, которая в том числе позволяет применять

методы дополнительного уплотнения потока обрабатываемой информации [16, 17], что схематически представлено на рис. 2.

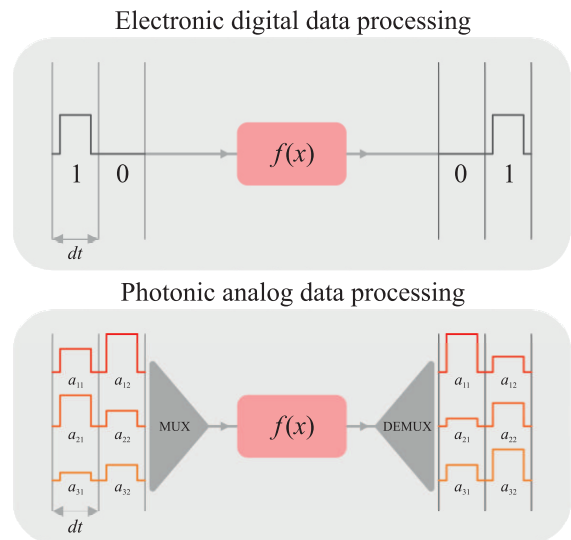


Рис. 2. (Цветной онлайн) Сравнение обработки информации в электронном ускорителе и в ФТЯ. В ФТЯ входной вес представлен в аналоговой форме, в отличие от цифрового представления в электронном ускорителе. Также ФТЯ предоставляет возможность дополнительного уплотнения потока информации с помощью мультиплексирования разных длин волн в один канал – позволяет одновременно умножать несколько входных векторов на весовую матрицу

В качестве поканальных модуляторов излучения входного вектора в работе рассматривались микрокольцевые модуляторы и модуляторы Маха–Цендера. К преимуществам первых можно отнести компактные геометрические размеры (до 10 мкм в диаметре [18]) и высокие частоты модуляции [19], к недостаткам же – высокую добротность резонанса [20], что приводит к необходимости термической подстройки модулятора [18] после изготовления. Модуляторы Маха–Цендера, в свою очередь, имеют существенно большие геометрические размеры [21], что негативно скажется на возможном масштабировании системы. По вышеуказанным причинам в данной работе был сделан выбор в пользу микрокольцевых модуляторов, для которых было экспериментально показано представление веса с точностью до 9 бит [22].

Для корректного детектирования выходного сигнала в некогерентной схеме необходимо сглаживать биения излучения на детекторе, возникающие из-за наличия близких длин волн в одном канале выходного вектора. Для этого на выходе детектора были использованы фильтр низких частот и трансимпедансный усилитель, рабочие параметры которых подбирались путем минимизации коэффициента битовых ошибок (КБО) при МВУ.

В качестве весовых элементов матрицы были выбраны пленки GeSbTe (GST) толщиной 10 нм, помещенные поверх кремниевого волновода – аморфное состояние ($n = 3.98 + 0.042i$ [23]) с большим пропусканием было взято за единицу, а кристаллическое с малым пропусканием – за ноль ($n = 6.49 + 1.054i$ [23]). Затухание волноводной моды в структуре было рассчитано методом конечных разностей. Точность задания матричного веса была взята меньшей или равной 6 бит [11], что является максимальным полученным на данный момент в эксперименте значением для интегральных GST-наноструктур.

При расчетах также были учтены реалистичные потери во всех элементах схемы и дисперсия направленных ответвителей (НО), так как дисперсия остальных элементов схемы существенно меньше (кроме микрокольцевых модуляторов, но их дисперсия компенсируется термодостройкой). Дисперсия НО была рассчитана исходя из выражения [24]:

$$P_{\text{cross}} = \kappa P_{\text{in}} = \sin^2\left(\frac{\pi \delta n}{\lambda}\right) P_{\text{in}}. \quad (1)$$

Здесь P_{in} – мощность излучения на входе ответвителя, P_{cross} – мощность в параллельном волноводе после прохождения ответвителя, κ – коэффициент связи НО, δn – разность эффективных показателей преломления симметричной и антисимметричной мод в

связанной секции НО, которая также была рассчитана методом конечных разностей. Коэффициенты связи κ в НО зависят от положения конкретного элемента в фотонной матрице, что отражено на рис. 3. Итоговая связь между входной и выходной мощностями излучения в элементе задается в виде:

$$P_{\text{out}} = P_{\text{in}} \frac{a_{ij}}{(N - i + 1)(M - j + 1)}, \quad (2)$$

где a_{ij} – вес элемента, задаваемый пленкой GST, N, M – число строк и столбцов в матрице, а i, j – индексы строки и столбца элемента соответственно. Подобные коэффициенты связи необходимы для получения каждым элементом матрицы одинаковой мощности излучения [11].

