

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА
ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи

Кузьмин Сергей Сергеевич

**Методы машинного обучения для характеристики
квантовых состояний и программирования
линейно-оптических интерферометров**

Специальность 1.3.19. Лазерная физика

ДИССЕРТАЦИЯ
на соискание учёной степени
кандидата физико-математических наук

Научный руководитель:
кандидат физико-математических наук
Страупе Станислав Сергеевич

Москва — 2026

Оглавление

	Стр.
Введение	5
Глава 1. Линейно-оптические интерферометры	17
1.1 Теоретическое описание	17
1.1.1 Преобразование когерентных состояний	17
1.1.2 Преобразование фоковских состояний	18
1.2 Дизайн линейно-оптических схем	19
1.2.1 Стандартные дизайны	20
1.2.2 Блочный дизайн	23
1.3 Изготовление	27
1.3.1 Литография	27
1.3.2 Фемтосекундная лазерная печать	27
1.4 Методы томографии преобразования	28
1.4.1 Когерентные состояния	28
1.4.2 Фоковские состояния	30
Глава 2. Обучение модели машинного обучения	
линейно-оптического интерферометра	34
2.1 Формулировка проблемы	35
2.2 Алгоритм обучения	37
2.2.1 Вычисление метрики	38
2.2.2 Вычисление градиента метрики	38
2.3 Численные симуляции	40
2.3.1 Модель — чёрный ящик	42
2.3.2 Модель с априорным знанием	45
2.4 Эксперимент	48
2.4.1 Дизайн эксперимента	48
2.4.2 Калибровка целевого интерферометра	50
2.4.3 Целевая модель интерферометра	53
2.4.4 Детали обучения и результаты эксперимента	53
2.5 Выводы по главе	57

Глава 3. Переход в новое признаковое пространство фаз для	
управления линейным оптическим интерферометром	59
3.1 Математические основы	59
3.2 Алгоритмы обучения	61
3.2.1 Обучение на всех слоях	61
3.2.2 Послойное обучение	64
3.3 Численные симуляции	66
3.3.1 Симуляция обучения на всех слоях	66
3.3.2 Симуляция послойного обучения	68
3.4 Практические аспекты экспериментальной реализации	74
3.5 Выводы по главе	76
Глава 4. Симуляция квантовых систем с использованием	
формализма тензорных сетей	78
4.1 Формализм тензорных сетей	79
4.1.1 Сингулярное разложение	79
4.1.2 Тензорные разложения	80
4.1.3 Алгоритм TT-SVD	82
4.2 Обзор литературы по тензорным сетям	85
4.3 Представление квантовых состояний в формате Tensor Train .	88
4.3.1 Разложение Шмидта	88
4.3.2 Представление чистых квантовых состояний в TT-формате	90
4.3.3 Представление оператора плотности квантового состояния в TT-формате	91
4.3.4 Примеры представления состояний	92
4.4 Преобразование квантовых состояний в TT-формате	93
4.4.1 Применение однокубитных гейтов	93
4.4.2 Применение двухкубитных гейтов	95
4.4.3 Ортонормализация вагонов тензорного поезда	98
4.4.4 Скалярное произведение квантовых состояний в TT-формате	102
4.4.5 Вычисление матрицы плотности одного кубита	105
4.5 Численные эксперименты	106

4.5.1	Случайные квантовые схемы	107
4.5.2	Распределение Портера-Томаса	109
4.5.3	Эффективное двухкубитное фиделити	111
4.5.4	Многокубитное фиделити	112
4.5.5	Распределение точности и чистоты состояний	114

Глава 5. Обучение тензорной сетевой модели квантового состояния с использованием однокубитных измерений 117

5.1	Обзор литературы	117
5.2	Тензорно-сетевая модель квантового состояния	120
5.2.1	Модель измерений	120
5.2.2	Обучаемая модель	121
5.2.3	Функция потерь	123
5.3	Численные симуляции	124
5.3.1	Симуляции на случайных смешанных состояниях . . .	124
5.3.2	Симуляции на реалистичном физическом устройстве .	127
5.4	Выводы по главе	130
Заключение		132
Список сокращений и условных обозначений		135
Список литературы		137
Список рисунков		151

