

DOI: <https://doi.org/10.30898/1684-1719.2026.2.12>

УДК: 621.383

ФОТОННЫЕ ТЕХНОЛОГИИ ДЛЯ РЕАЛИЗАЦИИ НЕЙРОМОРФНЫХ ВЫЧИСЛЕНИЙ: РЕЗУЛЬТАТЫ, НАПРАВЛЕНИЯ ИССЛЕДОВАНИЙ И АППАРАТНОЕ ОБЕСПЕЧЕНИЕ ИИ

А.А. Мачнев^{1,2}, И.В. Мельников², Е.Р. Надеждин², В.А. Соколов², Д.Д. Ступин²

¹Национальный исследовательский технологический университет «МИСиС»,
119049, Москва, Ленинский проспект, д. 4, стр. 1

²Московский физико-технический институт
(национальный исследовательский университет),
141701, Московская область, г. Долгопрудный, Институтский переулок, д. 9

Статья поступила в редакцию 21 июля 2025 г.

Аннотация. Внедрение фотонных технологий в индустрию высокопроизводительных вычислений позволяет серьезно трансформировать подходы к созданию новых поколений сверхбыстродействующих вычислительных систем, преодолевая присущие традиционным электронным вычислительным системам и технологиям ограничения с точки зрения энергоэффективности, распараллеливания и масштабируемости. Используя уникальные свойства светового излучения с точки зрения обеспечения высокой пропускной способности, малых временных задержек и возможности параллельной обработки больших информационных массивов и потоков, фотонные вычислительные системы представляют собой многообещающую альтернативу для выполнения сложных вычислительных операций, в том числе, в приложениях, предполагающих использование технологий искусственного интеллекта. В представленной работе синтезируются последние достижения в области фотонных нейроморфных архитектур, аналоговых оптических вычислений, вычислений на фотонных резервуарах и интегрированных

фотонных схемах, а также дается представление о будущих направлениях исследований и потенциальных применениях этих систем для обеспечения следующего поколения аппаратного обеспечения сверхвысокопроизводительных вычислений.

Ключевые слова: фотоника, фотонные технологии, высокопроизводительные вычисления, нейроморфные вычисления, искусственный интеллект, фотонные интегральные схемы, параллельные вычисления, нейронные сети.

Автор для переписки: Ступин Дмитрий Дмитриевич, ddstupin@yandex.ru

Введение

Технологии искусственного интеллекта (ИИ) в настоящее время находят все более широкое применение в самых разных областях человеческой деятельности. Помимо решения целого ряда насущных проблем нашей цивилизации, технологии ИИ оказывают мощное стимулирующее воздействие на развитие средств и технологий высокопроизводительных вычислений. До недавнего времени возможности решения большинства задач, связанных с интеллектуальной обработкой информации, в значительной степени обеспечивались «классическими» компьютерными системами либо симбиозом классических вычислительных систем с специализированными устройствами (видеокарты, процессоры цифровой обработки, системы на ПЛИС и так далее). Однако непрерывный рост требований к процессам обработки информации, включая лавинообразное увеличение потребностей в обработке сверхбольших массивов и потоков данных («большие данные», (Data Science, Big Data etc.) приводит к необходимости повышения производительности вычислительных систем, обеспечивая при этом разумный энергетический баланс производимых вычислений: современные суперкомпьютеры потребляют мощности порядка единиц и десятков мегаватт, что переводит задачи высокопроизводительных вычислений из разряда чисто научно-технических в разряд сложных и ресурсоемких научно-инженерных проблем.

Развитие фотоники, ориентированной на использование в информационных технологиях управления параметрами оптического излучения, может обеспечить новые возможности в плане реализации высокопроизводительных вычислений, с одной стороны, за счет использования процессов распространения электромагнитных волн, характеризующихся максимально возможными скоростями передачи/приема информации, а с другой – за счет возможности обеспечения высокой энергетической эффективности реализации этих процессов, например, вследствие малого затухания сигналов в специализированных оптических средах. При этом возможность управлять параметрами оптических сред позволяет реализовывать аналоговые вычисления, эквивалентные по своей производительности цифровым вычислениям на наиболее «продвинутых» суперкомпьютерах, но с значительно меньшими энергетическими затратами.

Именно возможность оптимизации энергетических затрат при проведении высокопроизводительных вычислений делает фотонные технологии весьма привлекательными для использования в перспективных вычислительных системах, включая системы и комплексы, предполагающие использование методов и технологий ИИ. Многочисленные попытки исследователей и разработчиков смоделировать интеллектуальные процессы, происходящие в человеческом мозге, с помощью цифровых вычислений постоянно наталкиваются на проблемы энергоэффективности: самые современные суперкомпьютеры для воспроизведения даже относительно примитивных интеллектуальных процессов потребляют мощность порядка единиц мегаватт в то время, как человеческому мозгу для реализации аналогичных процедур достаточно нескольких десятков ватт.

Одним из возможных направлений преодоления «энергетического проклятия» цифровых вычислительных технологий, заключающегося в необходимости повышения энергопотребления вычислительных устройств на несколько мегаватт при увеличении производительности на порядок, является направление, связанное с развитием нейроморфных вычислений. Такого рода вычисления, по мнению многих экспертов, в максимальной степени

аналогичны процессам обработки информации, происходящим в человеческом мозге, что делает их привлекательными для использования в системах с ИИ. По мнению авторов настоящей статьи, конвергенция нейроморфных вычислений с фотонными технологиями позволяет сформировать новые подходы к организации вычислительных процессов в системах, использующих методы и технологии ИИ, и получить, как результат, новые научно-технические решения, позволяющие создать высокопроизводительные энергоэффективные интеллектуальные вычислительные системы следующего поколения для использования в различных областях человеческой жизнедеятельности.

1. Типы высокопроизводительных вычислений и вычислительные архитектуры, в которых используются фотонные технологии

1.1. Матрично-векторное умножение и крупномасштабные фотонные процессоры

Операция умножения матрицы на вектор (MVM) является краеугольным камнем в большинстве архитектур нейронных сетей, где веса и входные сигналы объединяются для получения выходных сигналов, которые затем передаются на выходные устройства посредством различных функций активации. Традиционные электронные схемы выполняют MVM последовательно, требуя существенной энергии и большого времени вычислений, особенно, если речь идет о вычислениях, производимых с матричными структурами большой размерности. Фотонные системы, в отличие от электронных схем, могут выполнять операции MVM параллельно по нескольким каналам, используя присущий световым потокам параллелизм и обеспечивая при этом радикальное сокращение затрат вычислительных ресурсов различных видов (время, энергетика и т.д.). Одной из наиболее многообещающих архитектур в этой области является крупномасштабный фотонный матричный процессор, предложенный Брюкерхоффом-Плюкельманном и др. [1]. Предложенная на рис. 1 система использует обработку матриц с временным мультиплексированием для виртуального расширения размера фотонных перекрестных массивов,

что позволяет производить вычисления с матрицами размером $16\,000 \times 64$ без увеличения физического размера чипа [2]. Это достигается с помощью некогерентного оптического накопления с временным мультиплексированием, при этом обеспечивается точность около 98,9 % с оптическими импульсами длительностью ~ 1 нс.

Как показано на рис. 1, в процедуре обработки можно выделить три ключевых элемента, (A), (B) и (C), соответственно. Целесообразно представить краткие комментарии к каждому:

(A) Мультиплексированное по времени умножение вектора на матрицу (MVM). Осуществляется разложение MVM большого масштаба на несколько небольших MVM, которые вычисляет фотонная матрица перекрестных линий. Затем происходит суммирование промежуточных результатов путем накопления заряда внутри фотодетектора. Таким образом, не требуется никакой дополнительной электронной обработки;

(B) Концепция активной ячейки матрицы с произвольными значениями, когда ячейки матрицы спроектированы с использованием MZM в качестве веса матрицы. Таким образом, вес может модулироваться так же быстро, как и входной вектор, что имеет решающее значение для такой архитектуры. Более того, MZM позволяет выполнять эталонные вычисления без потерь;

(C) Эскиз полного матричного процессора. Процессор вычисляет несколько MVM параллельно, кодируя каждый вектор в своем диапазоне длин волн. Перед отправкой векторов в матрицу перекрестных линий они мультиплексируются вместе, а затем снова демультиплексируются для получения отдельных выходных векторов. Наконец, выполняется сбалансированное детектирование для получения правильного результата.

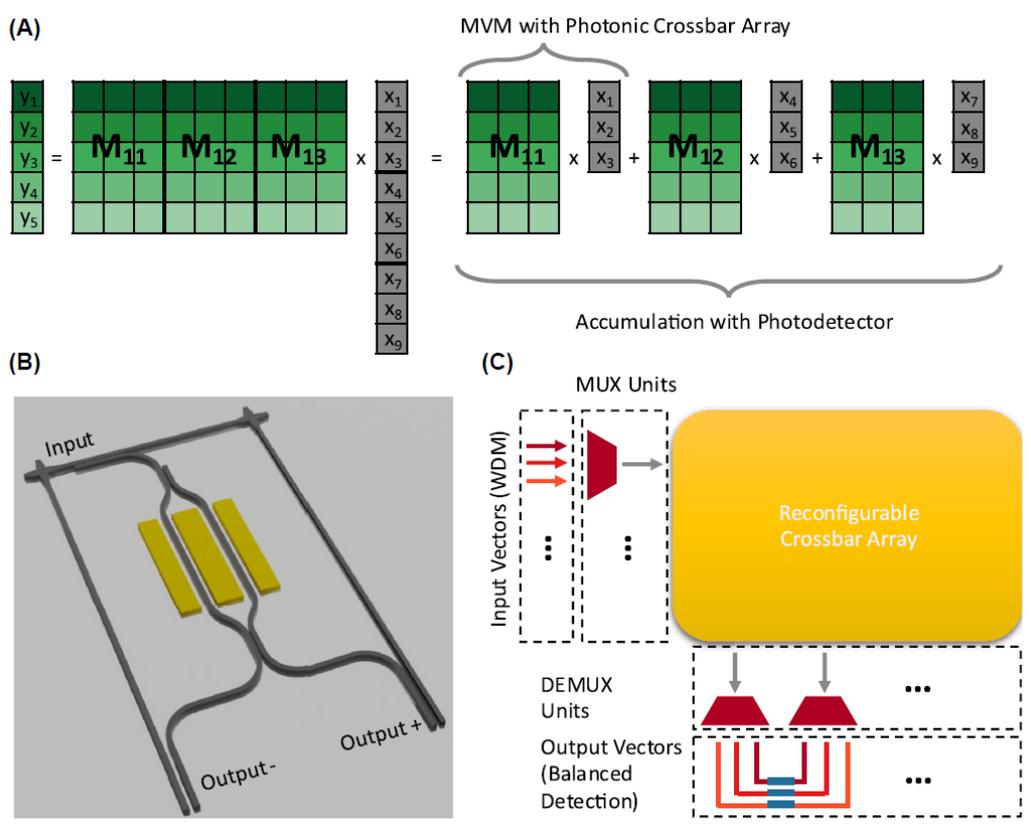


Рис. 1. Схема фотонного умножения вектора на матрицу большого масштаба [2].

Такая конструкция, реализованная на кристалле площадью 51,2 мм², может выполнять более 110 триллионов (примерно 10¹⁴) операций умножения и накопления (MAC) в секунду, что делает ее идеально подходящей для высокопроизводительных приложений ИИ в реальном времени, таких как автономное вождение, обработка естественного языка и распознавание изображений в реальном времени.

Ключевым компонентом архитектуры фотонного матричного процессора являются модуляторы Маха-Цендера (MZM), которые расположены в конфигурации кроссбара для формирования основных вычислительных элементов. Каждый элемент в кроссбаре может быть индивидуально запрограммирован путем установки нужной фазы входящего света, определяя вес соответствующего элемента матрицы. Матричное умножение выполняется с помощью интерференционных картин, генерируемых MZM, в которых интенсивность выходного света кодирует результаты умножения. Для достижения высокоточного умножения система использует комбинацию методов когерентной и некогерентной модуляции света, улучшая отношение сигнал/шум и минимизируя

ошибки, вызванные фазовыми несоответствиями или оптическими потерями. Эта структура дополняется использованием временного мультиплексирования (TDM) для увеличения эффективного размера матрицы путем кодирования различных элементов в отдельных временных интервалах, тем самым достигая высокой степени параллелизма без необходимости привлечения дополнительных физических ресурсов.

Архитектура кроссбара на основе MZM обладает высокой степенью реконфигурируемости, что позволяет динамически перенастраивать весовые коэффициенты сети посредством оптической настройки. Это достигается с помощью интегрированных фазовращателей на основе термооптических или электрооптических эффектов, которые изменяют эффективный показатель преломления волноводов в плечах MZM, тем самым управляя фазой света, распространяющегося через них. Термооптические фазовращатели используют локализованные нагревательные элементы для изменения показателя преломления кремниевых волноводов, в то время как электрооптические фазовращатели используют такие материалы, как ниобат лития или кремний-органические гибридные соединения, для достижения быстрой маломощной фазовой модуляции. Эти фазовращатели соединяются с фотодетекторами на выходах MZM, которые преобразуют оптические сигналы обратно в электрический домен для последующей обработки или дальнейших оптических вычислений. Тщательно управляя фазовыми сдвигами в каждом MZM, в рамках такой структуры можно реализовать широкий спектр матричных операций, от простых умножений до сложных операций свертки.

1.2. Повышение параллелизма вычислений за счет мультиплексирования длин волн и мод

Производительность системы может дополнительно повыситься за счет интеграции на чипе мультиплексоров и демультимплексоров длин волн, которые позволяют обрабатывать несколько сигналов с разными оптическими длинами волн одновременно. Каждая оптическая длина волны соответствует отдельному

входному вектору, что позволяет системе выполнять несколько матричных умножений параллельно. Этот подход с мультиплексированием по длине волны (WDM) дополняется использованием оптических усилителей для компенсации вносимых потерь и поддержания целостности сигнала по всему чипу. Для этой цели обычно используются усилители на основе легированного эрбием волокна (EDFA) и/или полупроводниковые оптические усилители (SOA), которые обеспечивают усиление на рабочих длинах волн при минимизации шума и искажений. Эти усилители интегрированы в чип с использованием конструкций на основе волноводов, что обеспечивает компактное усиление с малыми потерями, сохраняющее когерентность оптических сигналов (рис. 2 [3]).

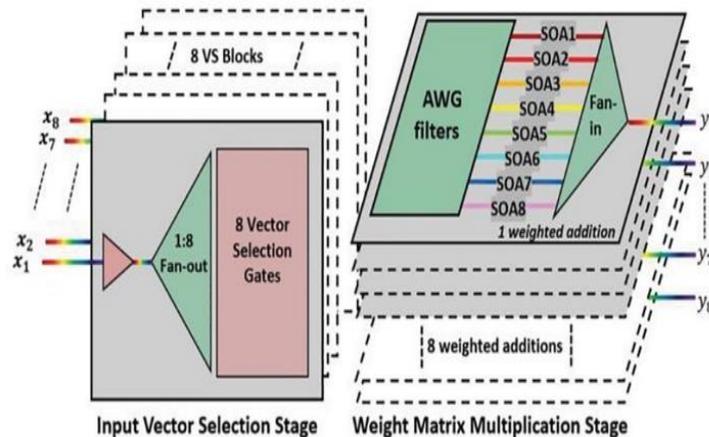


Рис. 2. Однонаправленная оптическая нейронная сеть, основанная на частотном мультиплексировании (WDM) и полупроводниковых оптических усилителях (SOA) [3].