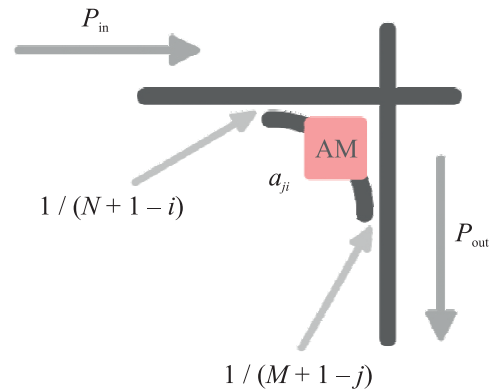


Рис. 3. (Цветной онлайн) Схематичное представление весового элемента ФТЯ. Стрелками отмечены направление распространения излучения и коэффициенты связи направленного ответвителя

При моделировании в систему вносились искусственные шумы и смещения не более половины битового интервала во входные веса и веса матрицы. Это позволило приблизить условия моделирования к реальным и оценить количество битовых ошибок на выходе системы, т.е. ее стабильность и точность работы. Все рабочие параметры модели суммированы в табл. 1.

Таблица 1. Параметры ФТЯ

Параметры модели	Значение
Частота модуляции	до 10 ГГц
Центральная частота лазера	193.1 ТГц
Интервал между частотами	от 0.1 ТГц [14]
Срез фильтра низких частот	18 ГГц
Сопrotивление усилителя	2 кОм
Входные веса	до 9 бит [22]
Матричные веса	до 6 бит [11]

3. Результаты. С использованием построенной модели было продемонстрировано умножение случайного входного вектора на случайную матрицу с

весами в диапазоне от 0 до 1 с помощью ФТЯ размерами от 2×2 до 9×9 , подробнее см. дополнительные материалы. С помощью квантователя (с точностью до 6 бит) на выходе усилителя получена картина сигнала в цифровом домене, что позволило рассчитать показатель КБО МВУ аналогового ФТЯ по отношению к контрольному МВУ, произведенному пакетом `numpy`. Битовой ошибкой считалось попадание выходного веса аналогового МВУ и выходного веса контрольного МВУ в различные уровни квантования. Минимизация КБО использовалась для дополнительной оптимизации элетронной обвязки детектора.

Результат умножения фотонным ядром показывает высокую точность попадания в доверительный интервал выходного вектора с предельно возможной точностью 6 бит, которая ограничивается сверху битовой точностью задания весов матрицы. По результатам численного моделирования можно сказать, что увеличение точности задания входных весов и весов матрицы выше 5 бит практически не оказывает влияния на КБО, так как начинает существенно влиять дисперсия НО.

Для изучения применимости ФТЯ к задачам глубокого обучения был рассмотрен вопрос эффективности переноса базовых слоев НС в оптический аналоговый домен. Для этого было проведено сравнение эффективности операции свертки с помощью ФТЯ с эталоном, полученным с помощью программного пакета `numpy`. Для этого была проведена свертка различных изображений из датасета CIFAR-10 с помощью четырех горизонтальных и вертикальных фильтров 3×3 , которые были перенесены в фотонную матрицу размером 9×4 элементов, подробнее см. дополнительные материалы. Входной вектор был получен в результате преобразования из двумерной матрицы, соответствующей участку изображения, в одномерный массив значений с последующей нормировкой. Из-за невозможности в выбранной архитектуре реализовать отрицательные веса в матрице ФТЯ проводилась постобработка результата свертки с помощью пакета `numpy`, используя следующее выражение:

$$B_{\text{real}} = 2B_{\text{core}} - IA_{\text{in}}, \quad (3)$$

где B_{real} – необходимый выходной вектор, B_{core} – результат умножения в ФТЯ, I – единичная матрица, а A_{in} – входной вектор. В будущем реализовать отрицательные веса возможно либо с помощью модоконвертирующей метаповерхности [25], либо с помощью подходов балансного детектирования из архитектуры микрокольцевого весового банка [26]. Результаты

численного моделирования операции свертки представлены на рис. 4.