Введение

Квантовые технологии быстро переходят от лабораторных демонстраций к инженерным системам, в которых ключевую роль играют воспроизводимость, программируемость и масштабируемость аппаратных платформ. Методы линейной оптики и интегральной фотоники используются во многих современных экспериментах и прототипах квантовых устройств. Многомодовые интерферометры позволяют реализовывать унитарные преобразования оптических мод и служат основой для обработки квантовой информации, генерации и анализа многомерных запутанных состояний, а также выполнения фотонных вычислений. Современные фотонные схемы, как правило, содержат большое число управляемых параметров, прежде всего фазовых сдвигов. Это обеспечивает программируемость таких устройств, однако одновременно повышает сложность их калибровки, настройки и последующей эксплуатации.

В общем случае программирование линейно-оптического интерферометра заключается в выборе управляющих параметров, при которых устройство реализует заданное преобразование, чаще всего представленное унитарной матрицей. Для некоторых «идеальных» универсальных архитектур разработаны аналитические алгоритмы программирования, основанные на декомпозиции унитарных матриц. Вместе с тем в реальных фотонных чипах и оптических системах возникают технологические отклонения, паразитные связи, потери и дрейфы параметров, вследствие чего реализуемое преобразование может значительно расходиться с проектным. Для более устойчивых к дефектам архитектур аналитическое преобразование «целевая матрица $\rightarrow$ набор фаз» как правило отсутствует. В этом случае программирование целевого преобразования превращается в итерационную оптимизацию на физическом устройстве, предполагающую многократную реконструкцию (томографию) преобразования, которая быстро становится ресурсоёмкой и нестабильной с увеличением размерности интерферометра.

Преодолеть эти ограничения можно с помощью построения «цифрового двойника» устройства: численной модели, согласованной с экспериментальными данными и используемой для быстрого расчёта управляющих параметров без обращения к физическому устройству (без томографии) на

каждом шаге оптимизации. Инструментом для такого подхода служат методы машинного обучения, в которых параметры модели подбираются по обучающей выборке вида «управляющие параметры $\rightarrow$ наблюдаемый отклик». Для фотонных интерферометров это означает обучение модели на основе набора матриц преобразования, реконструированных при различных настройках фаз. Затем подбор фаз, реализующих целевое преобразование, осуществляется с использованием обученной модели.

Помимо задач программирования линейно-оптических устройств существует фундаментальная проблема эффективной характеристики (томографии) квантовых состояний. Полная квантовая томография требует экспоненциально растущего с размером системы числа измерений, что ограничивает применимость стандартных протоколов для квантовых устройств промежуточного масштаба с шумом (noisy intermediate-scale quantum, NISQ). Это мотивирует развивать методы неполной характеристики, которые используют априорные предположения о структуре состояния и нацелены на предсказание экспериментально наблюдаемых величин, а не на восстановление полного описания состояния. В этой области важную роль играет формализм тензорных сетей, позволяющий компактно параметризовать широкий класс физических состояний, а также используются подходы, близкие к парадигме вероятно приблизительно корректного обучения (probably approximately correct, PAC) и методам теневой томографии, в которых для предсказания результатов будущих измерений достаточно сравнительно небольшого набора измерений.

Диссертация посвящена применению методов машинного обучения к двум связанным задачам квантовых технологий: построению и обучению моделей для линейно-оптических интерферометров и алгоритмов их программирования, а также обучению компактных тензорно-сетевых моделей квантовых состояний по ограниченному набору результатов локальных измерений для предсказания вероятностей исходов будущих измерений. Общей методологической основой работы является перенос ресурсоёмких процедур настройки и характеристики в пространство обучаемых моделей, использующих априорные структурные предположения о физической системе и обеспечивающих более эффективное масштабирование с ростом числа параметров системы.