Более того, недавние достижения в области мультиплексирования мод (MDM) [4] позволили достичь еще больших уровней параллелизма за счет использования различных пространственных мод в одном волноводе для кодирования нескольких потоков данных. Эта технология реализуется путем проектирования многомодовых волноводов, в которых каждая мода действует как независимый канал для передачи информации. Преобразователи мод и мультиплексоры используются для выборочного возбуждения и соединения этих мод, позволяя системе выполнять параллельные вычисления на различных пространственных каналах одновременно. Например, в многомодовой решетке MZM каждая мода может независимо модулироваться и обрабатываться,

эффективно увеличивая вычислительную мощность системы без увеличения ее физического размера (см. рис. 3 [4]). Такие многомодальные подходы особенно полезны для приложений, требующих высокой пропускной способности и низкого энергопотребления, таких как высокоскоростные системы связи и вычисления с использованием нейронных сетей в реальном времени.

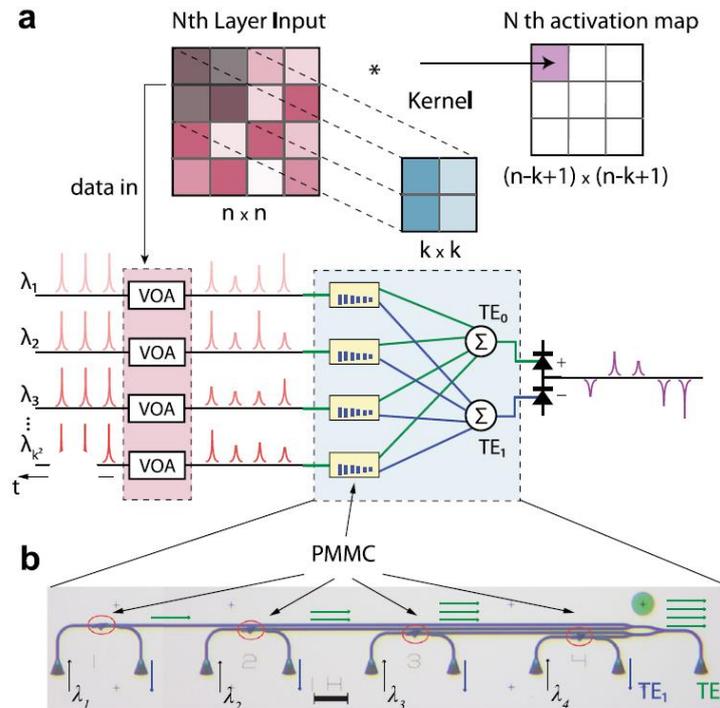


Рис. 3. Оптическая нейронная сеть на основе MDM и WDM [4].

1.3. Гибридные фотонно-электронные архитектуры

Для дальнейшего повышения вычислительной мощности фотонных систем разрабатываются гибридные фотонно-электронные архитектуры, в которых делается попытка объединить сильные стороны обеих технологий. Эти гибридные системы используют высокоскоростные и маломощные характеристики фотонных схем для операций с интенсивным использованием данных, а электронные компоненты – для управления и программирования. Одним из наиболее перспективных подходов считается использование гибридных фотонно-электронных матричных умножителей. В этих устройствах операции умножения выполняются в оптическом диапазоне, а функции накопления и активации реализуются в электронных схемах. Такая конфигурация позволяет использовать преимущества оптического параллелизма, сохраняя

гибкость электронной обработки, что, в частности, дает возможность реализовывать сложные модели нейронных сетей с высокой степенью точности.

В этих гибридных системах фотонный матричный умножитель состоит из массива MZM либо микрокольцевых резонаторов (MRR), которые выполняют умножение входного вектора на весовую матрицу [5-8]. Затем оптические сигналы с помощью фотодетекторов преобразуются в электрический домен, где они суммируются и пропускаются через нелинейные функции активации, такие как выпрямленные линейные единицы (ReLU) или сигмоидальные функции. Такой подход снижает энергопотребление и временную задержку, связанные с полностью электронными реализациями, при этом сохраняя способность выполнять сложные нелинейные операции. Использование гибридных архитектур также упрощает интеграцию фотонных схем с существующими электронными системами, что позволяет встраивать фотонные ускорители в обычные вычислительные платформы, такие как CPU, GPU и FPGA.

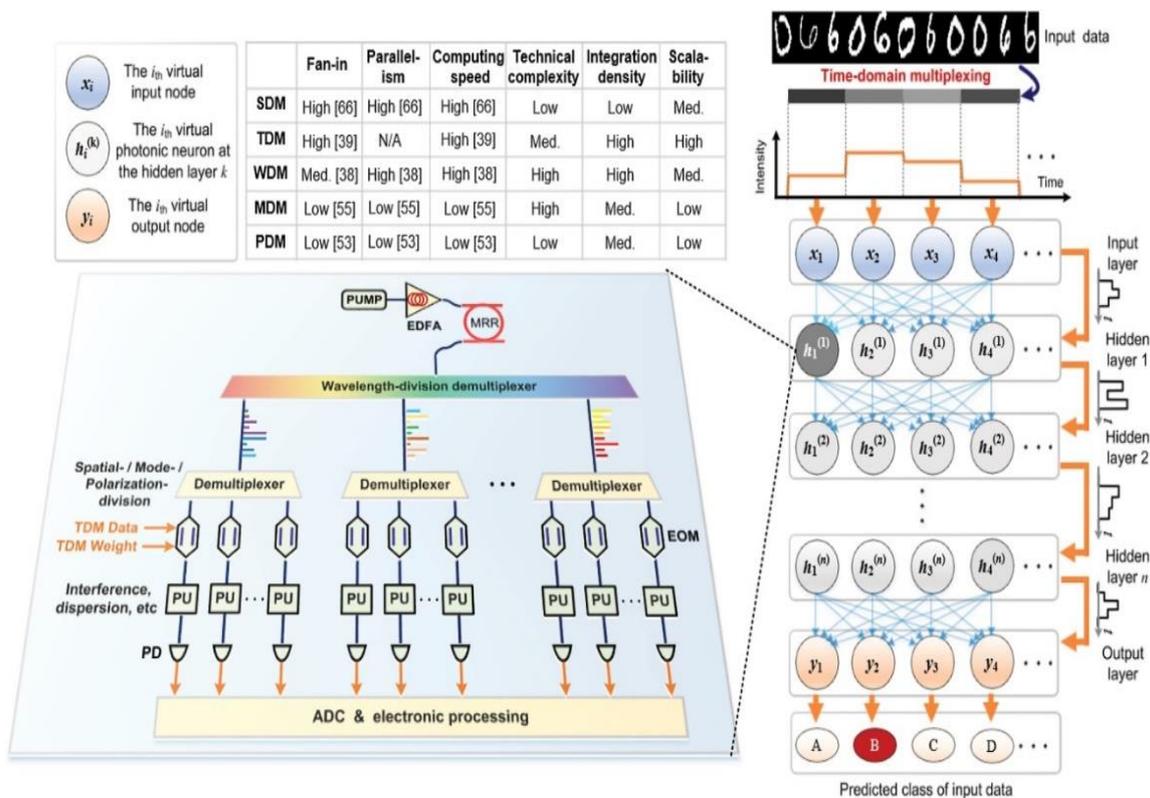


Рис. 4. Схема оптической нейронной сети, использующей гибридное мультиплексирование [8]; обозначения: EOM – электрооптический модулятор, PU – процессор, PD – фотодетектор, ADC – АЦП.

1.4. Аналоговые оптические вычисления и нелинейные фотонные архитектуры

Аналоговые оптические вычисления представляют собой принципиально иной подход к реализации математических операций, предполагающий выполнение вычислительных процедур как прямое следствие взаимодействий света и вещества. Вместо использования дискретных цифровых логических вентилей аналоговые оптические схемы позволяют управлять амплитудой, фазой и поляризацией света для реализации таких операций, как умножение матриц, операция свертки и преобразование Фурье. Дэн и др. [10-11] классифицируют аналоговые оптические вычислительные платформы по четырем основным категориям: платформы когерентной интеграции, платформы некогерентной интеграции, платформы пространственного распространения и платформы оптоволокна. Каждая из этих платформ поддерживает различные вычислительные модели, включая оптические нейронные сети (ONN), оптические резервуарные вычисления и оптические машины Изинга [12], предлагая определенные преимущества в зависимости от приложения.

Платформы когерентной интеграции особенно подходят для выполнения линейных операций, таких как умножение матрицы на вектор и преобразование Фурье, благодаря их способности использовать эффекты интерференции. Эти платформы используют фазово-когерентные источники света, такие как лазерные диоды или частотно-стабилизированные микрогребенки, для генерации высокостабильных оптических полей, которые взаимодействуют внутри волноводных структур или оптических компонентов свободного пространства. Полученные интерференционные картины кодируют желаемые вычисления, причем амплитуда и фаза выходных полей представляют результаты вычислений. Этот подход очень эффективен для выполнения матричных операций, поскольку несколько входных векторов могут быть одновременно преобразованы с использованием одного когерентного источника света. Точность этих вычислений определяется свойствами когерентности источника

света и стабильностью оптических компонентов, что делает критически важным минимизацию фазового шума и температурных колебаний в системе.

Напротив, некогерентные интеграционные платформы используют интенсивность света, а не его фазу для кодирования информации, что делает их более устойчивыми к колебаниям параметров окружающей среды. Эти платформы идеально подходят для реализации нелинейных операций, таких как порогово-логическая обработка, функции активации и нелинейные преобразования, которые необходимы для моделей нейронных сетей. Некогерентные системы часто полагаются на оптические нелинейности, такие как эффект Керра, двухфотонное поглощение или дисперсия свободных носителей, для достижения требуемой нелинейности. Эти нелинейности обычно реализуются с использованием таких материалов, как кремний, фосфид индия или халькогенидные стекла, которые демонстрируют сильную оптическую нелинейность на «телекоммуникационных» длинах волн. Например, эффект Керра, который приводит к изменению показателя преломления, зависящему от интенсивности падающего электрического поля, может использоваться для реализации полностью оптических переключателей, логических вентилях и модуляторов. Эти устройства имеют решающее значение для создания реконфигурируемых фотонных схем, которые могут адаптироваться к различным вычислительным задачам в реальном времени.

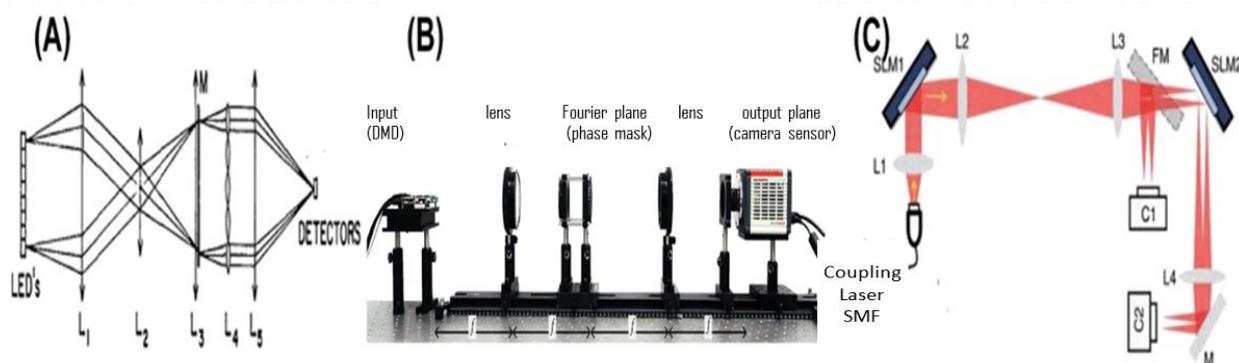


Рис. 5. Развитие ONN основанных на SDM:
 (А) первая система мультиплексирования [9],
 (В) выполнение операции конволюции на основе 4-х линз [10],
 (С) полностью оптическая нелинейная нейронная сеть [11].

Платформы пространственного мультиплексирования (SDM) (рис. 5), характерным примером которых являются космические информационно-телекоммуникационные платформы, используют оптику свободного пространства для реализации сложных пространственных преобразований, включая передачу и преобразование больших информационных потоков. В этих платформах используются линзы, зеркала, дифракционные решетки и пространственные модуляторы света для управления пространственным профилем световых пучков, что позволяет выполнять свертки и другие пространственно-зависимые операции [9-11]. Одним из распространенных применений платформ космического базирования являются оптические корреляторы, которые используются для распознавания образов и обработки изображений. В оптическом корреляторе входное изображение кодируется в пространственный модулятор света, который модулирует амплитуду и фазу лазерного луча. Затем модулированный луч проходит через линзу или систему линз, которая выполняет двумерное преобразование Фурье обрабатываемого изображения для последующего анализа и обработки.

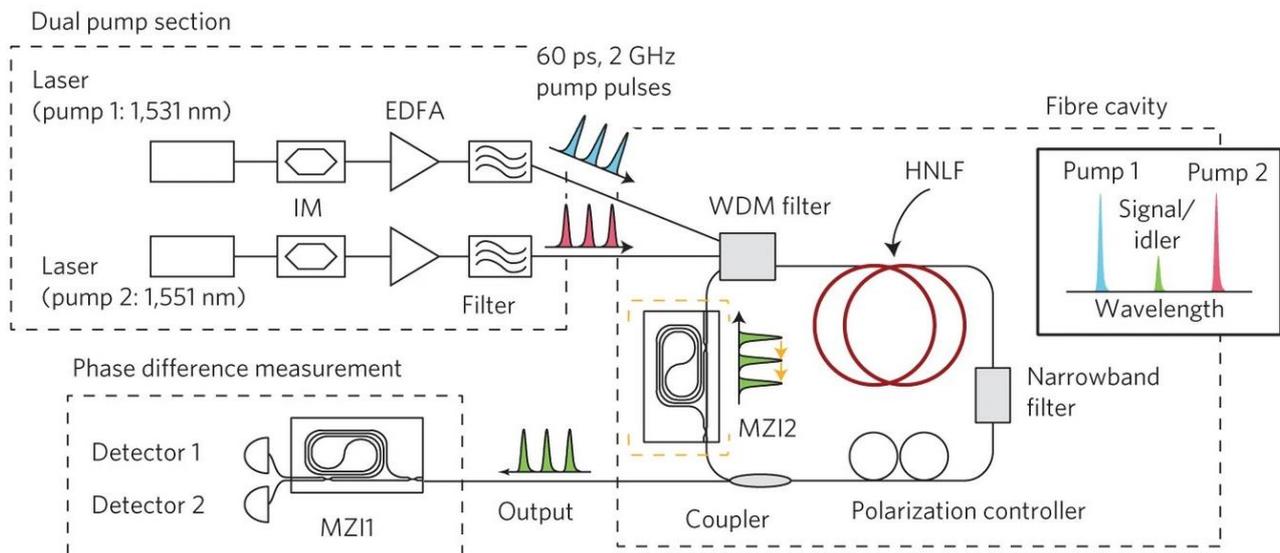


Рис. 6. Пример оптического резервуара, использующего нелинейное оптическое волокно (HNLF) [12].

Выходные данные оптических корреляторов, важнейшего компонента космических телекоммуникационных платформ, могут быть дополнительно обработаны для идентификации и классификации характерных признаков

в сложных визуальных входных данных. Оптические корреляторы работают, выполняя свертку входного изображения с набором predetermined шаблонов, хранящихся в плоскости Фурье, где производится сравнение компонентов пространственной частоты изображения и шаблонов. Эта свертка может использоваться для обнаружения определенных шаблонов, таких как края, углы или текстуры, которые затем подаются в нижестоящие нейронные сети для анализа более высокого уровня. Распределение выходной интенсивности коррелятора, которое представляет собой корреляцию между входным изображением и шаблонами, может быть получено с помощью камеры или матрицы фотодетекторов и обработано в реальном времени. Последние достижения в области пространственных модуляторов света (SLM) и фазовых голограмм позволили создать реконфигурируемые оптические корреляторы, которые могут динамически обновлять свои шаблоны на основе обратной связи в реальном времени, что делает их подходящими для адаптивных систем распознавания изображений.