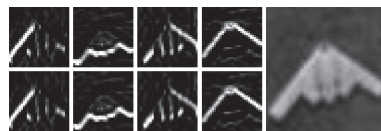


Рис. 4. Сравнение сверток. Верхний ряд изображений – контрольная свертка с помощью `numpy`. Нижний ряд – свертка с помощью ФТЯ. Справа исходное изображение из CIFAR-10

После демонстрации работоспособности модели в задачах прямого прохождения СНС было проведено исследование потенциала масштабирования кроссбар-массива. Для оценки предельного размера ФТЯ при заданных параметрах модели была рассчитана мощность на выходе системы при максимальных значениях входных и матричных весов (т.е. единичных) – оптический бюджет, также было проведено сравнение мощности минимального полезного сигнала на выходе ФТЯ (при максимальных входных весах и диагональной матрице с минимальными отличными от нуля весами) и мощности оптических шумов в системе (при максимальных входных весах и нулевых матричных весах). Подробнее об используемых матрицах и принципе получения максимального размера массива см. дополнительные материалы. Результаты расчетов представлены на рис. 5. Немонотонность зависимости полезного сигнала от масштабов кроссбар-матрицы можно объяснить эффектом деструктивных биений из-за наличия в выходном канале множества длин волн.

Как критерий достижения предельного размера ФТЯ было выбрано отношение минимального сигнала к шуму 1:1, т.е. разница в мощности в 3дБм – подобное отношение позволит различать два минимальных ненулевых уровня сигнала. Итоговые данные предельно возможного размера ФТЯ изображены на рис. 6. Немонотонное уменьшение размера массива с увеличением точности вычислений вытекает из эффекта деструктивных биений при расчетах минимального полезного сигнала. Оценка проводилась в зависимости от точности задания матричных весов, соответственно, от предельной точности итоговых вычислений. В данном моделировании ключевыми параметрами являются потери на каждом элементе в системе – они сведены в таблице 2.

Для понимания практической применимости ФТЯ необходимо оценить его энергоэффективность и производительность в тераоперациях в секунду. Эти показатели увеличиваются с ростом размера

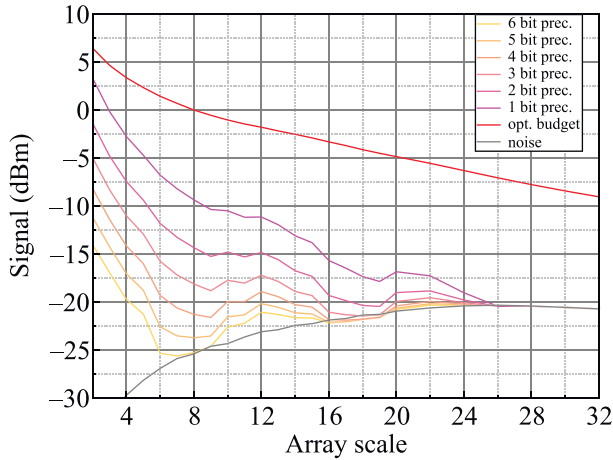


Рис. 5. (Цветной онлайн) Результат численной оценки минимального полезного сигнала в зависимости от размера по одной стороне массива с квадратной матрицей. Минимальный полезный сигнал существенно зависит от точности задания весов в массиве

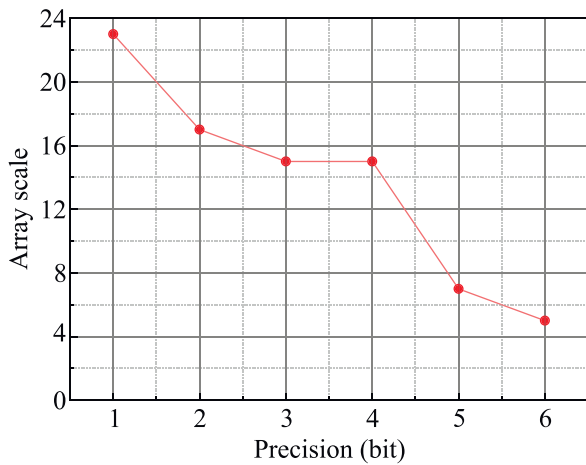


Рис. 6. (Цветной онлайн) Результат численной оценки максимального размера массива по одной стороне в зависимости от точности задания матричных весов