В главе 1 представлен обзор работ, посвящённых линейно-оптическим интерферометрам и их универсальным архитектурам, а также реконструкции реализуемых ими преобразований. В главе 2 рассматривается обучение архитектурно-независимой модели линейно-оптического интерферометра по экспериментальным данным и её последующее использование для программирования экспериментального устройства. Приведены результаты экспериментальной проверки предложенного подхода, а также исследовано влияние факторов, существенных для его практического применения, включая паразитные взаимные влияния (кросс-токи) и ограничения схемы управления. В главе 3 предложены методы повышения эффективности обучения модели интерферометра и его программирования. Они основаны на использовании специальных признаков, зависящих от фаз в плечах интерферометра, и линеаризации зависимости реализуемого преобразования от этих признаков. В главе 4 представлен обзор тензорных сетей и рассмотрены их применения для моделирования квантовых состояний. В главе 5 предложен и исследован протокол обучения тензорно-сетевой модели неизвестного квантового состояния из заданного класса состояний по результатам измерений, задаваемых операторами, представимыми в виде тензорных произведений однокубитных операторов. Обсуждаются условия применимости метода и его масштабируемость. В заключении сформулированы основные результаты работы и обозначены направления дальнейших исследований.

Актуальность работы связана с двумя задачами. Первая состоит в разработке и исследовании методов машинного обучения для характеристики квантовых систем, в частности для построения компактных тензорно-сетевых моделей квантовых состояний по ограниченному набору измерений. Вторая задача состоит в калибровке программируемых линейно-оптических интерферометров с учётом технологических отклонений при изготовлении и паразитных взаимных влияний (кросс-токов). Решение этой задачи важно для создания и отладки квантовых фотонных вычислительных устройств.

Степень разработанности темы можно оценить как высокую. Для линейно-оптических интерферометров разработаны универсальные архитектуры, с помощью которых можно реализовывать произвольные целевые унитарные преобразования, а также практические методы их

программирования и томографической реконструкции реализуемых ими преобразований.

Классические архитектуры универсальных интерферометров и соответствующие схемы реализации унитарных преобразований были предложены Реком (Reck) [1] в 1994 году и Клементсом (Clements) [2] в 2016 году.

Для экспериментальной реконструкции преобразований линейно-оптических сетей разработано множество методов томографии, включая подход «super-stable», предложенный Лэйнгом (Laing) и О’Брайеном (O’Brien) [3] в 2012 году, метод прямой характеристики, представленный Рахими-Кешари (Rahimi-Keshari) [4] в 2013 году, а также оптимизационные методы реконструкции, например подход PhaseLift, рассмотренный в работе Суэсса (Suess) [5], опубликованной в 2020 году.

Направление программируемых фотонных процессоров и универсальных интегральных фотонных схем также активно развивается, что отражено, например, в работе Харриса (Harris) [6], опубликованной в 2018 году. В этой области рассматриваются вопросы калибровки с учётом устойчивости к параметрам изготовления и воспроизводимости результатов реализации преобразований.

Традиционные методы настройки и томографии требуют значительного числа измерений и итераций оптимизации, что стимулирует развитие альтернативных подходов. В рамках таких подходов модель устройства обучается на экспериментальных данных и затем используется для подбора управляющих параметров без дорогостоящей итеративной оптимизации непосредственно на физическом чипе.

В задаче характеристики квантовых состояний степень разработанности темы также можно оценить как высокую, однако традиционная полная томография ограничена экспоненциальным ростом числа измерений с увеличением размерности системы. Существенный теоретический вклад в понимание возможности «обучаемой» характеристики внесла работа Ааронсона (Aaronson) [7] 2007 года, посвящённая РАС-обучаемости квантовых состояний. Дальнейшее развитие получили методы теневой томографии и классических теней, позволяющие предсказывать множество свойств квантовой системы по относительно малому числу измерений; соответствующие результаты были получены Хуангом (Huang), Куэнгом (Kueng) и Прескиллом (Preskill) [8] в 2020 году. Параллельно сформировался мощный класс

структурных моделей на основе тензорных сетей, широко применяемых для компактного описания и реконструкции квантовых состояний и процессов при наличии ограниченной эффективной сложности. Таким образом, в литературе хорошо разработаны как методы построения программируемых фотонных схем и их характеристики, так и подходы к обучению компактных моделей квантовых состояний; при этом остаётся актуальной разработка конструктивных протоколов, одновременно вычислительно эффективных и совместимых с экспериментальными ограничениями, такими как шум и неполные измерения.