Платформы на основе волокон, напротив, используют уникальные свойства оптических волокон, такие как дисперсия и нелинейность, для выполнения временных и спектральных преобразований [12-14]. Эти платформы особенно хорошо подходят для распознавания временных образов, где входной сигнал представляет собой изменяющуюся во времени форму волны, которую необходимо классифицировать или анализировать. Тщательно проектируя дисперсию и нелинейность волокна, платформа может реализовывать сложные временные фильтры, которые преобразуют входной сигнал в многомерное представление, подходящее для моделей машинного обучения. Одним из заметных применений платформ на основе волокон является вычисление с использованием технологии «оптического резервуара» (рис 6). Оптический резервуар состоит из длинной нелинейной волоконной петли, которая отображает входной сигнал в многомерное пространство состояний [12]. Это многомерное преобразование позволяет, помимо решения задач распознавания объектов, эффективно классифицировать и прогнозировать

временные шаблоны, такие как речевые сигналы, данные временных рядов и хаотические последовательности. Затем ответ резервуара считывается с помощью набора фотодетекторов и обрабатывается линейным классификатором для получения конечного вывода.

1.5. Оптические нейронные сети и нейроморфные вычисления

Фотонные нейронные сети (PNN) используют свойства света для реализации нейроподобных операций, обеспечивая сверхбыстрые вычисления с низким энергопотреблением для задач машинного обучения. Разработка однофотонных персептронов, как продемонстрировали Мэнси Тан и др. [15-16], представляет собой значительный прогресс в технологии PNN. Их конструкция использует источник в виде солитонного кристалла микрогребенки для одновременной генерации нескольких оптических частот, обеспечивая компактные и энергоэффективные нейроподобные операции. Микрогребенка генерирует ряд дискретных, равномерно распределенных спектральных линий, каждая из которых представляет отдельный нейрон в сети. Модулируя амплитуду и фазу каждой линии, система может реализовывать широкий спектр линейных и нелинейных преобразований, что позволяет строить крупномасштабные фотонные нейронные сети с минимальной аппаратной сложностью.

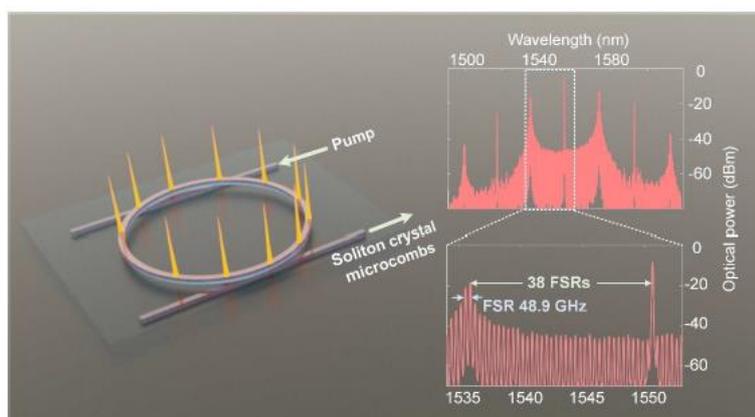


Рис. 7. Солитонный кристалл типа оптической микрогребенки и спектр генерируемого оптического сигнала. Солитонный кристалл генерируется в интегральном 4-х портовом кольцевом микрорезонаторе (MRR) с межмодовым интервалом 49 ГГц [17-18].

Солитонные микрогребенки генерируются с использованием нелинейных микрорезонаторов, поддерживающих стабильные солитонные состояния, в которых баланс между дисперсией и нелинейностью приводит к образованию локализованных импульсов. Эти импульсы циркулируют внутри микрорезонатора, генерируя широкий спектр частот посредством четырехволнового смешения (см. рис. 7). Каждый частотный компонент микрогребенки может быть независимо модулирован и направлен с использованием селективных по длине волны компонентов, таких как микрокольцевые резонаторы или настраиваемые фильтры, что позволяет реализовать сложные архитектуры нейронных сетей на одном чипе (см. рис. 8). Использование солитонных микрогребенок значительно повышает масштабируемость фотонных нейронных сетей, поскольку одна микрогребенка может заменить несколько дискретных лазерных источников, уменьшая размер, энергопотребление и сложность системы. Более того, солитонные микрогребенки демонстрируют высокую стабильность и когерентность, гарантируя, что различные частотные каналы остаются фазово-синхронизированными и свободными от перекрестных помех даже в крупномасштабных реализациях.

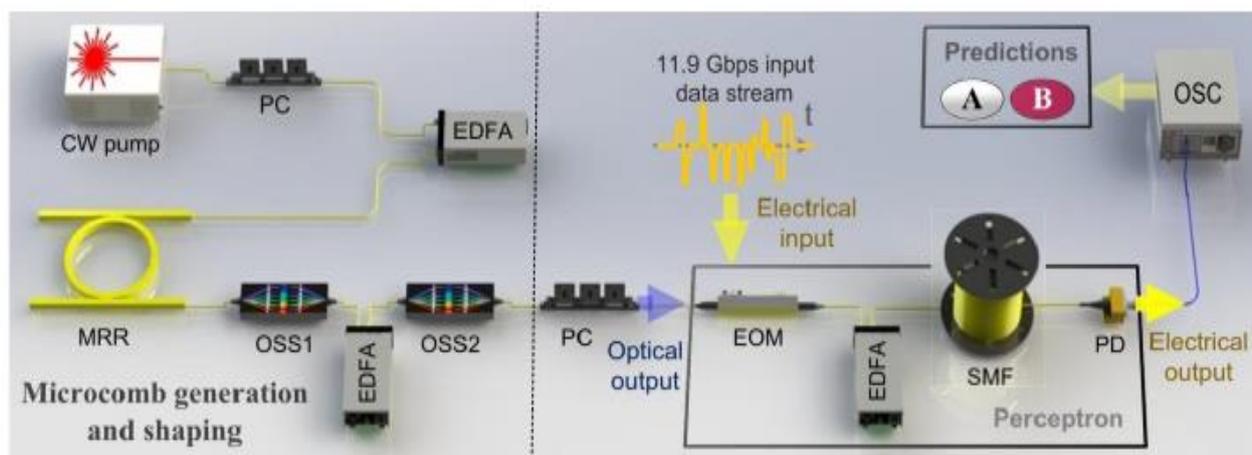


Рис. 8. Экспериментальная реализации оптического персептрона, обозначения:
 CW pump – непрерывный лазер; PC – контролер поляризации;
 EDFA – эрбиевый волоконный усилитель; MRR – кольцевой микрорезонатор;
 OSS – оптический формирователь спектра; EOM – электрооптический модулятор Маха-Цендера; SMF – одномодовое волокно; PD – фотодетектор;
 OSC – высокоскоростной осциллограф [20-23].

Ключевым преимуществом фотонных нейронных сетей является их способность реализовывать как линейные, так и нелинейные операции в оптической области, что позволяет строить глубокие нейронные сети (DNN) с несколькими слоями взаимосвязанных нейронов. Линейные операции, такие как умножение матриц, обычно выполняются с использованием массивов MZM или MRR, в то время как нелинейные операции реализуются с использованием оптических нелинейностей, таких как двухфотонное поглощение, перекрестная фазовая модуляция или нелинейные оптические кристаллы. Например, двухфотонное поглощение в полупроводниковом волноводе может использоваться для реализации нелинейной функции активации путем модуляции выходной интенсивности на основе квадрата входной мощности. Такой подход позволяет напрямую реализовывать нелинейности в оптической области, устраняя необходимость в электронной обработке и снижая общее энергопотребление сети.

В дополнение к стандартным архитектурам нейронных сетей фотонные системы также хорошо подходят для реализации нейронных сетей со спайками (SNN), создание которых во многом связано с поведением биологических нейронов. SNN используют дискретные спайки для кодирования информации, при этом время и частота спайков являются носителями вычислительной информации [24-25]. Фотонные реализации SNN используют сверхбыстрые оптические импульсы для представления спайков, при этом фотонные нейроны и синапсы построены с использованием нелинейных элементов, таких как насыщающиеся поглотители, бистабильные лазерные диоды и резонансные туннельные диоды (см. рис. 9). Эти компоненты позволяют генерировать, модулировать и передавать оптические спайки, что позволяет строить крупномасштабные нейронные сети со спайками, которые работают на гигагерцовых частотах. SNN особенно эффективны для решения задач, требующих распознавания временных образов, таких как обработка речи, анализ видео и слияние датчиков, где точная синхронизация событий имеет решающее значение.

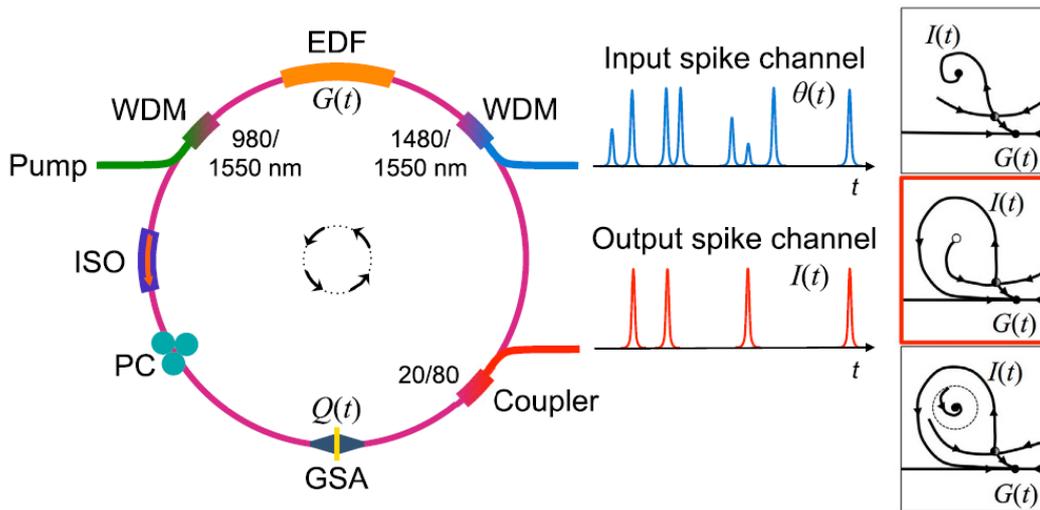


Рис. 9 Полностью оптический волоконный нейрон на основе графена в качестве насыщающегося поглотителя [24].

1.6. Вычисления на основе фотонных резервуаров

Вычисления на основе резервуаров фотонов (PRC) – это новая нейроморфная архитектура, которая использует динамические свойства оптических резервуаров для выполнения многомерных преобразований входных данных. В отличие от традиционных нейронных сетей, где веса каждого слоя обучаются по методу градиентного спуска, системы резервуарных вычислений состоят из фиксированного резервуара со случайными соединениями; за этими соединениями следует слой считывания, который обучается с помощью простой линейной регрессии. Эта архитектура радикально упрощает процесс обучения, поскольку необходимо настроить только веса считывания, что делает ее идеальной для аппаратных реализаций. В PRC входной сигнал вводится в резервуар, который состоит из сети связанных нелинейных узлов, таких как полупроводниковые лазеры, оптические волокна или микрокольцевые резонаторы. Реакция резервуара – это многомерное нелинейное преобразование входных данных, которое можно использовать для классификации сложных шаблонов, прогнозирования данных временных рядов или решения задач оптимизации.

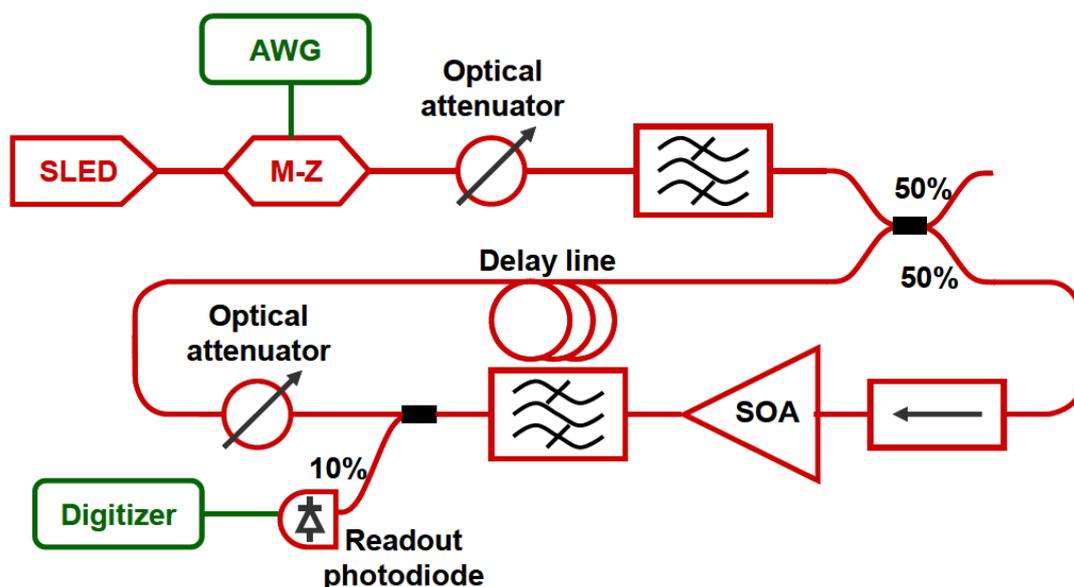


Рис. 10. Реализация фотонного резервуара на основе волоконной линии задержки [26].

Одна из наиболее широко используемых реализаций фотонных резервуаров основана на полупроводниковых лазерах с задержкой (см. рис. 10). В этих системах одномодовый полупроводниковый лазер работает в условиях наличия оптической обратной связи с задержкой, реализованной с помощью волоконного контура. Благодаря наличию контура обратной связи, создается ряд запаздывающих во времени реплик входного сигнала. Динамика лазера, которая описывается уравнениями Ланга-Кобаяши [27-28], генерирует многомерное пространство состояний, чувствительное к небольшим изменениям на входе. Регулируя длину задержки, глубину обратной связи и ток инжекции, резервуар можно настроить на демонстрацию широкого спектра динамических поведений, от периодических колебаний до хаотических режимов. Эта динамика используется для выполнения сложных вычислений, при этом выход резервуара считывается с помощью фотодетектора и обрабатывается линейным классификатором. Резервуары с задержкой успешно применялись для таких задач, как распознавание произнесенных цифр, прогнозирование временных рядов и реконструкция хаотических сигналов, демонстрируя свою универсальность и вычислительную мощность.

Другим перспективным подходом к созданию PRC является использование интегрированных фотонных схем на основе микрокольцевых резонаторов или волноводных решеток. В этих системах резервуар состоит из сети связанных резонаторов, где коэффициенты связи и резонансные частоты случайным образом распределены для создания разнообразного набора шаблонов отклика. Входной сигнал вводится в один или несколько узлов сети, и полученные интерференционные картины обнаруживаются на различных выходных узлах. Интерференционные картины кодируют многомерное преобразование входа, которое может использоваться для задач классификации или регрессии. Интегрированные фотонные резервуары обладают высокой масштабируемостью, поскольку количество узлов и соединений можно легко увеличить, добавив больше резонаторов или пересечений волноводов. Более того, использование интегральной фотоники позволяет реализовать реконфигурируемые резервуары, где коэффициенты связи и резонансные частоты можно динамически регулировать с помощью термооптических или электрооптических элементов настройки.