ФТЯ. Соответственно, необходимо использовать данные моделирования по определению максималь-

Таблица 2. Потери на каждом элементе ФТЯ

Потери	Значение, дБм
Лазер (входная мощность)	10 [27]
Волновод	-180/м
Направленный ответвитель	-0.1 [28]
Пересечение волноводов	-0.03 [29]
Шум в пересечении волноводов	-37 [29]
Микрокольцевой модулятор	-0.1 [30]
GST-поглотитель	-0.5 [31]

ного размера ФТЯ. С использованием современных экспериментальных данных по энергопотреблению активных элементов фотонной матрицы из табл. 3 были получены оценки энергоэффективности системы в зависимости от битности матричных весов и, соответственно, предельной точности вычислений. Для обеспечения постоянного потока входных данных и загрузки весовых данных нейросетевой модели необходима развитая иерархия памяти с глобальным буфером в виде DRAM, через который идет обмен информацией с электронной вычислительной частью, а также локальным объемом SRAM для быстрого доступа к входным и матричным весам. Более подробный анализ иерархии памяти и ее параметров может быть найден в работах [32, 33]. Энергопотребление обращения в память [34] также было учтено в модели.

Таблица 3. Энергопотребление элементов ФТЯ

Элемент	Энергопотребление
Лазер	10 мВт, КПД до 25% [35]
Микрокольцевой модулятор	500 фДж/бит с термоподстройкой 40 фДж/бит без термоподстройки [18]
GST-поглотители	20 пДж/переключение [23]
Детектор и усилитель	2.3 пДж/бит [36]
Обращение в память	3.9 пДж/бит [34]

При расчетах принималось, что матричные элементы переключаются один раз в 1000 циклов умножения, что примерно соответствует задаче свертки изображения размером 32 × 32 пикселей. При расчете оценок производительности ФТЯ в тераоперациях в секунду (ТОВС) за единичную операцию были взяты операции умножения и сложения, производимые в ходе МВУ. Оценки энергоэффективности и производительности представлены на рис. 7, из которого можно сделать вывод о повышении производительности ФТЯ при уменьшении точности вычислений, а также наличии минимального значения энергопотребления ядра при точности вычислений 4 бита.

4. Заключение. По результатам моделирования ФТЯ на архитектуре кроссбар показано, что энергетически наиболее оптимально использование ФТЯ размером 15 × 15 при использовании микрокольцевых модуляторов входного вектора со скоростью модуляции 10 ГГц и точности задания матричных весов 4 бита – при этом производительность ядра составит порядка 4 ТОВС при энергопотреблении менее 0.1 пДж/бит. Производительность одного ФТЯ сравнима с электронными ускорителями, что при нали-

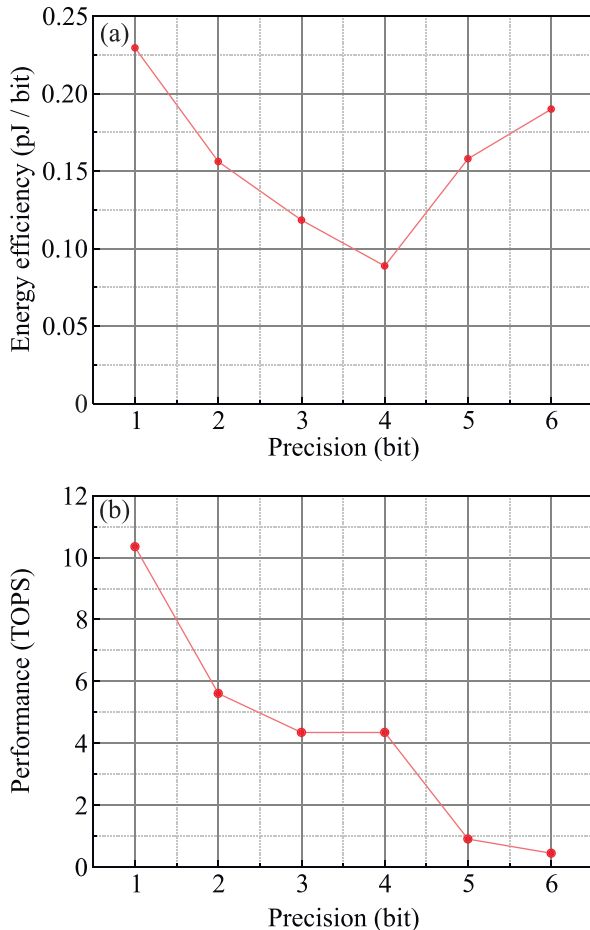


Рис. 7. (Цветной онлайн) (а) – Результат численной оценки энергоэффективности системы в пикоджоулях на операцию в зависимости от требуемой точности вычислений. (б) – Результат численной оценки производительности системы в ТОВС в зависимости от требуемой точности вычислений