Целью диссертационной работы является разработка, анализ и экспериментальная проверка методов машинного обучения для (i) идентификации и программирования линейно-оптических интерферометров по экспериментальным данным и (ii) эффективной характеристики квантовых состояний на основе компактных тензорно-сетевых моделей, обучаемых по ограниченному набору локальных измерений.

В рамках поставленной цели были сформулированы следующие **задачи**:

1. Разработать архитектурно-независимый алгоритм обучения цифровой модели реконфигурируемого линейно-оптического интерферометра по обучающей выборке, содержащей пары «фазовые настройки – реконструированная матрица преобразования» (восстановленная по экспериментальным измерениям).
2. Выполнить экспериментальную проверку подхода к программированию фотонного процессора на основе обученной модели для случая малого числа мод ($N = 4$), исследовав влияние реалистичных факторов, включая кросс-токи, ограничение числа фазовых слоёв и особенности используемых измерительных процедур.
3. Разработать улучшенный алгоритм обучения модели интерферометра, использующий специально построенные признаки (линеаризацию зависимости преобразования от фазовых параметров), и исследовать его свойства, включая требования к объёму данных, устойчивость к шуму реконструкции и области применимости (в частности, для неглубоких схем).
4. Разработать протокол обучения тензорно-сетевой модели неизвестного многокубитного квантового состояния, использующий только

факторизованные однокубитные измерения, с целью предсказания вероятностей исходов будущих измерений.

5. Исследовать численную эффективность и масштабируемость тензорно-сетевого протокола, включая зависимость качества предсказаний от размера обучающей выборки и ранговых ограничений модели, и определить область применимости метода для состояний, возникающих в типичных для NISQ-устройств схемах.

Объектом исследования являются методы машинного обучения и вычислительные протоколы, применяемые для (i) идентификации, калибровки и программирования линейно-оптических интерферометров и фотонных процессоров по экспериментальным данным и (ii) эффективной характеристики квантовых состояний на основе обучаемых компактных моделей. **Предметом исследования** являются алгоритмы построения цифровых моделей интерферометров (включая архитектурно-независимое и линеаризованное обучение) и тензорно-сетевых моделей квантовых состояний, а также исследование их точности, устойчивости к экспериментальным неидеальностям (шум, ошибки реконструкции, кросс-токи, ограниченные измерения) и масштабируемости по числу мод или кубитов и объёму обучающей выборки.

Научная новизна заключается в следующих положениях:

1. Разработан архитектурно-независимый алгоритм обучения цифровой модели реконфигурируемого линейно-оптического интерферометра по обучающей выборке вида «фазовые настройки – реконструированная матрица преобразования».
2. Для указанного алгоритма получены аналитические выражения градиентов целевой функции (на основе нормы Фробениуса) по параметрам модели интерферометра, что позволяет эффективно применять градиентные методы оптимизации при обучении.
3. Предложена практическая схема проекции промежуточных оценок параметров модели на унитарное множество (через полярное разложение) в ходе обучения, обеспечивающая корректность модели при численной оптимизации.
4. Численно исследованы условия успешной идентификации модели интерферометра и выявлены ограничения «black-box» обучения

при росте размерности, связанные с трудностями невыпуклой оптимизации.