2. Проблемы, возникающие в области фотонных нейроморфных вычислений и направления развития этих технологий

Несмотря на значительный прогресс в области фотонных вычислений, имеет место ряд проблем в реализации полностью интегрированных фотонных систем ИИ. Одним из главных препятствий является разработка эффективных нелинейных компонентов, таких как чисто оптические переключатели и модуляторы, которые могут быть бесшовно интегрированы с линейными фотонными схемами. Нелинейности необходимы для реализации функций активации в нейронных сетях, которые имеют решающее значение для захвата сложных шаблонов в данных.

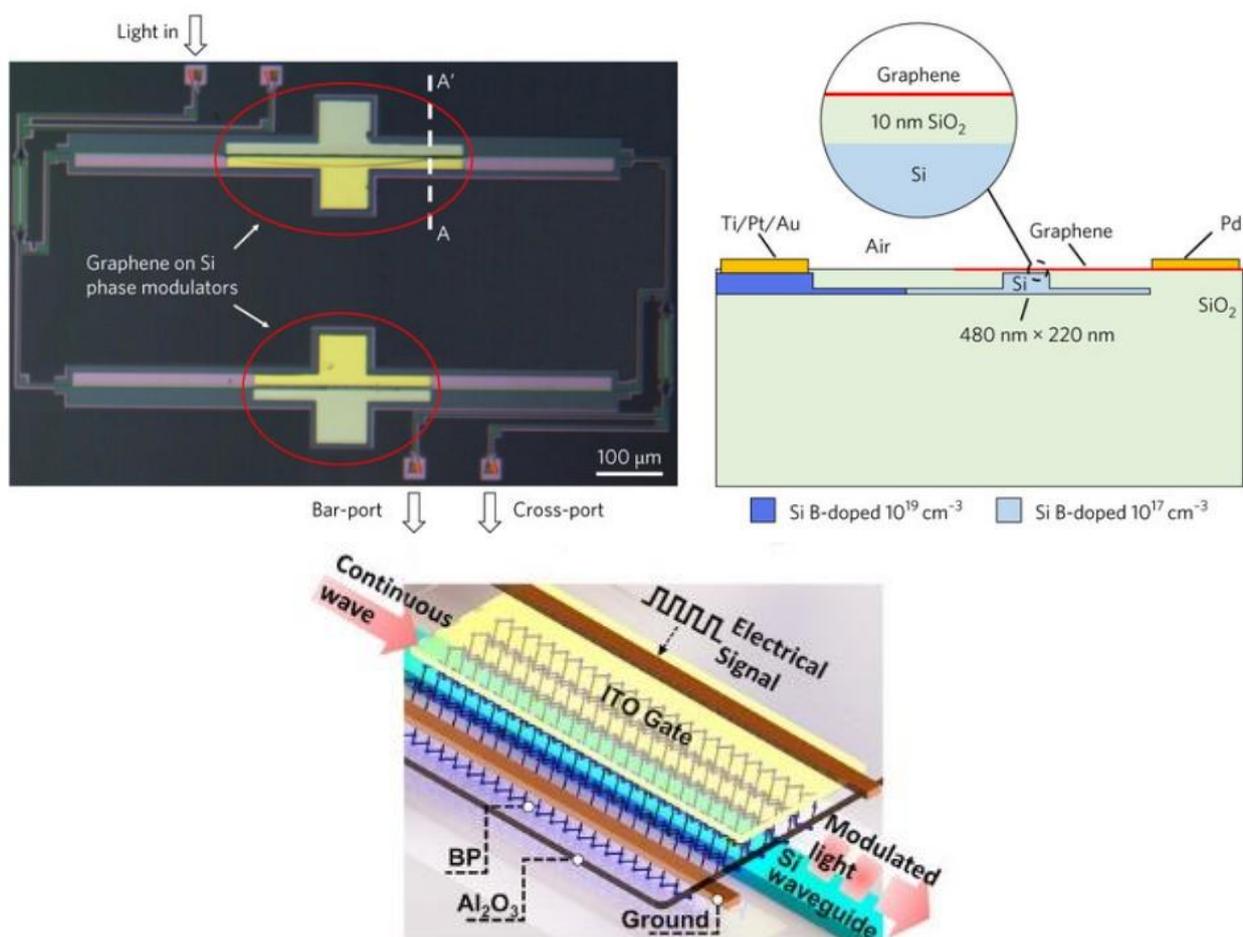


Рис. 11. 2D-интегральный фазовый модулятор на основе структуры графен-изолятор-кремний [29-31].

Текущие нелинейные компоненты часто страдают от высокого энергопотребления и ограниченной скорости, что, в свою очередь, ограничивает масштабируемость фотонных нейронных сетей. Ожидается, что достижения в области материаловедения, такие как разработка маломощных нелинейных оптических материалов и использование нанофотонных структур, решат эти проблемы и позволят построить крупномасштабные фотонные нейронные сети с маломощными высокоскоростными нелинейными элементами (см. рис. 11).

Еще одной проблемой является интеграция оптической памяти, которая может хранить промежуточные результаты и обеспечивать обратную связь в нейронных сетях. В отличие от электронных систем, где память и логика могут быть размещены на одном чипе, фотонные системы часто требуют отдельных компонентов памяти, которые с точки зрения физической реализации отличаются от вычислительных блоков. Такое разделение увеличивает задержку

и энергопотребление, ограничивая потенциальные преимущества фотонных вычислений. Целый ряд недавно проведенных исследований [32-36] был посвящен разработке компактной, высокоскоростной фотонной памяти с использованием таких материалов, как халькогениды с фазовым переходом и ячейки памяти на основе интегрированных резонаторов. Эти технологии обещают достижение фотонной памяти высокой плотности с низкой задержкой, которая может быть интегрирована с фотонными нейронными сетями [32-36].

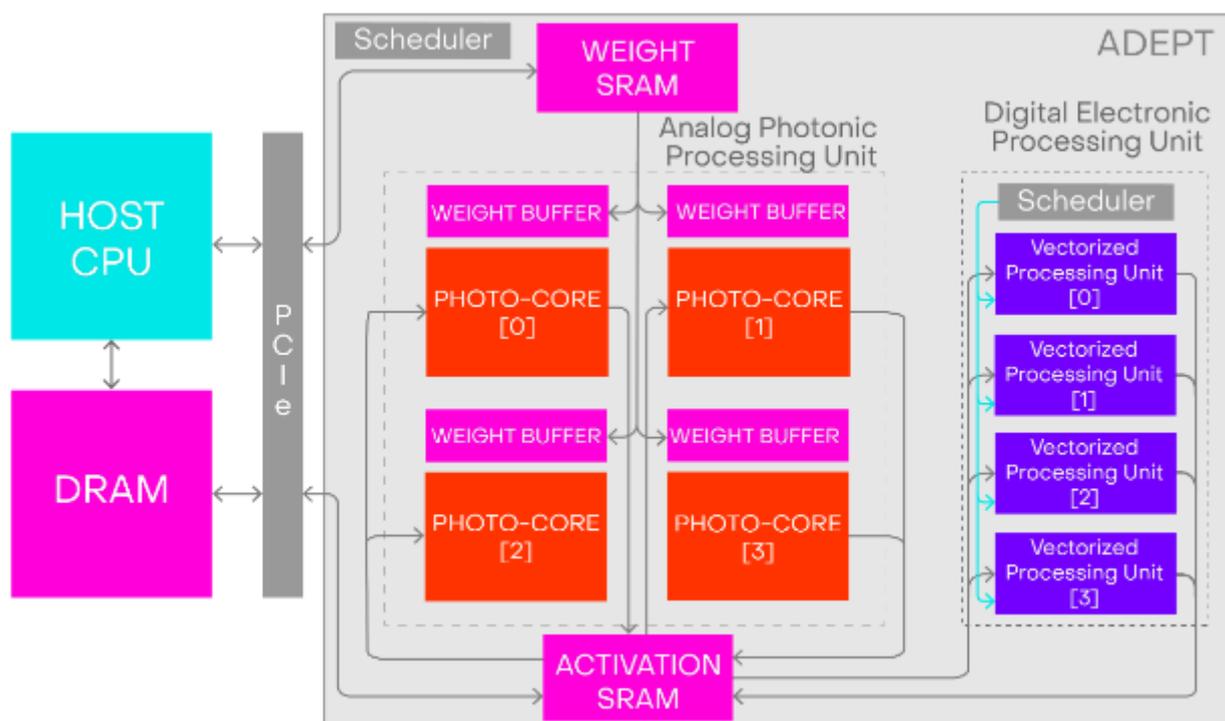


Рис. 12. Схема и функциональные процессы в электрооптической вычислительной системе [37].

Будущее фотонных вычислений представляется в разработке гибридных оптоэлектронных платформ, которые объединяют лучшее из фотоники и электроники (см. рис. 12). Такие платформы сочетают высокую пропускную способность и параллелизм фотонных систем для операций с интенсивным использованием данных, одновременно используя электронные схемы для задач, требующих программируемости и сложного управления. Гибридные архитектуры могут устранить ограничения чисто фотонных систем, такие как отсутствие эффективных нелинейных компонентов и потребность в надежных элементах памяти. Интегрируя фотонные и электронные компоненты на одном

кристалле, эти платформы могут достичь беспрецедентного уровня производительности, открывая новые приложения в области ИИ, высокопроизводительных вычислений и т.д.

3. Достижения в области фотонных интегральных схем и их роль в аппаратном обеспечении высокопроизводительных вычислений, включая ИИ

Быстрое развитие фотонных интегральных схем (ФИС) сыграло решающую роль в развитии архитектур фотонных вычислений, предоставив универсальную платформу для интеграции нескольких оптических и электронных компонентов на одном чипе. Кремниевая фотоника, в частности, стала перспективным решением для изготовления крупномасштабных фотонных схем высокой плотности с использованием хорошо зарекомендовавшего себя процесса изготовления структур типа «комплементарный металл-оксид-полупроводник» (КМОП). Следует отметить, что в последнее время был реализован целый ряд исследований, посвященный улучшению производительности и функциональности ключевых фотонных компонентов, таких как источники света на кристалле, модуляторы, волноводы и фотодетекторы, которые образуют фундаментальные строительные блоки ФИС.

3.1. Разработка эффективных источников света

Одной из основных проблем в кремниевой фотонике является разработка эффективных источников света на кристалле, поскольку кремний является материалом с непрямой запрещенной зоной и, таким образом, не может эффективно излучать свет. Чтобы устранить это ограничение, исследователи изучили различные материалы и методы, такие как интеграция полупроводников III-V на кремниевых подложках, использование совместимых с кремнием излучателей света, таких как кремниевые нанокристаллы, и применение нелинейных оптических процессов, таких как комбинационное рассеяние. Интеграция полупроводников III-V, таких как фосфид индия (InP) или арсенид галлия (GaAs),

на кремнии оказалась одним из наиболее эффективных решений, позволяющих изготавливать высокопроизводительные лазеры непосредственно на кремниевых чипах. Эти гибридные лазеры III-V/Si демонстрируют высокую выходную мощность, низкие пороговые токи и широкую перестройку длины волны, что делает их идеальными для использования в системах WDM (см. рис. 13 [38]).

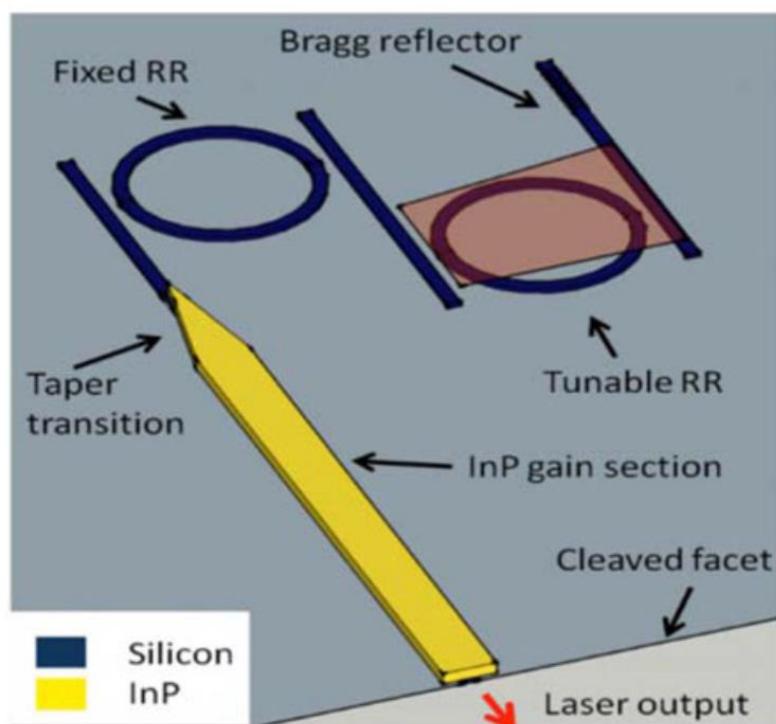


Рис. 13. Схема гибридного III-V/Si лазера [38].

Другой перспективный подход предполагает использование кремниевых нанокристаллов и кремния, легированного эрбием (Er), в качестве светоизлучающих материалов. Кремниевые нанокристаллы формируются путем имплантации ионов кремния в матрицу диоксида кремния, создавая наномасштабные кластеры, которые демонстрируют эффективную фотолюминесценцию из-за эффектов квантового ограничения. Эти нанокристаллы могут использоваться для изготовления светодиодов (LED) и усилителей, которые полностью совместимы со стандартными КМОП-процессами, что позволяет интегрировать источники света с существующими электронными схемами. С другой стороны, легирование эрбием позволяет использовать линию излучения 1,54-мкм ионов эрбия, которая лежит в телекоммуникационном диапазоне длин волн. Легируя кремниевые волноводы или резонаторы эрбием, исследователи

продемонстрировали усилители и лазеры на кристалле с умеренным усилением и низким уровнем шума, обеспечивая совместимое с кремнием решение для генерации и усиления света (см. рис. 14).

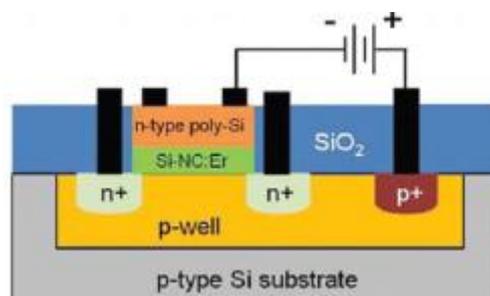


Рис. 14. Поперечное сечение Si-NC: Er LED [39].

3.2. Современные и перспективные модуляторы света

Помимо источников света, важную роль в управлении потоком оптических сигналов в ФИС играют модуляторы. Модуляторы на основе кремния, такие как интерферометрические MZM и MRR-модуляторы, достигли значительного улучшения производительности в последние годы, обеспечивая, в частности, в телекоммуникационных приложениях скорость передачи данных, превышающую 100 Гбит/с на канал. Принцип работы модуляторов основан на управлении изменениями показателя преломления кремния с помощью эффекта плазменной дисперсии, когда свободные носители инжектируются или истощаются в волноводе, чтобы вызвать соответствующие изменения эффективного показателя преломления. Это приводит к изменению фазы распространяющегося света, обеспечивая высокоскоростную модуляцию оптического сигнала. Основными ограничениями кремниевых модуляторов являются их относительно высокое энергопотребление и ограниченная электрооптическая полоса пропускания. Чтобы преодолеть эти проблемы, исследователи рассматривают альтернативные материалы, такие как ниобат лития (LiNbO_3), графен и кремний-органические гибридные (SOH) материалы, которые обеспечивают высокую эффективность модуляции и более быстрое время отклика.