чии множества ядер в фотонном ускорителе даст существенный прирост не только в энергоэффективности, но и производительности.

Дополнительно увеличить производительность и энергоэффективность устройства может помочь многопараметрическая оптимизация ФТЯ методами обратной разработки [37, 38], а также использование нескольких пакетов длин волн в параллельном режиме. Также необходима оптимизация отдельных элементов системы, а именно НО, пересечений волноводов, модуляторов и т.д. Это поможет снизить количество шумов в системе и, как следствие, увеличить потенциал масштабирования устройства. Использование метаповерхностей для модовой конверсии позволит снизить потери излучения в весовых элементах, что тоже положительно скажется на максимальном размере матрицы.

Финансирование работы. Исследование выполнено в рамках научной программы Национального центра физики и математики, направление # 1 “Национальный центр исследования архитектур суперкомпьютеров. Этап 2023–2025” и при поддержке Некоммерческого фонда развития науки и образования “Интеллект”.

Конфликт интересов. Авторы данной работы заявляют, что у них нет конфликта интересов.

1. W. Aspray, *John von Neumann and the origins of modern computing*, MIT Press, Cambridge (1990).
2. F. Amato, A. López, E. M. Pena-Méndez, P. Vañhara, A. Hampl, and J. Havel, *J. Appl. Biomed.* **11**, 47 (2013).
3. Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, *IEEE* **13**, 844 (2014).
4. А. А. Попкова, А. А. Федянин, *Письма в ЖЭТФ* **118**, 513 (2023).
5. J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, *Pattern Recognit.* **77**, 354 (2018).
6. H. J. Yoo, *IEIE SPC* **4**, 35 (2015).
7. C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, and J. Cong, *Proceedings of the 2015 ACM/SIGDA international symposium on field-programmable gate arrays* **161**, Association for Computing Machinery, N.Y. (2015).
8. K. Ovtchahrov, O. Ruwase, J. Y. Kim, J. Fowers, K. Strauss, and E. S. Chung, *Microsoft Research Whitepaper* **2**, 1 (2015).
9. А. И. Мусорин, А. С. Шорохов, А. А. Чежегов, Т. Г. Балуйан, К. Р. Сафронов, А. В. Четвертухин, А. А. Грунин, А. А. Федянин, *Успехи физических наук* **193**, 1284 (2023).
10. M. Miscuglio and V. J. Sorger, *Appl. Phys. Rev.* **7**, 31404 (2020).
11. F. Brücknerhoff-Plückelmann, J. Feldmann, C. D. Wright, H. Bhaskaran, and W. Pernice, *J. Appl. Phys.* **129**, 151103 (2021).
12. J. Zheng, A. Khanolkar, P. Xu, S. Colburn, S. Deshmukh, J. Myers, J. Frantz, E. Pop, J. Hendricksin, J. Doyle, N. Boechler, and A. Majumdar, *Opt. Mater. Express* **8**, 1551 (2018).
13. M. Delaney, I. Zeimpekis, D. Lawson, D. Hewak, and O. Muskens, *Adv. Funct. Mater.* **30**, 2002447 (2020).
14. J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, and H. Bhaskaran, *Nature* **589**, 7840 (2021).
15. N. Youngblood *IEEE J. Sel. Top. Quantum Electron.* **29**, 1 (2022).
16. B. Dong, S. Aggarwal, W. Zhou, U. E. Ali, N. Farmakidis, J. S. Lee, Y. He, X. Li, D. L. Kwong,