5. Показано, что использование априорной информации о внутренней структуре устройства (инициализация, близкая к проектным базисным преобразованиям) существенно расширяет область применимости метода по числу мод и повышает устойчивость обучения.
6. Исследована устойчивость обучения модели интерферометра к ошибкам реконструкции матриц преобразования (шуму томографии) и определены режимы, в которых точность данных становится лимитирующим фактором.
7. Выполнена экспериментальная реализация и проверка подхода к программированию фотонного процессора на основе обученной модели для случая $N = 4$.
8. Экспериментально исследовано влияние кросс-токов между фазовыми элементами на качество программирования и корректность обучаемой модели интерферометра.
9. Продемонстрирована работоспособность подхода при ограниченной управляемости схемы (один фазовый слой) и при использовании различных экспериментальных измерительных процедур, отличающихся от «идеализированной» полной реконструкции.
10. Предложен линеаризованный (признаковый) алгоритм обучения модели линейно-оптического интерферометра, в котором зависимость преобразования от фазовых параметров переводится в форму, допускающую обучение методами линейной алгебры.
11. Для линеаризованного подхода исследованы требования к объёму обучающих данных и влияние шума, а также выделены области применимости по глубине интерферометрической структуры.
12. Разработан вычислительно реализуемый протокол обучения тензорно-сетевой модели неизвестного многокубитного квантового состояния по данным факторизованных однокубитных измерений.
13. В рамках предложенного тензорно- сетевого протокола задача сводится к обучению модели с учителем по результатам доступных измерений и затем оцениванию вероятности исходов последующих измерений без восстановления полной матрицы плотности.

14. Проведено численное исследование точности и масштабируемости протокола, зависимости качества обучения от объёма обучающей выборки и ограничений на тензорные ранги модели. Для рассмотренных сценариев показано, что необходимый объём данных может расти в ряде случаев линейно от числа кубитов.

Теоретическая и практическая значимость определяется двумя обстоятельствами. Во-первых, работа связана с разработкой и экспериментальной проверкой новых методов программирования линейно-оптических интерферометров и характеристики квантовых состояний. Во-вторых, полученные результаты могут быть использованы при создании и отладке квантовых вычислительных устройств, а также практической настройке линейно-оптических интерферометров.

Методология диссертационного исследования включает в себя аналитическое моделирование, численные симуляции и экспериментальную проверку в линейно-оптическом эксперименте.

В задаче программирования линейно-оптических интерферометров рассматривались схемы с блочной архитектурой, в которой чередовались слои фазовых элементов и блоки смешивания мод. Параметры модели такого интерферометра оценивались в парадигме задачи обучения с учителем по выборке вида «фазовые параметры — реконструированная матрица преобразования».

Для подбора параметров модели применялись градиентные методы оптимизации целевой функции, основанной на норме Фробениуса. Унитарность промежуточных оценок обеспечивалась их проекцией на множество унитарных матриц с помощью полярного разложения.

Экспериментальная часть включала формирование обучающей выборки на основе результатов томографии преобразования интерферометра при различных установленных фазах и оценку качества его программирования. Для оценки использовались выбранные метрики близости между реализуемым и целевым преобразованиями.

В линеаризованном варианте модели интерферометра использовались специально построенные признаки, позволяющие свести этап обучения модели к решению задачи наименьших квадратов.

В задачах характеристики квантовых состояний методология опиралась на тензорно-сетевое моделирование и постановку задачи обучения с

учителем для предсказания вероятностей исходов измерений по данным факторизованных однокубитных измерений; оптимизация параметров тензорной модели выполнялась стохастическими градиентными методами по среднеквадратичной функции потерь, а качество обучения оценивалось на независимых тестовых выборках и по метрикам согласия предсказанных и истинных вероятностей.

Положения, выносимые на защиту:

1. Обучение цифровой модели реконфигурируемого линейно-оптического интерферометра по выборке «фазовые настройки — реконструированная матрица преобразования» осуществимо без априорного знания базисных матриц для размерностей до $N = 6$ включительно.
2. Программирование фотонного процессора на основе обученной модели экспериментально реализуемо для $N = 4$ в условиях, характерных для реальной аппаратуры, включая наличие перекрёстного влияния между фазовыми элементами, наличие только одного фазового слоя и использование ограниченного набора экспериментальных данных для обучения.
3. В линейаризованном (признаковом) подходе к обучению модели интерферометра режим обучения «по всем слоям» требует объёма обучающей выборки, масштабирующегося как $O(N^L)$, где N — число оптических мод, L — глубина интерферометрической структуры.
4. В линейаризованном (признаковом) подходе режим послойного обучения приводит к линейной по числу мод масштабируемости объёма обучающей выборки для одного слоя, что обеспечивает применимость метода для интерферометров с большим числом слоёв по сравнению с режимом «по всем слоям».
5. Обучение тензорно-сетевой модели многокубитного квантового состояния по данным факторизованных однокубитных измерений позволяет предсказывать вероятности исходов новых измерений без восстановления полной матрицы плотности. В численных экспериментах зависимость минимально необходимого объёма обучающих данных для достижения заданного качества предсказания от числа

кубитов хорошо аппроксимируется линейной зависимостью в рассмотренном диапазоне количества кубитов.