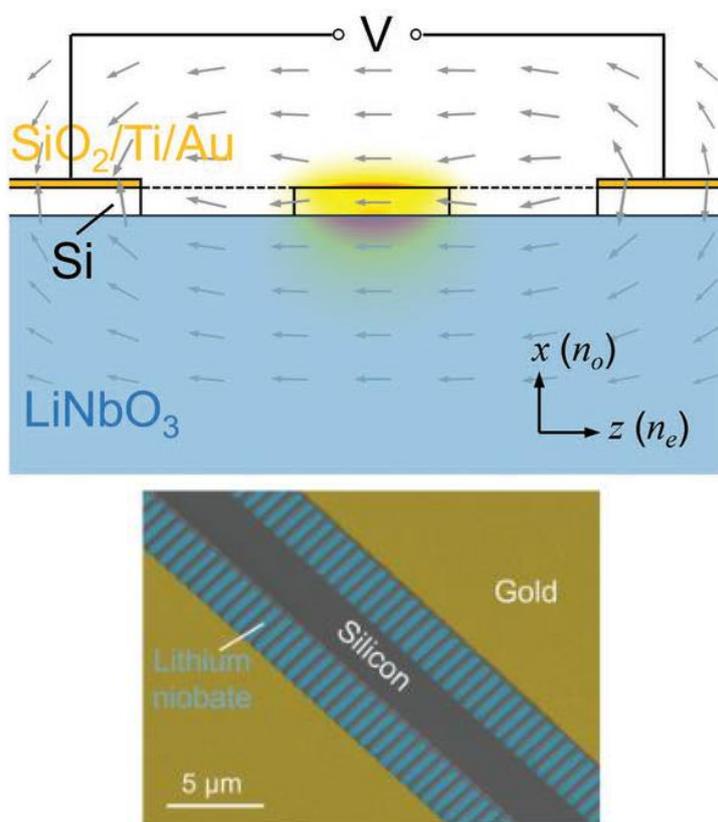


Рис. 15 Гибридный электрооптический модулятор на основе Si-on-LiNbO₃ [40].

Например, модуляторы на основе ниобата лития известны своим высоким электрооптическим коэффициентом и низкими вносимыми потерями, что делает их идеальными для внедрения высокоскоростных маломощных модуляторов в фотонные схемы. Недавние прорывы в технологии тонкопленочного ниобата лития (TFLN) позволили изготовить компактные интегрированные модуляторы с полосой пропускания, превышающей 100 ГГц. Эти модуляторы изготавливаются путем присоединения тонких слоев LiNbO₃ к кремниевым или сапфировым подложкам, что обеспечивает плотное оптическое ограничение и эффективную электрооптическую модуляцию. Аналогичным образом, модуляторы на основе графена используют уникальные электронные свойства графена, такие как его высокая подвижность носителей и перестраиваемый уровень Ферми, для достижения сверхбыстрой модуляции с минимальным потреблением энергии. Способность графена сильно взаимодействовать со светом в широком диапазоне длин волн делает его привлекательным материалом для широкополосных модуляторов и фотодетекторов.

3.3. Датчики на основе оптических волноводов и фотонные нейроморфные компоненты

Волноводы образуют основу фотонных интегральных схем, позволяя передавать световые потоки между различными компонентами с низкими потерями и высокой эффективностью. Последние достижения в области проектирования волноводов были сосредоточены на минимизации потерь распространения, оптимизации ограничения мод и обеспечении сложной геометрии маршрутизации. Волноводы типа «кремний на изоляторе» (SOI), которые состоят из кремниевой сердцевины с высоким показателем преломления, окруженной оболочкой из кремния с меньшим (низким) показателем преломления, обеспечивают эффективное ограничение мод и низкие потери распространения на телекоммуникационных длинах волн (см. рис. 16). Для дальнейшего снижения потерь исследователи разработали передовые методы изготовления, такие как химико-механическая полировка (CMP) и литография глубокого УФ-излучения, которые позволяют точно контролировать размеры волновода и шероховатость поверхности.

(a) Smart Ring Resonator Sensor

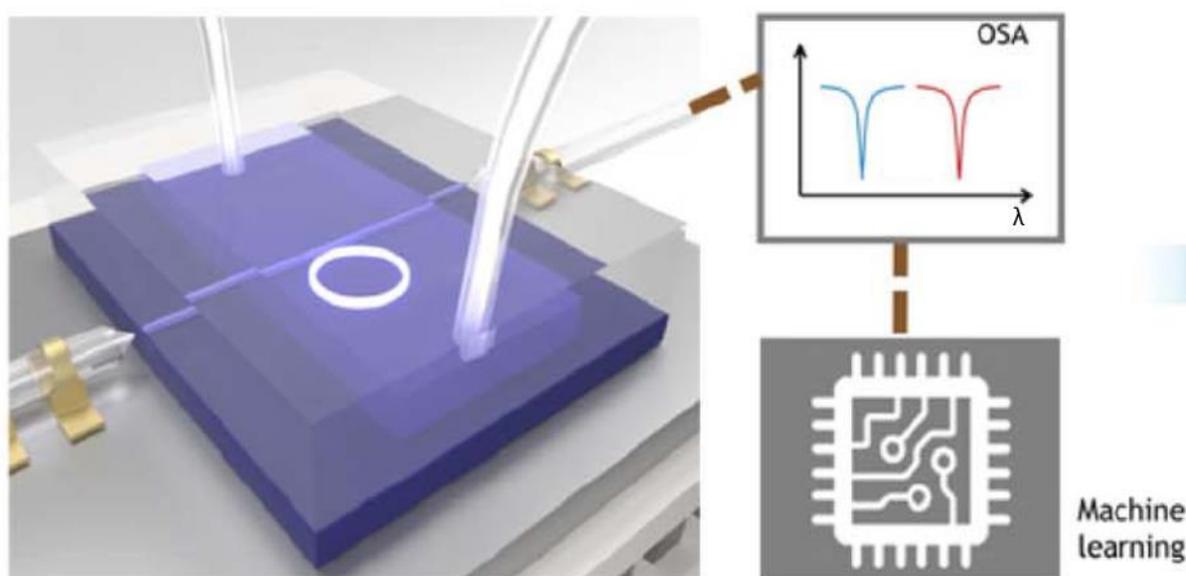


Рис. 16. «Умный» кольцевой резонатор [41].

В дополнение к стандартным одномодовым волноводам, многомодовые и фотонно-кристаллические волноводы были исследованы на предмет их способности поддерживать несколько пространственных мод и выполнять сложные операции обработки сигналов. Многомодовые волноводы могут использоваться для реализации схем мультиплексирования с разделением мод (MDM), где каждая пространственная мода служит независимым каналом для передачи информации. Объединив MDM с WDM и TDM, исследователи смогли реализовать сверхвысокопроизводительные системы связи с совокупной скоростью передачи данных, превышающей 1 Тбит/с. Фотонно-кристаллические волноводы, с другой стороны, используют периодические изменения показателя преломления для создания фотонных запрещенных зон, которые ограничивают и манипулируют светом способами, невозможными при использовании обычных волноводов. Эти волноводы используются для реализации компактных оптических фильтров, переключателей и линий задержки, которые необходимы для построения интегрированных фотонных схем.

Датчики на основе оптических волноводов также привлекают значительное внимание исследователей, благодаря своим применениям в химическом и биологическом зондировании. Вследствие функционализации поверхности кремниевого волновода с помощью определенного рецептора устройство может обнаруживать изменения показателя преломления или поглощающих свойств окружающей среды, что позволяет обнаруживать биомолекулы, газы и другие аналиты без использования меток. Эти датчики обычно реализуются с использованием кольцевых резонаторов, интерферометров Маха-Цендера или фотонных кристаллических полостей, которые демонстрируют высокую чувствительность к изменениям в локальной среде. Последние достижения в области проектирования датчиков были сосредоточены на повышении чувствительности и селективности этих устройств с помощью таких методов, как поверхностный плазмонный резонанс (SPR), щелевые волноводные структуры и покрытия метаповерхностей (см. рис. 16).

Одно из наиболее перспективных применений фотонных схем на основе волноводов – нейроморфные вычисления, в которых фотонные схемы используются для реализации синаптических весов, функций активации и взаимосвязей между фотонными нейронами. Синаптические веса могут быть реализованы с использованием настраиваемых оптических аттенуаторов, таких как микрокольцевые резонаторы или переменные оптические соединители, которые модулируют интенсивность света, проходящего через синапс. Динамически регулируя коэффициент связи или резонансную частоту этих компонентов, синаптический вес может быть настроен в реальном времени, что обеспечивает адаптивное обучение и пластичность в фотонных нейронных сетях. Фотонные нейроны, которые имитируют поведение биологических нейронов, обычно реализуются с использованием нелинейных оптических элементов, таких как насыщаемые поглотители, оптические бистабильные устройства или резонансные туннельные диоды. Эти компоненты генерируют оптические пики в ответ на входящие сигналы, что позволяет строить крупномасштабные пиковые (спайковые) нейронные сети (SNN), которые могут выполнять сложные временные и пространственные задачи распознавания образов [42].

3.4. Интегрированные фотодетекторы и фотонные системы считывания

Фотодетекторы являются критически важными компонентами в PIC, отвечающими за преобразование оптических сигналов в электрические сигналы для взаимодействия с электронными схемами. Кремниевые фотодетекторы, такие как p-i-n фотодиоды и лавинные фотодиоды (APD), широко используются из-за их высокой чувствительности, низкого уровня шума и совместимости с КМОП-процессами. P-i-n фотодиоды работают, генерируя пары «электрон-дырка», когда фотоны поглощаются во внутренней области устройства, создавая фототок, пропорциональный интенсивности падающего света. APD, с другой стороны, используют ударную ионизацию для достижения внутреннего усиления, что приводит к более высокой чувствительности и более быстрому времени

отклика по сравнению с обычными фотодиодами. Однако кремниевые фотодетекторы ограничены длинами волн менее 1,1 мкм из-за ширины запрещенной зоны кремния. Для расширения диапазона обнаружения до более длинных волн в кремний интегрируются такие материалы, как германий (Ge) и арсенид индия-галлия (InGaAs), что позволяет осуществлять фотодетектирование на телекоммуникационных длинах волн (1,3-1,55 мкм) (см. рис. 17).

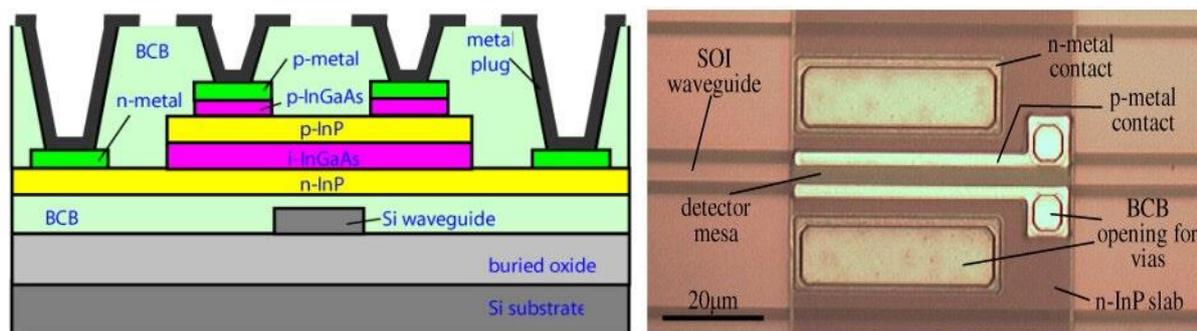


Рис. 17. Поперечное сечение и внешний вид InGaAs PIN-фотодиода, интегрированного с SOI-волноводом [43].

Интегрированные фотодетекторы необходимы для реализации оптических систем считывания в фотонных нейронных сетях и оптических ячейках памяти. В фотонных нейронных сетях выход каждого нейрона или синапса должен быть преобразован в электрический сигнал для дальнейшей обработки или для сопряжения с электронными схемами. Обычно это достигается с помощью массивов фотодиодов или фототранзисторов, которые интегрированы в выходные волноводы сети. Чтобы минимизировать перекрестные помехи и ухудшение сигнала, в современных конструкциях фотодетекторов используются такие методы, как затухающая связь, когда оптический сигнал поглощается затухающим полем, простирающимся от волновода в детектор, и вертикальная интеграция, когда детектор размещается над или под слоем волновода для уменьшения площади схемы.

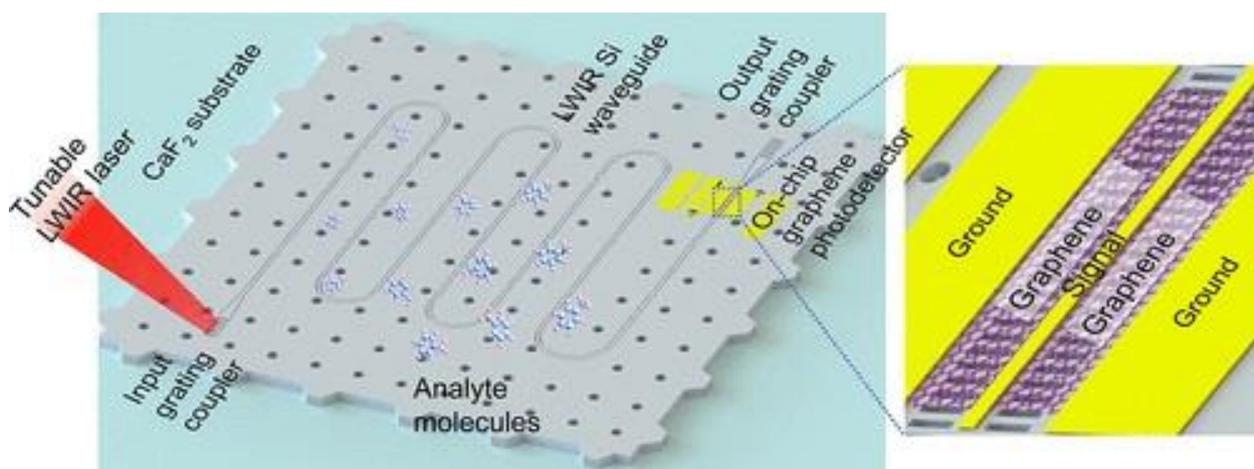


Рис. 18. Схема волновода на основе структуры графен/кремний/галид [44].

Недавние достижения в технологии фотодетекторов были сосредоточены на улучшении скорости, чувствительности и плотности интеграции этих устройств. Одним из перспективных подходов является использование плазмонных фотодетекторов, которые используют поверхностные плазмонные резонансы для концентрации света в наномасштабных объемах, улучшая поглощение и чувствительность детектора (см. рис. 18). Включая плазмонные наноструктуры, такие как металлические нанорешетки или наночастицы, в активную область фотодетектора, исследователи продемонстрировали устройства со временем отклика менее 10 пс и полосой пропускания, превышающей 100 ГГц. Другой подход предполагает использование 2D-материалов, таких как графен, дихалькогениды переходных металлов (TMD) или черный фосфор, которые демонстрируют высокую подвижность, широкополосное поглощение и настраиваемые электронные свойства. Эти материалы используются для изготовления сверхтонких фотодетекторов, совместимых с кремниевыми волноводами, что позволяет разрабатывать высокопроизводительные фотонные схемы с уменьшенной площадью.

3.5. Передовые фотонные запоминающие устройства и чисто оптические логические элементы

Важным компонентом в реализации практических фотонных нейронных сетей является разработка элементов фотонной памяти, которые могут

эффективно хранить и вызывать оптическую информацию. В отличие от обычных электронных систем, где операции памяти и логики могут выполняться с использованием одних и тех же кремниевых транзисторов, фотонные системы требуют специализированных устройств, которые способны сохранять определенные оптические состояния в течение нужного времени или управлять процессами изменения этих состояний. Это отличие возникает из-за различных физических принципов управления электронными и фотонными технологиями хранения информации. Технологии фотонной памяти изучаются с использованием различных подходов, включая хранение на основе резонаторов, материалов с фазовым переходом (PCM) и оптомеханических осцилляторов. Каждая из этих технологий предполагает наличие определенных преимуществ и недостатков в зависимости от области применения.