- C.D. Wright, W. Pernice, and H. Bhaskaran, *Nature Photon.* **17**, 1080 (2023).
17. F. Brücknerhoff-Plückelmann, J. Feldmann, H. Gehring, W. Zhou, C.D. Wright, H. Bhaskaran, and W. Pernice, *Nanophotonics* **11**, 4063 (2022).
 18. P. Dong, R. Shafiha, S. Liao, H. Liang, N.N. Feng, D. Feng, G. Li, X. Zheng, A.V. Krishnamoorthy, and M. Asghari, *Opt. Express* **18**, 10941 (2010).
 19. Y. Zhang, H. Zhang, J. Zhang, J. Liu, L. Wang, D. Chen, N. Chi, X. Xiao, and S. Yu, *Photonics Res.* **10**, 1127 (2022).
 20. W. Bogaerts, P. De Heyn, T. van Vaerenbergh, K. De Vos, S.K. Selvaraja, T. Claes, P. Dumon, P. Bienstmann, D. van Thourhout, and R. Baets, *Laser Photonics Rev.* **6**, 47 (2012).
 21. M.R. Watts, W.A. Zortman, D.C. Trotter, R.W. Young, and A.L. Lentine, *IEEE J. Sel. Top. Quantum Electron.* **16**, 159 (2010).
 22. W. Zhang, C. Huang, H.T. Peng, S. Bilodeau, A. Jha, E. Blow, T.F. de Lima, B.J. Shastri, and P. Prucnal, *Optica* **9**, 579 (2022).
 23. D. Wu, X. Yang, N. Wang, L. Lu, J. Chen, L. Zhou, and B. Rahman, *Nanophotonics* **11**, 3437 (2022).
 24. A. Yariv, *IEEE J. Quantum Electron.* **9**, 919 (1973).
 25. C. Wu, H. Yu, S. Le, R. Peng, I. Takeuchi, and M. Li, *Nature Commun.* **12**, 96 (2021).
 26. A.N. Tait, M.A. Nahmias, B.J. Shastri, and P.R. Pruchal, *J. Light. Technol.* **32**, 4029 (2014).
 27. S. Keyvaninia, G. Roelkens, D. van Thourhout, C. Jany, M. Lamponi, A. Le Lieprve, F. Lelarge, D. Make, G.H. Duan, D. Bordel, and J.M. Fedeli, *Opt. Express* **21**, 3784 (2013).
 28. M.A. Al-Qadasi, L. Chrostowski, B.J. Shastri, and S. Shekhar, *APL Photonics* **7**, 20902 (2022).
 29. S. Wu, X. Mu, L. Cheng, S. Mao, and H.Y. Fu, *Micromachines* **11**, 326 (2020).
 30. S. Manipatruni, K. Preston, L. Chen, and M. Lipson, *Opt. Express* **18**, 18235 (2010).
 31. Z. Yu, J. Zheng, P. Xu, W. Zhang, and Y. Wu, *IEEE Photonics Technol. Lett.* **30**, 250 (2017).
 32. C. Demirkiran, F. Eris, G. Wang, J. Elmhurst, N. Moore, N.C. Harris, A. Basumallik, V.J. Reddi, A. Joshi, and D. Bunadar, *J. Emerg. Technol. Comput. Syst.* **19**, 1 (2023).
 33. D. Sturm and S. Mozaeni, *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, **1**, EDA Consortium, San Jose (2023).
 34. M. O'Connor, N. Chatterjee, D. Lee, J. Wilson, A. Agrawal, S.W. Keckler, and W.J. Dally, *Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture*, **41**, Association for Computing Machinery, N.Y. (2017).
 35. B.R. Koch, E.J. Norberg, J.E. Roth, B. Kim, A. Ramaswamy, R.S. Guzzon, J. Hutchinson, J.H. Shin, J. Imamura, B. Gomez, G. Fish, and A. Fang, *Novel In-Plane Semiconductor Lasers XIII. – SPIE* **9002**, 72 (2014).
 36. S. Nayak, A.H. Ahmed, A. Sharkia, A.S. Ramani, S. Mirabbassi, and S. Shekhar, *IEEE Trans. Circuits Syst. I: Regul. Pap.* **66**, 3162 (2019).
 37. К.Р. Сафронов, В.О. Бессонов, А.А. Федянин, *Письма в ЖЭТФ* **114**, 360 (2021).
 38. M. Wei, X. Lin, K. Xu, Y. Wu, C. Wang, Z. Wang, K. Lei, K. Bao, J. Li, L. Li, E. Li, and H. Lin, *Nanophotonics* **13**, 2183 (2024).