Достоверность эксперимента обеспечивается использованием современных оптических компонентов, приборов и измерительного оборудования. Достоверность численных симуляций опирается на применение распространённых и протестированных программных средств (компиляторов, библиотек алгоритмов). Наблюдается сходство результатов эксперимента и численных расчётов, а также согласие с аналитически ожидаемым поведением. Результаты находятся в соответствии с результатами, полученными другими авторами.

Личный вклад. Все результаты работы получены автором лично или при его непосредственном участии. Автор принимал участие в постановке задач, разработке алгоритмов, программной реализации методов, обработке результатов экспериментов, представленных в диссертационной работе.

Апробация работы. Основные результаты, представленные в диссертационной работе, докладывались на следующих международных и всероссийских конференциях:

1. HOLOEXPO 2023 20-я Международная конференция по голографии и прикладным оптическим технологиям, Сочи, Россия, 2023 год,
2. HOLOEXPO 2024 21-я Международная конференция по голографии и прикладным оптическим технологиям, Казань, Россия, 2024 год,
3. VIII International Conference on Quantum Technologies (ICQT), Москва, Россия, 2025 год,
4. HOLOEXPO 2025 22-я Международная конференция по голографии и прикладным оптическим технологиям, Уфа, Россия, 2025 год,
5. 7-ая Школа молодых учёных в рамках Российского форума «Микроэлектроника», Сочи, Россия, 2025 год,

а также на семинарах кафедры квантовой электроники физического факультета МГУ имени М. В. Ломоносова и семинарах Российского Квантового Центра.

Публикации. Основные результаты работы опубликованы в 4 печатных работах, в том числе в 4 статьях (общим объёмом 3.1 п.л.) в

рецензируемых научных журналах, удовлетворяющих «Положению о присуждении учёных степеней в МГУ имени М. В. Ломоносова».

Статьи по теме диссертации, опубликованные в рецензируемых научных изданиях, рекомендованных для защиты в диссертационном совете МГУ по специальности и отрасли наук:

1. Kuzmin S.S., Dyakonov I.V., Kulik S.P. Architecture agnostic algorithm for reconfigurable optical interferometer programming // Optics Express. — 2021. — Vol. 29, № 23. — pp. 38429–38440. — Импакт-фактор 3,3 (JIF) / 0.75 п.л. / вклад соискателя: теоретическая разработка и программная реализация алгоритма обучения модели машинного обучения линейно-оптического интерферометра, выполнение численных симуляций, EDN: STTIRC.
2. Kuzmin S.S., Mikhailova V.I., Dyakonov I.V., Straupe S.S. Learning the tensor network model of a quantum state using a few single-qubit measurements // Physical Review A. — 2024. — Vol. 109, № 5. — p. 052616. — Импакт-фактор 2,9 (JIF) / 0.55 п.л. / вклад соискателя: теоретическая разработка и программная реализация алгоритма обучения модели машинного обучения квантового состояния в формате тензорной сети, выполнение численных симуляций, EDN: YRNJQN.
3. Kuzmin S.S., Dyakonov I.V., Straupe S.S. Leveraging a feature-based model for linear optical interferometer control // Physical Review A. — 2025. — Vol. 112, №. 5. — p. 053515. — Импакт-фактор 2,9 (JIF) / 0.55 п.л. / вклад соискателя: теоретическая разработка и программная реализация алгоритма обучения модели машинного обучения линейно-оптического интерферометра, выполнение численных симуляций, EDN: SILNRE.
4. Kondratyev I.V., Urusova K.N., Argenchiev A.S., Klushnikov N.S., Kuzmin S.S., Skryabin N.N., Golikov A.D., Kovalyuk V.V., Goltsman G.N., Dyakonov I.V., Straupe S.S., Kulik S.P. Effective programming of a photonic processor with complex interferometric structure // Physical Review Applied. — 2026. — Vol. 25, № 3. — p. 034072. — Импакт-фактор 4,4 (JIF) / 1.25 п.л. / вклад соискателя: теоретическая разработка и программная реализация алгоритма