Резонаторные запоминающие устройства, включающие микрокольцевые резонаторы, фотонные кристаллические полости и связанные резонаторные волноводы, используют принцип оптического резонанса для хранения информации. Когда оптический сигнал попадает в резонатор, он циркулирует внутри полости, создавая стоячую волну, которая может сохраняться в течение времени, определяемого добротностью резонатора (Q). Высокодобротные резонаторы, которые демонстрируют низкие потери, могут хранить свет в течение сотен пикосекунд, что делает их пригодными для краткосрочного хранения и буферизации. Сохраненный свет можно считывать, настраивая резонансную частоту резонатора с помощью тепловых или электрооптических настроечных элементов. Хотя резонаторные запоминающие устройства обеспечивают высокую скорость работы и компактный размер, они обладают некоторыми недостатками, к которым, в частности, можно отнести ограниченное время хранения информации и высокую чувствительность к колебаниям параметров окружающей среды.

Материалы с фазовым переходом (PCM), такие как германий-сурьма-теллурид (GST) и диоксид ванадия (VO_2), обеспечивают альтернативный подход к созданию устройств фотонной памяти, используя способность материала

переключаться между аморфным и кристаллическим состояниями. Этот фазовый переход, который может быть вызван оптически или электрически, приводит к значительному изменению показателя преломления материала, что позволяет хранить двоичную информацию. При интеграции в фотонные волноводы или резонаторы РСМ могут действовать как реконфигурируемые оптические переключатели, модуляторы и элементы памяти. Основным преимуществом РСМ является их энергонезависимость, поскольку материал сохраняет свое фазовое состояние даже при отключении питания. Это свойство делает фотонную память на основе РСМ идеальной для долгосрочного хранения и энергонезависимых нейроморфных вычислений. Недавние достижения продемонстрировали многоуровневую память РСМ, в которой несколько дискретных состояний показателя преломления могут быть достигнуты посредством частичной кристаллизации, что позволяет хранить многобитовую информацию в одной ячейке памяти [49].

Оптомеханические осцилляторы (ОМО) представляют собой третий подход к фотонной памяти, где информация хранится в форме механических колебаний внутри наномеханического резонатора. Эти устройства связывают оптические и механические степени свободы посредством давления излучения, что позволяет хранить оптическую информацию в виде механических колебаний. Когда оптический импульс попадает в оптомеханическую полость, он возбуждает механическую моду, которая колеблется с определенной частотой, эффективно сохраняя оптическую информацию. Механическое состояние можно считывать, посылая зондирующий свет через полость и обнаруживая результирующий фазовый сдвиг, вызванный колеблющимся зеркалом. ОМО предлагают сверхдлительное время хранения и высокую чувствительность, но ограничены своими относительно низкими рабочими частотами и восприимчивостью к тепловому шуму.

Интеграция фотонной памяти с логическими вентилями и другими вычислительными элементами имеет решающее значение для создания полностью оптических процессорных устройств, которые могут выполнять

сложные вычисления без преобразования сигналов обратно в электронную область. Полностью оптические логические вентили, такие как вентили AND, OR, NOT и XOR [45-48], могут быть реализованы с использованием нелинейных оптических эффектов, таких как четырехволновое смешение (FWM), перекрестная фазовая модуляция (XPM) и двухфотонное поглощение. Эти нелинейности обычно достигаются с использованием таких материалов, как кремний, фосфид индия или халькогенидные стекла, которые демонстрируют сильную оптическую нелинейность на телекоммуникационных длинах волн. Например, базовый вентиль AND может быть реализован с использованием FWM в кремниевом волноводе, где два входных сигнала объединяются для генерации третьего сигнала на другой длине волны, представляющего логическую операцию AND. Аналогично, XPM может быть использован для реализации вентиля NOT, где фаза выходного сигнала инвертируется при наличии входного сигнала.

3.6. Новые методы мультиплексирования для крупномасштабных фотонных систем

Для полного использования параллелизма фотонных систем были разработаны передовые методы мультиплексирования, которые используют несколько степеней свободы оптического излучения, таких как длина волны, пространство, время, мода и поляризация. Каждый метод мультиплексирования позволяет кодировать и передавать несколько независимых потоков данных по одному оптическому каналу, значительно увеличивая емкость данных и вычислительную мощность фотонных систем.

Мультиплексирование с разделением по длине волны является одним из наиболее широко используемых методов мультиплексирования, в котором разные длины волн (или цвета) света используются для передачи отдельных потоков данных. WDM обычно реализуется с использованием массивов mRR или тонкопленочных фильтров, которые осуществляют разделение излучения по различным длинам волн и ввод этого модифицированного излучения в отдельные волноводы. Объединив WDM с временным мультиплексированием

(TDM), исследователи продемонстрировали системы, способные обрабатывать сотни входных каналов одновременно, достигая совокупной скорости передачи данных, превышающей 10 Тбит/с). Одной из ключевых проблем в системах WDM является точный контроль интервала между длинами волн и стабильности, поскольку небольшие изменения температуры или дефекты изготовления могут привести к перекрестным помехам между соседними каналами [50-72].

Мультиплексирование с разделением мод (MDM) – еще один эффективный метод, который использует различные пространственные моды в волноводе для кодирования независимых потоков данных. В системах MDM специально разработанные многомодовые волноводы используются для поддержки нескольких пространственных мод, каждая из которых может переносить отдельный канал данных. Моды могут быть выборочно возбуждены и объединены с помощью преобразователей мод и мультиплексоров, которые обычно реализуются с использованием конических волноводов, многомодовых интерференционных (MMI) ответвителей или фазовых решеток. MDM обеспечивает значительное увеличение емкости данных без необходимости использования дополнительных волноводов, что делает его идеальным для плотных фотонных схем и систем связи [73]. Однако основным ограничением MDM являются перекрестные межмодовые помехи, когда мощность передается между соседними модами из-за несовершенств в геометрии волновода или изменений показателя преломления. Чтобы минимизировать перекрестные помехи, исследователи разработали усовершенствованные модовые фильтры и селективные по модам ответвители, которые могут разделять и изолировать отдельные моды с высокой точностью.

Мультиплексирование с поляризационным разделением (PDM) использует состояние поляризации света для кодирования двух независимых потоков данных на одной и той же длине волны. Используя поляризационные делители и вращатели луча, можно модулировать горизонтальное и вертикальное состояния поляризации световой волны по отдельности, удваивая эффективную емкость данных системы. PDM обычно используется в сочетании с WDM и MDM для

достижения сверхвысокой емкости систем связи. Однако состояния поляризации очень чувствительны к колебаниям окружающей среды, таким как изменения температуры и механическое напряжение, которые могут вызывать вращение поляризации и снижать производительность системы. Для решения этой проблемы были разработаны волокна и волноводы, сохраняющие поляризацию, а также активные контроллеры поляризации для стабилизации состояний поляризации и обеспечения надежной передачи данных [73-75].

Объединение этих методов мультиплексирования позволяет создавать гибридные мультиплексные системы, которые могут обрабатывать тысячи независимых каналов одновременно. Например, гибридная система WDM-MDM-PDM может использовать сотни длин волн, каждая из которых несет несколько пространственных мод, а каждая мода несет два состояния поляризации, что приводит к экспоненциальному увеличению емкости данных. Такие гибридные системы изучаются для приложений в высокопроизводительных вычислениях, оптических межсоединениях и нейроморфных вычислениях, где способность обрабатывать большие объемы данных параллельно имеет решающее значение.

3.7. Фотонные нейроморфные вычисления для повышения эффективности решения задач с использованием ИИ

Фотонные нейроморфные вычисления направлены на имитацию структуры и функций человеческого мозга путем внедрения искусственных нейронов и синапсов с использованием фотонных устройств. Этот подход предполагает наличие значительных преимуществ по сравнению с обычными электронными нейронными сетями, включая сверхбыструю скорость обработки, низкое энергопотребление и высокую пропускную способность данных. Фотонные нейроморфные системы обычно реализуются с использованием трех основных компонентов: фотонных нейронов, фотонных синапсов и фотонных соединений [76-87].

Фотонные нейроны – это нелинейные элементы, которые имитируют поведение биологических нейронов, генерируя оптические пики в ответ на входные сигналы. Эти нейроны обычно реализуются с использованием насыщающихся поглотителей, оптических бистабильных устройств или нелинейных резонаторов, которые демонстрируют пороговое поведение, похожее на функции активации в нейронных сетях. Например, насыщающийся поглотитель может использоваться для реализации сигмоидальной функции активации, где выходная интенсивность медленно увеличивается при низких входных мощностях и насыщается при высоких входных мощностях. Аналогично, бистабильные резонаторы могут использоваться для реализации функции выпрямленного линейного блока (ReLU), где выход равен нулю при низких входных мощностях и линейно увеличивается выше определенного порога. Эти функции активации имеют решающее значение для захвата сложных взаимосвязей между входными характеристиками в глубоких нейронных сетях.

Фотонные синапсы – это настраиваемые оптические элементы, которые модулируют силу связи между нейронами. Синапсы могут быть реализованы с использованием оптических аттенюаторов, таких как переменные оптические соединители или микрокольцевые резонаторы, которые контролируют интенсивность света, проходящего через синапс. Динамически регулируя коэффициент связи или резонансную частоту этих элементов, синаптический вес можно настраивать в режиме реального времени, что обеспечивает адаптивное обучение и пластичность в фотонных нейронных сетях. Недавние исследования были сосредоточены на разработке энергонезависимых синапсов с использованием материалов с изменением фазы или оптоэлектронных модуляторов, которые могут сохранять свое состояние даже при отключении питания. Такие энергонезависимые синапсы необходимы для реализации долговременной памяти и снижения энергопотребления фотонных нейронных сетей.

Фотонные соединения связывают нейроны и синапсы, образуя сложные сетевые топологии. Эти соединения обычно реализуются с использованием

волноводных переходов, соединителей и многопортовых интерферометров, которые направляют свет между различными слоями сети. Проблема проектирования фотонных соединений заключается в минимизации потерь и перекрестных помех при сохранении компактности. Для решения этой проблемы исследователи разработали волноводные переходы с низкими потерями и многоуровневые фотонные сети, в которых различные слои сети реализованы на отдельных плоскостях чипа. Эта трехмерная интеграция позволяет создавать плотные соединения с минимальными потерями, что позволяет создавать крупномасштабные фотонные нейронные сети, способные решать сложные задачи ИИ.

4. Прогноз влияния передовых фотонных технологий на аппаратное обеспечение высокопроизводительных вычислений, включая ИИ

Развитие передовых фотонных технологий, по оценкам многих специалистов, может произвести революцию в оборудовании ИИ, позволяя использовать сверхбыстрые, энергоэффективные нейронные сети, которые могут обрабатывать огромные объемы и потоки данных параллельно. Фотонные вычислительные системы предлагают уникальные преимущества с точки зрения скорости, пропускной способности и энергоэффективности, что делает их идеальными для ускорения задач глубокого обучения, распознавания образов и различных видов оптимизации. Однако для полной реализации потенциала фотонных нейроморфных вычислений необходимо решить ряд технических проблем.

4.1. Проблемы и «вызовы» в плане внедрения фотонных нейроморфных вычислений

Несмотря на значительные достижения в области фотонных нейроморфных вычислений, остается несколько ключевых проблем в достижении крупномасштабных, полностью интегрированных фотонных систем, которые могут соответствовать или превосходить свои электронные

аналоги как по производительности, так и по эффективности. Эти проблемы возникают из-за различных факторов, включая сложность интеграции нелинейных компонентов, потребность в стабильных и маломощных источниках света и ограничения, налагаемые текущими технологиями изготовления.

Одним из главных препятствий в проблематике фотонных вычислений является эффективная реализация нелинейностей, которые необходимы для выполнения функций активации в нейронных сетях. В отличие от электронных систем, где транзисторы могут легко реализовывать нелинейные операции, такие как умножение и порогово-логическая обработка, фотонные системы требуют специальных материалов и конфигураций устройств для достижения аналогичных эффектов. Нелинейности в фотонных схемах обычно реализуются с использованием эффекта Керра, двухфотонного поглощения или дисперсии свободных носителей, все эти явления требуют высокой оптической интенсивности для достижения значительных нелинейных откликов. Это приводит к высокому энергопотреблению и ограничивает масштабируемость фотонных нейронных сетей. Недавние исследования были сосредоточены на разработке новых материалов, таких как графен, дихалькогениды переходных металлов (TMD) и органические электрооптические полимеры, которые демонстрируют сильную нелинейность при более низких уровнях мощности. Эти материалы были интегрированы в волноводы и резонаторы для создания компактных маломощных нелинейных устройств, но необходима дальнейшая работа по оптимизации их производительности и интеграции с существующими фотонными схемами.

Еще одним серьезным «вызовом» является разработка стабильных маломощных источников света, которые можно интегрировать непосредственно в кремниевые чипы. Хотя полупроводниковые лазеры типа «III-V» (АЗВ5) обеспечивают высокую эффективность и широкую перестраиваемость длины волны, их интеграция с кремнием остается сложной проблемой из-за большого несоответствия решеток и разницы коэффициентов теплового расширения между материалами. Для преодоления этой проблемы были исследованы

гибридные методы интеграции, такие как прямое склеивание пластин и эпитаксиальный рост, но эти методы часто приводят к высокой плотности дефектов и снижению «выхода годных». Альтернативным подходом является использование совместимых с кремнием источников света, таких как кремниевые рамановские лазеры, кремниевые волноводы, легированные эрбием, или кремний-органические гибридные усиливающие среды. Кремниевые рамановские лазеры, например, используют вынужденное рамановское рассеяние для генерации когерентного света в кремниевом волноводе, что устраняет необходимость во внешних лазерных источниках. Однако эти устройства требуют высокой мощности накачки и ограничены с точки зрения выходной мощности и эффективности.

Интегральные фотонные схемы также сталкиваются с проблемами, связанными с изготовлением и упаковкой. Точное выравнивание волноводов, резонаторов и детекторов имеет решающее значение для минимизации потерь и обеспечения высокой целостности сигнала. Современные методы изготовления, такие как электронно-лучевая литография и фотолитография в глубоком УФ-излучении, позволяют добиться субволновой точности при формировании рисунка фотонных устройств, но небольшие изменения в геометрии волновода могут привести к значительному ухудшению производительности. Для решения этой проблемы исследователи разработали передовые процессы изготовления, такие как химико-механическая полировка (ХМП), для сглаживания шероховатости боковой стенки и снижения потерь на рассеяние в кремниевых волноводах. Кроме того, методы многослойной интеграции, при которых различные фотонные компоненты изготавливаются на отдельных слоях, а затем складываются вместе, использовались для создания трехмерных фотонных схем с минимальными перекрестными помехами и высокой плотностью интеграции [88-90].

Другой важной областью исследований является разработка эффективных фотонных соединений, которые могут связывать несколько фотонных схем вместе для формирования крупномасштабных нейронных сетей.

Фотонные соединения должны обеспечивать маршрутизацию оптических сигналов с малыми потерями между различными слоями сети, сохраняя при этом целостность сигнала. Чтобы добиться этого, исследователи изучили использование волноводных перекрестков, вертикальных ответвителей и многоуровневых фотонных сетей. Волноводные перекрестки, где два волновода пересекаются под прямым углом, особенно сложны из-за потенциальных потерь на отражение и рассеяние. Для «смягчения» этих эффектов, усовершенствованные конструкции перекрестков используют конические волноводы и антибликовые покрытия для минимизации потерь и поддержания точности сигнала. Вертикальные ответвители, которые соединяют волноводы на разных слоях многоуровневой фотонной схемы, являются еще одним важным компонентом для построения трехмерных фотонных сетей. Эти ответвители используют затухающую связь между вертикально уложенными волноводами для передачи света между слоями с минимальными потерями и перекрестными помехами.

4.2. Роль вычислений с использованием фотонных резервуаров в обработке сигналов в реальном времени

Фотонные резервуарные вычисления (PRC) появились как эффективный подход к обработке сигналов в реальном времени и машинному обучению, предлагая простую, но эффективную архитектуру для обработки сложных временных и пространственных данных. В отличие от традиционных нейронных сетей, где веса каждого нейрона должны корректироваться в процессе обучения, системы резервуарных вычислений используют фиксированную сеть нелинейных узлов, при этом обучается только выходной слой. Это значительно упрощает процесс обучения и делает резервуарные вычисления в каком-то смысле идеальными для аппаратных реализаций с использованием фотонных устройств.

В типичном фотонном резервуаре (рис. 19) входной сигнал вводится в сеть связанных нелинейных узлов, таких как полупроводниковые лазеры, микрокольцевые резонаторы или волноводные решетки. Динамика резервуара,

которая регулируется взаимодействиями между узлами, преобразует вход в многомерное представление, которое может захватывать сложные временные паттерны. Это многомерное отображение затем считывается с помощью набора фотодетекторов, а полученные сигналы обрабатываются с помощью линейного классификатора для получения конечного вывода. Поскольку сам резервуар не обучен, основная проблема в PRC заключается в проектировании резервуара с правильными свойствами, такими как высокая нелинейность, богатая динамика и большое пространство состояний.

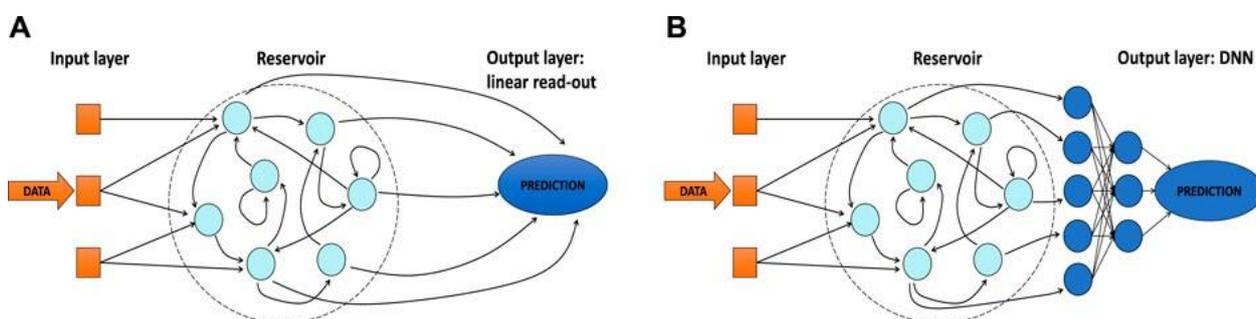


Рис. 19. Упрощенное представление общей резервуарной вычислительной системы, с слоем ввода, резервуаром и линейным считыванием в выходном слое (А). Упрощенное представление сети, которая использует резервуар в качестве предпроцессора. В выходном слое линейный слой считывания заменен DNN с целью выполнения прогноза (В). [91-92].

Одна из наиболее широко изученных реализаций PRC основана на полупроводниковых лазерах с задержкой (рис. 20). В этих системах один полупроводниковый лазер подвергается задержанной оптической обратной связи через волоконный контур или встроенный волновод, создавая ряд задержанных во времени реплик входного сигнала. Динамика лазера, которая регулируется уравнениями Ланга-Кобаяши [27-28], генерирует многомерное пространство состояний, которое чувствительно к небольшим изменениям на входе. Регулируя длину задержки, глубину обратной связи и ток инжекции, резервуар можно настроить на демонстрацию широкого спектра динамических режимов функционирования («поведений»), от периодических колебаний до хаотических режимов.

На рис. 20 SL – полупроводниковый лазер с накачкой постоянным током. Во входном слое маска $m(t)$ осуществляет кодировку входящего образца u_k ,

который, в свою очередь, инжектируется в резервуар посредством модулятора Маха-Цендера (MZM). Расстояние между нодами резервуара θ , время задержки τ , виртуальные ноды обозначены светло-голубыми кружками. Прогнозируемые в резервуаре целевые данные обозначены \hat{y}_k (A). По сути, это упрощенное представление сети, которая использует резервуар в качестве предпроцессора r , связанного с DNN, нодам соответствуют кружки синего цвета (B).

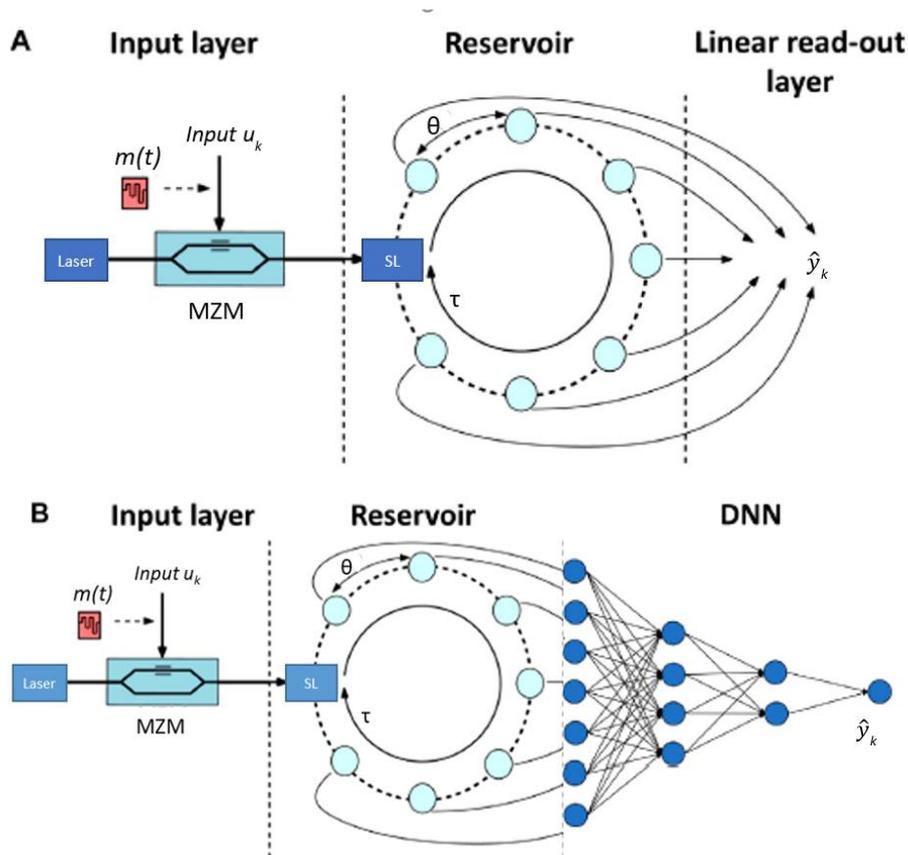


Рис. 20. Упрощенная архитектура резервуарной вычислительной системы, основанной на линии задержки, с слоем ввода, резервуаром и линейных выходным считывающим слоем.

Такая архитектура может быть использована для выполнения сложных вычислений, при этом выход резервуара считывается с помощью фотодетектора и обрабатывается линейным классификатором. Резервуары с задержкой успешно применялись для таких задач, как распознавание произнесенных цифр, прогнозирование временных рядов и реконструкция хаотических сигналов, демонстрируя свою универсальность и вычислительную мощность.

Другая перспективная реализация PRC основана на интегрированных фотонных схемах с использованием микрокольцевых резонаторов или волноводных решеток. В этих системах резервуар состоит из сети связанных резонаторов, где коэффициенты связи и резонансные частоты случайным образом распределены для создания разнообразного набора шаблонов отклика. Входной сигнал вводится в один или несколько узлов сети, и полученные интерференционные картины обнаруживаются на различных выходных узлах. Интерференционные картины кодируют многомерное преобразование входа, которое может использоваться для задач классификации или регрессии. Интегрированные фотонные резервуары обладают высокой масштабируемостью, поскольку количество узлов и соединений можно легко увеличить, добавив больше резонаторов или пересечений волноводов. Более того, использование интегральной фотоники позволяет реализовать реконфигурируемые резервуары, где коэффициенты связи и резонансные частоты можно динамически регулировать с помощью термооптических или электрооптических элементов настройки.

4.3. Практические приложения и перспективы применения фотонных вычислений

Развитие фотонных вычислительных систем способно оказать преобразующее влияние на широкий спектр приложений, от высокопроизводительных вычислений и телекоммуникаций до искусственного интеллекта и автономных систем. Сверхбыстрые скорости обработки и параллелизм фотонных схем делают их идеальными для задач, требующих анализа данных в реальном времени, таких как распознавание изображений и речи, обработка естественного языка и управление роботами. В частности, фотонные нейронные сети (PNN) и фотонные резервуарные вычислительные системы обеспечивают значительные преимущества для приложений машинного обучения, поскольку они могут обрабатывать большие наборы данных с минимальной задержкой и энергопотреблением.

Одно из самых перспективных применений фотонных вычислений находится в области телекоммуникаций, где спрос на высокоскоростную передачу и обработку данных продолжает расти. Фотонные схемы могут использоваться для реализации сверхбыстрых функций обработки сигналов, таких как оптическая коммутация, фильтрация и модуляция, что позволяет разрабатывать оптические сети с высокой пропускной способностью и скоростью передачи данных 1 Тбит/с и более. Кроме того, фотонные соединения изучаются как решение для снижения энергопотребления и задержки центров обработки данных, где узкое место в коммуникации между серверами является основным ограничением. Заменяя электрические соединения фотонными соединениями, центры обработки данных могут достичь значительно более высокой пропускной способности и более низкого энергопотребления, что позволяет создавать более эффективные и масштабируемые вычислительные архитектуры.

Другая ключевая область применения – автономные системы, где способность обрабатывать данные, получаемые в реальном времени от большого числа разнообразных сенсоров, имеет решающее значение для таких задач, как навигация, распознавание объектов и принятие решений. Фотонные нейроморфные системы с их высокоскоростными возможностями обработки и низким энергопотреблением хорошо подходят для реализации сложных нейронных сетей, необходимых для этих задач. Например, фотонные нейронные сети могут использоваться для реализации сверточных нейронных сетей (CNN) для распознавания изображений или рекуррентных нейронных сетей (RNN) для прогнозирования и управления временными рядами. Эти системы могут быть интегрированы в автономные транспортные средства, беспилотные летательные аппараты и робототехнику, обеспечивая принятие решений в реальном времени и адаптивное поведение в динамических средах.

Заключение

Представленная публикация позволяет однозначно утверждать, что внедрение вычислительных процессов, реализованных на основе фотонных технологий, способно оказать серьезное влияние на развитие и проектирование перспективных средств аппаратного (в будущем, возможно, и аппаратно-программного) обеспечения ИИ, предлагая уникальные преимущества с точки зрения скорости, распараллеливания вычислений и энергоэффективности. Несмотря на наличие определенных (порою – значительных) технических и технологических проблем, последние достижения в области фотонных материалов, устройств и архитектур приблизили нас к реализации крупномасштабных фотонных нейронных сетей и вычислительных систем. В следующем десятилетии, вероятно, будет наблюдаться непрерывный прогресс в этой области с разработкой гибридных оптоэлектронных платформ, усовершенствованной фотонной памяти и полностью интегрированных фотонных процессоров, способных выполнять сложные задачи ИИ с беспрецедентной скоростью. По мере развития этих технологий, они откроют новые приложения и возможности, сформируют подходы к новому типу высокопроизводительных вычислений, при котором свет, а не электроны, будет управлять наиболее важными и трудоемкими вычислительными процессами, обеспечивая решение критически важных для устойчивого развития человеческой цивилизации задач.

Литература

1. Brückerhoff-Plückelmann F. et al. Broadband photonic tensor core with integrated ultra-low crosstalk wavelength multiplexers // *Nanophotonics*. – 2022. – Vol. 11. No. 17. – P. 1–10.
2. Brückerhoff-Plückelmann F. et al. A large scale photonic matrix processor enabled by charge accumulation // *Nanophotonics*. – 2023. – Vol. 2. – P. 812. <https://doi.org/10.1515/nanoph-2022-0441>
3. Shi B., Calabretta N., Stabile R. Deep neural network through an InP SOA-based photonic integrated cross-connect // *IEEE J. Sel. Top. Quantum Electron.* – 2020. – Vol. 26, No. 1. – P. 1–11.
4. Wu C. et al. Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network // *Nat. Commun.* – 2021. – Vol. 12, No. 96. – P. 1–8.
5. Wetzstein G. et al. Inference in artificial intelligence with deep optics and photonics // *Nature*. – 2020. – Vol. 588. No. 7836. – P. 39–47.
6. Feldmann J. et al. Parallel convolutional processing using an integrated photonic tensor core // *Nature*. – 2021. – Vol. 589. No. 7840. – P. 52–58.
7. Xu X. et al. 11 TOPS photonic convolutional accelerator for optical neural networks // *Nature*. – 2021. – Vol. 589. No. 7840. – P. 44–51.
8. Zhou T. et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit // *Nat. Photonics*. – 2021. – Vol. 15. No. 5. – P. 367–373.
9. Goodman J. W., Dias A. R., Woody L. M. Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms // *Opt. Lett.* – 1978. – Vol. 2. No. 1. – P. 1–3.
10. Chang J. et al. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification // *Sci. Rep.* – 2018. – Vol. 8. No. 1. – P. 12324.
11. Zuo Y. et al. All-optical neural network with nonlinear activation functions // *Optica*. – 2019. – Vol. 6. No. 9. – P. 1132–1137.

12. Inagaki T. et al. Large-scale Ising spin network based on degenerate optical parametric oscillators // Nat Photon/ – 2016. 10(6): 415–9. <https://doi.org/10.1038/nphoton.2016.68>
13. Larger L. et al. Photonic information processing beyond turing: An optoelectronic implementation of reservoir computing // Opt Express. – 2012. – 20(3):3241–9. <https://doi.org/10.1364/OE.20.003241>
14. Takano K. et al. Compact reservoir computing with a photonic integrated circuit // Optics Express. – 2018. – 26(22): 29424–39. <https://doi.org/10.1364/OE.26.029424>
15. M. Tan. Microwave and RF Photonic Fractional Hilbert Transformer Based on a 50 GHz Kerr Micro-Comb // Journal of Lightwave Technology. – 2019. – Vol. 37. No. 24. – P. 6097-6104. <https://doi.org/10.1109/JLT.2019.2946606>
16. Tan M. et al. RF and microwave fractional differentiator based on photonics // IEEE Transactions on Circuits and Systems II: Express Briefs. Early Access. – 2020. – V. 67. – I. 11. – P. 2767 – 2771. <https://doi.org/10.1109/TCSII.2020.2965158>
17. Kippenberg T. J., Holzwarth R., Diddams S. A. Microresonator-based optical frequency combs // Science. – 2011. –Vol. 332. – P. 555-559 .
18. Savchenkov A. A. et al. Tunable optical frequency comb with a crystalline whispering gallery mode resonator // Physics Review Letters. –2008. –Vol. 101. – P. 093902.
19. Pasquazi A. et al. Micro-combs: a novel generation of optical sources // Physics Reports. – 2018. – Vol. 729. –P. 1-81.
20. Ferrera M. et al. Low-power continuous-wave nonlinear optics in doped silica glass integrated waveguide structures // Nat. Photonics. –2008. –Vol. 2 (12). – P. 737-740.
21. Peccianti M. et al. Demonstration of a stable ultrafast laser based on a nonlinear microcavity // Nat. Commun. – 2012. –Vol. 3 765. – P. 1-6.
22. Pasquazi A. et al. Stable, dual mode, high repetition rate mode-locked laser based on a microring resonator // Opt. Exp. – 2012. – Vol. 20. No. 24. – P. 27355-27362.