обучения модели машинного обучения линейно-оптического интерферометра, обработка результатов эксперимента, EDN: PWUGQS.

Объем и структура работы. Диссертация состоит из введения, 5 глав и заключения. Полный объём диссертации составляет 154 страницы, включая 51 рисунок. Список литературы содержит 137 наименований.

Глава 1. Линейно-оптические интерферометры

1.1 Теоретическое описание

Линейно-оптические схемы — это оптические сети, построенные из светоделителей и фазовых задержек, которые могут быть использованы для экспериментальной реализации любого унитарного преобразования [1], которое связывает состояния света в начале и в конце схемы. В них обязательно исключается использование оптических усилителей и других нелинейных элементов.

Линейные оптические интерферометры стремительно превращаются в важнейший инструмент как в квантовой оптике [9], так и в оптической обработке информации [6]. Рост интереса к линейной оптике связан с тем, что технологии интегральной фотонной фабрики становятся всё более доступными для научного сообщества.

Для идеальной линейно-оптической схемы без оптических потерь матрица U , которая описывает схему унитарна. Однако на практике из-за потерь матрица U является подматрицей большей унитарной матрицы. Несмотря на это, знание эффективной матрицы передачи U однозначно характеризует линейное преобразование мод внутри рассматриваемого набора входных и выходных каналов. Линейно-оптическая схема, у которой N входных и N выходных пространственных мод, описывается матрицей $U = (U_{ij})$ размера $N \times N$.

1.1.1 Преобразование когерентных состояний

Линейно-оптическая схема преобразует произвольное многомодовое когерентное состояние света $|\alpha_1, \alpha_2, \dots, \alpha_N\rangle$, которое подаётся на N входных мод в другое многомодовое когерентное состояние $|\beta_1, \beta_2, \dots, \beta_N\rangle$ на выходе схемы:

$$\beta_j = \sum_{i=1}^N U_{ij} \alpha_i. \quad (1.1)$$

1.1.2 Преобразование фоковских состояний

Унитарная матрица U размера $N \times N$, которая описывает линейно-оптическую схему, связывает операторы рождения $a_i^\dagger$ и уничтожения a_i фотонов [10; 11] в i -й моде линейно-оптической схемы в её начале и конце:

$$\begin{pmatrix} \hat{a}'_1 \\ \vdots \\ \hat{a}'_N \end{pmatrix} = U^\dagger \begin{pmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_N \end{pmatrix}, \quad \begin{pmatrix} \hat{a}'_1{}^\dagger \\ \vdots \\ \hat{a}'_N{}^\dagger \end{pmatrix} = U^T \begin{pmatrix} \hat{a}_1{}^\dagger \\ \vdots \\ \hat{a}_N{}^\dagger \end{pmatrix}. \quad (1.2)$$

С точки зрения задачи *boson sampling* [12] особый интерес представляет действие подобной схемы на состояния поля с определённым числом фотонов (фоковские состояния). Действие линейной оптической схемы $\hat{U}$ на состояние с n_i фотонами в i -м входной моде можно описать следующим образом:

$$\hat{U}|n_1, \dots, n_N\rangle = \prod_{i=1}^N \frac{1}{\sqrt{n_i!}} \left(\sum_{k_i=1}^N U_{k_i i} a_{k_i}^\dagger \right)^{n_i} |0\rangle^{\otimes N}. \quad (1.3)$$