23. Pasquazi A. et al. Self-locked optical parametric oscillation in a CMOS compatible microring resonator: a route to robust optical frequency comb generation on a chip // *Opt. Exp.* –2013. – Vol. 21. No. 11. – P. 13333-13341.
24. Shastri B. J. et al. Spike processing with a graphene excitable laser // *Sci. Rep.* – 2016. – Vol. 6. – P. 19126.
25. Feldmann J. et al. All-optical spiking neurosynaptic networks with self-learning capabilities // *Nature.* – 2019. Vol. 569. – P. 208-214.
26. Duport F. et al. All-optical reservoir computing // *Opt. Express.* – 2012. 20(20):22783–95. <https://doi.org/10.1109/JSTQE.2018.2836985>
27. Lang R., Kobayashi K. External optical feedback effects on semiconductor injection laser properties // *IEEE Journal of Quantum Electronis.* – 1980. – Vol. 16. – No. 3. –P. 347-355. <https://doi:10.1109/JQE.1980.1070479>
28. Alsing P. M., Kovanis V., Gavrielides A., Erneux T. Lang and Kobayashi phase equation // *Physical Review A.* – 1996. –Vol. 53 (6). 4429-4434. <https://doi.org/10.1103/PhysRevA.53.4429>
29. Huang L. et al. Mid-infrared modulators integrating silicon and black phosphorus photonics // *Ang. Mater. Today Adv.* 2021. 12. 100170 <https://doi.org/10.1016/j.mtadv.2021.100170>
30. Soriano V. et al. Graphene–silicon phase modulators with gigahertz bandwidth // *Nat. Photonics.* – 2018. –Vol. 12. 40.
31. Qiao Q. et al. Multifunctional mid-infrared photonic switch using a MEMS-based tunable waveguide coupler // *Opt. Lett.* – 2020. – Vol. 45. 5620. <https://doi.org/10.1364/OL.400132>
32. Sagheer A. Kotb M. Time series forecasting of petroleum production using deep LSTM recurrent networks // *Neurocomputing.* – 2019. – V. 323. – P. 203–13. <https://doi.org/10.1016/j.neucom.2018.09.082>
33. Chimmula VKR. Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks // *Chaos. Solitons & Fractals.* – 2020. – Vol. 135:109864. <https://doi.org/10.1016/j.chaos.2020.109864>

34. Hochreiter S. Schmidhuber J. Long short-term memory // *Neural Comput*/ – 1997. – Vol. 9:1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>
35. LeCun Y. Bengio Y. Hinton G. Deep learning // *Nature*. – 2015. – Vol. 521:436–44. <https://doi.org/10.1038/nature14539>
36. Voulodimos A. Doulamis N. Doulamis A. Protopapadakis E. Deep learning for computer vision: A brief review // *Comput intelligence Neurosci*. – 2018
37. Demirkiran C. Eris F. Wang G. Elmhurst J. Moore N. Harris NC. et al. An electro-phonic system for accelerating deep neural networks // – P. 01126. arXiv preprint (2021):arXiv:2109. <https://doi.org/10.48550/arXiv.2109.01126>
38. Xu S., Ren Z., Dong B., J. Zhou, W. Liu, C. Lee. Mid-Infrared Silicon-on-Lithium-Niobate Electro-Optic Modulators Toward Integrated Spectroscopic Sensing Systems // *Adv. Opt. Mater.* – 2023. 11. – P. 2202228.
39. Li Z., Zhang H., B. Nguyen T. T., et al. // *Photonics Res.* – 2021, 9, B38.
40. Zhou J., Zhang Z., Dong B., et al. // *Opt. Express* – 2010. 18. 1756
41. Ma Y., Chang Y., Dong B., Wei J., W. Liu, C. Lee // *ACS Nano*. – 2021. – Vol. 15. 10084.
42. Rios C., Du Q., Zhang Y., et al. // *Sci. Rep.* – 2016. 6. 22616.
43. Shi B., N. Calabretta, Stabile R. Deep neural network through an InP SOA-based photonic integrated cross-connect // *IEEE J. Sel. Top. Quantum Electron.* – 2020. – Vol. 26. No. 1. – P. 1–11.
44. Xu Q., Soref R. Reconfigurable optical directed-logic circuits using microresonator-based optical switches, *Opt. Express*, – 2011. – Vol. 19. No. 6. – P. 5244–5259.
45. Yang L., Ji R., Ding J., Xu Q. On-chip CMOS-compatible optical signal processor // *Opt. Express*. – 2012. – Vol. 20, No. 12, – P. 13560–13565.
46. Tait A. N., Nahmias M. A., Shastri B. J., Prucnal P. R. Broadcast and weight: an integrated network for scalable photonic spike processing // *J. Lightwave Technol.* – 2014. – Vol. 32. No. 21. – P. 4029–4041.
47. Tait A. N., de Lima T. F., Zhou E., et al. Neuromorphic photonic networks using silicon photonic weight banks // *Sci. Rep.* – 2017. – Vol. 7. No. 1. – P. 7430.

48. Tait A. N., de Lima T. F., Nahmias M. A., Shastri B. J., Prucnal P. R. Multi-channel control for microring weight banks // *Opt. Express.* – 2000. – Vol. 24., No. 8. – P. 8895–8906.
49. Tait A. N., de Lima T. F., Nahmias M. A., B. J. Shastri, P. R. Prucnal. Continuous calibration of microring weights for analog optical networks // *IEEE Photon. Technol. Lett.* – 2016. – Vol. 28. – No. 8. – P. 887–890.
50. Tait A. N., Xu A. X., de Lima T. F., et al. Microring weight banks // *IEEE J. Sel. Top. Quantum Electron.* – 2016. – Vol. 22. – No. 6. – P. 312–325.
51. Huang C., Bilodeau S., de Lima T. F., et al. Demonstration of scalable microring weight bank control for large-scale photonic integrated circuits // *APL Photonics* – 2020. – Vol. 5. – No. 4. – P. 040803.
52. Xu S., Wang J., Zou W. Optical convolutional neural network with WDM-based optical patching and microring weighting banks // *IEEE Photon. Technol. Lett.* – 2021. – Vol. 33. No. 2. – P. 89–92.
53. Bangari V., Marquez B. A., Miller H., et al. Digital electronics and analog photonics for convolutional neural networks (DEAP-CNNs) // *IEEE J. Sel. Top. Quantum Electron.* – 2020. – Vol. 26. No. 1. – P. 1–13.
54. Huang C., Fujisawa S., de Lima T. F., et al. A silicon photonic-electronic neural network for fibre nonlinearity compensation // *Nat. Electron.* – 2021. – Vol. 4. No. 11. – P. 837–844.
55. Shi B., Calabretta N., Stabile R. Deep neural network through an InP SOA-based photonic integrated cross-connect // *IEEE J. Sel. Top. Quantum Electron.*, – 2020. – Vol. 26. – No. 1. – P. 1–11.
56. Mourgias-Alexandris G., Tsakyridis A., Passalis N., et al. An all-optical neuron with sigmoid activation function // *Opt. Express.* – 2019. – Vol. 27. No. 7. – P. 9620.
57. Shi B., Calabretta N., Stabile R. InP photonic integrated multi-layer neural networks: architecture and performance analysis // *APL Photonics.* – 2022. – Vol. 7. No. 1. –P. 010801.

58. Chakraborty I., Saha G. Sengupta, A., K. Roy. Toward fast neural computing using all-photonic phase change spiking neurons // *Sci. Rep.* – 2018. – Vol. 8. No. 1. – P. 12980.
59. Li X., Youngblood N., Zhou W., et al. On-chip phase change optical matrix multiplication core // in *IEEE Int. Electron Devices Meet.* – 2020. – P. 7.5.1–7.5.4.
60. Nahmias M. A., Peng H., de Lima T. F., et al. A teraMAC neuromorphic photonic processor // in *IEEE Photonics Conf. (IPC).* – 2018. – P. 1–2.
61. Shainline J. M., Buckley S. M., McCaughan A. N., et al. Circuit designs for superconducting optoelectronic loop neurons // *J. Appl. Phys.* – 2018. – Vol. 124. No. 15. – P. 152130.
62. Shastri J., Tait A. N., Prucnal P. R. A leaky integrate-and-fire laser neuron for ultrafast cognitive computing // *IEEE J. Sel. Top. Quantum Electron.* – 2013. – Vol. 19, No. 5, – P. 1–12.
63. Tait A. N., Chang J., Shastri B. J. Nahmias, M. A., Prucnal P. R. Demonstration of WDM weighted addition for principal component analysis // *Opt. Express* – 2015. – Vol. 23, No. 10, – P. 12758–12765.
64. Xu X., Moss D., Tan M., J. Wu, Morandotti R. Photonic perceptron based on a kerr microcomb for high-speed, scalable, optical neural networks // *Laser Photon. Rev.* – 2020. – Vol. 14. – No. 10. – P. 44–51.
65. Mourgias-Alexandris G., Dabos G., Passalis N., Totović A., A. Tefas, and N. Pleros. All-optical WDM recurrent neural networks with gating // *IEEE J. Sel. Top. Quantum Electron.* – 2020. – Vol. 26, No. 5, – P. 1–7.
66. Totovic A., Giamougiannis G., Tsakyridis A., et al. Programmable photonic neural networks combining WDM with coherent linear optics // *Sci. Rep.* – 2022. – Vol. 12, No. 1, p. 5605.
67. Xiang J., Tao Z., Li X., et al. Metamaterial-enabled arbitrary on-chip spatial mode manipulation // *Light Sci. Appl.* – 2022. – Vol. 11, No. 168, – P. 1–11.
68. Zhou Y., Braverman B., Fyffe A., et al. High-fidelity spatial mode transmission through a 1-km-long multimode fiber via vectorial time reversal // *Nat. Commun.*, – 2021. – Vol. 12, No. 1866, – P. 1–7.

69. Jiang W., Miao J., Li T. Compact silicon 10-mode multi/demultiplexer for hybrid mode- and polarization-division multiplexing system // *Sci. Rep.* – 2019. – Vol. 9, No. 13233, – P. 1–15.
70. Feldmann J., Youngblood N., C. D. Wright, H. Bhaskaran, W. H. P. Pernice // *Nature.* – 2019. – Vol. 569. 208.
71. Perez-Lopez D., Lopez A., P. DasMahapatra, J. Capmany // *Nat. Commun.* – 2020. –Vol. 11. 6359.
72. Prabhu M., Roques-Carmes C., Shen Y., N. Harris, L. Jing, J. Carolan, R. Hamerly, T. Baehr-Jones, M. Hochberg, V. ˇCeperi´c, J. D. Joannopoulos, D. R. Englund, M. Soljacic // *Optica.* – 2020. 7. 551.
73. Wu C., Yang X., H. Yu, R. Peng, Takeuchi I., Y. Chen, M. Li // *Sci. Adv.* – 2022. 8. 2956.
74. Xu S., Wang J., S. Yi, W. Zou // *Nat. Commun.* – 2022. 13. 7970.
75. Bai B., Yang Q., Shu H., L. Chang, F. Yang, B. Shen, Z. Tao, J. Wang, S. Xu, W. Xie, W. Zou, W. Hu, J. E. Bowers, X. Wang // *Nat. Commun.* – 2023. 14. 66.
76. Cheng J., Zhang W., W. Gu, H. Zhou, J. Dong, X. Zhang // *ACS Photonics.* – 2022. 10. 2173.
77. Xu X., Tan M., Corcoran B. et al. 11 TOPS photonic convolutional accelerator for optical neural networks // *Nature* – 2021. 589. 44.
78. Shen Y., Harris N. C., Skirlo S. et al. Deep Learning with Coherent Nanophotonic Circuits // *Nat. Photonics* – 2017, 11, 441.
79. Zhang H., Gu M., Jiang X. D. et al. An optical neural chip for implementing complex-valued neural network // *Nat. Commun.* – 2021. – Vol. 12. – P. 457.
80. Ashtiani F., Geers A. J., Aflatouni F. An on-chip photonic deep neural network for image classification // *Nature/* – 2022. – Vol. 606. –P 501.
<https://doi.org/10.1038/s41586-022-04714-0>
81. Tanomura R. Tang R, Umezaki T, Soma G, Tanemura T, Nakano Y. Scalable and robust photonic integrated unitary converter based on multiplane light conversion // *Phys Rev Appl.* – 2022. –Vol. 17(2):024071
<https://doi.org/10.1103/PhysRevApplied.17.024071>

82. Qu Y, Zhu H, Shen Y, Zhang J, Tao C, Ghosh P, et al. Inverse design of an integrated-nanophotonics optical neural network // *Sci Bull.* – 2020. –Vol. 65(14):1177–83. <https://doi.org/10.1016/j.scib.2020.03.042>
83. Zarei S, Marzban MR, Khavasi A. Integrated photonic neural network based on silicon metalines // *Opt Express.* – 2020. –Vol. 28(24):36668–84. <https://doi.org/10.1364/oe.404386>.
84. Khoram E., Chen A., Liu D, Ying L., Wang Q., Yuan M. et al. Nanophotonic media for artificial neural inference // *Photon Res.* – 2019.– Vol. 7(8):823–7. <https://doi.org/10.48550/arXiv.1810.07815>
85. Huang C, Sorger VJ, Miscuglio M, Al-Qadasi M, Mukherjee A, Lampe L, et al. Prospects and applications of photonic neural networks // *Adv Phys X.* – 2022. – Vol. 7: 1981155. <https://doi.org/10.1080/23746149.2021.1981155>
86. Wu J, Lin X, Guo Y, Liu J, Fang L, Jiao S, et al. Analog optical computing for artificial intelligence // *Engineering.* – 2021. –Vol. 10:133–45. <https://doi.org/10.1016/j.eng.2021.06.021>
87. Xu X, Tan M, Corcoran B, Wu J, Boes A, Nguyen TG, et al. 11 TOPS photonic convolutional accelerator for optical neural networks // *Nature.* – 2021.– Vol. 589:44–51. <https://doi.org/10.1038/s41586-020-03063-0>
88. Li Y, Xue Y, Tian L. Deep speckle correlation: A deep learning approach toward scalable imaging through scattering media // *Optica/* – 2018. –Vol. 5:1181–90. <https://arxiv.org/abs/1806.04139v2>
89. Brossollet C, Cappelli A, Carron I, Chaintoutis C, Chatelain A, Daudet L, et al. LightOn optical processing unit: Scaling-up AI and HPC with a non von Neumann co-processor (2021) // arXiv preprint arXiv:2107.11814.
90. Launay J, Poli I, Müller K, Carron I, Daudet L, Krzakala F, et al. Light-in-the loop: Using a photonics co-processor for scalable training of neural networks // arXiv preprint arXiv:2006.01475 (2020). <https://doi.org/10.48550/arXiv.2006.01475>
91. Tanaka G., Yamane T. et al. Recent advances in physical reservoir computing: A review // *Neural Networks.* – 2019. 115: 100–23. <https://doi.org/10.48550/arXiv.1808.04962>

92. Li Y., Yu H. et al. Graphene-Based Floating-Gate Nonvolatile Optical Switch // IEEE Photonics Technol. Lett. – 2016. – Vol. 28. – P. 284.
93. Rios C. et al. Integrated all-photonic non-volatile multi-level memory // Nat. Photonics. – 2015. – Vol. 9, – P. 725.

Для цитирования:

Мачнев А.А., Мельников И.В., Надеждин Е.Р., Соколов В.А., Ступин Д.Д. Фотонные технологии для реализации нейроморфных вычислений: результаты, направления исследований и аппаратное обеспечение ИИ. // Журнал радиоэлектроники. – 2026. – №. 2. <https://doi.org/10.30898/1684-1719.2026.2.12>