Для задачи *boson sampling* наибольший интерес представляет проекция данного состояния $\hat{U}|n_1, \dots, n_N\rangle$ на одно из фоковских состояний $|m_1, \dots, m_N\rangle$ на выходе линейной оптической схемы $\langle m_1, \dots, m_N | \hat{U} | n_1, \dots, n_N \rangle$. Следуя работе [12], получаем, что:

$$\begin{aligned} & \langle m_1, \dots, m_N | \hat{U} | n_1, \dots, n_N \rangle = \\ & \left(\prod_{i=1}^N n_i! \right)^{-1/2} \left(\prod_{j=1}^N m_j! \right)^{-1/2} \text{per } U[(m_1, \dots, m_N) | (n_1, \dots, n_N)], \end{aligned} \quad (1.4)$$

где per обозначает перманент матрицы:

$$\text{per } U = \sum_{\sigma} \prod_{i=1}^N U_{i\sigma_i} \quad (1.5)$$

и сумма ведётся по всевозможным перестановкам σ , а $U[(m_1, \dots, m_N)|(n_1, \dots, n_N)]$, обозначает матрицу $n \times n$, где $n = \sum_{i=1}^N n_i = \sum_{i=1}^N m_i$, составленную из элементов матрицы U по следующему принципу. Числа, стоящие в фигурных скобках после матрицы означают, что в новой матрице индекс определённой строки (столбца) должен встретиться заданное число раз (а именно индекс i -й строки матрицы Λ встречается ровно m_i раз, а индекс j -ого столбца матрицы Λ встречается ровно n_j раз). Например,

$$U[(1,1,1)|(0,2,1)] = \begin{pmatrix} u_{12} & u_{12} & u_{13} \\ u_{22} & u_{22} & u_{23} \\ u_{32} & u_{32} & u_{33} \end{pmatrix}. \quad (1.6)$$

Данная запись означает, что индекс первого столбца не используется ни разу, так как на первой позиции для столбцов стоит 0, в то время как на второй позиции для столбцов стоит 2 и поэтому индекс второго столбца используется 2 раза.

Формула (1.4) позволяет вычислить действие линейной оптической схемы на заданное начальное фоковское состояние и разложить итоговое состояние по фоковским состояниям. Понятно, что вероятности, с которыми будут реализовываться различные распределения фотонов по выходным модам линейно-оптической схемы, будут описываться квадратом модуля этого выражения. Согласно формуле (1.4), для расчёта этих вероятностей требуется вычисление перманентов, что является сложной задачей для классического компьютера [13]. В реальном эксперименте эти вероятности являются малыми [14], поэтому точность настройки требуемой унитарной матрицы должна быть высокой.

1.2 Дизайн линейно-оптических схем

Главное достоинство современных интегральных линейных интерферометров — перенастраиваемость (реконфигурируемость), то есть возможность изменять воздействие на входной оптический режим по запросу пользователя. Эта способность делает линейные оптические схемы особенно привлекательными для решения задач обработки информации.

В частности, реконфигурируемые интерферометры являются основными компонентами современных экспериментов по линейным оптическим квантовым вычислениям [15—17] и рассматриваются как аппаратные ускорители для приложений глубокого обучения [18; 19].

Кроме того, качество изготовления и улучшенная масштабируемость реконфигурируемых фотонных схем привели к появлению концепции программируемой фотонной матрицы — многоцелевой фотонной схемы, которая может иметь множество возможных применений с помощью низкоуровневого программирования устройства [20].

Универсальные многомодовые линейно-оптические схемы могут быть реализованы в виде многопортовых интерферометров и могут быть запрограммированы для реализации любого линейного преобразования между несколькими модами. Эти интерферометры, как правило, состоят из регулярной сетки светоделителей и фазовых задержек, которые можно варьировать, что обеспечивает простоту их изготовления и масштабируемость с использованием интегральных фотонных архитектур. Отметим, что универсальная линейно-оптическая схема, которая реализует произвольное унитарное преобразование из группы $U(N)$ должна содержать N^2 варьируемых действительных параметров.

1.2.1 Стандартные дизайны

Стандартные дизайны для универсального мультипортового интерферометра были предложены в работах [1] (*треугольный дизайн*, рисунок 1.2) и [2] (*прямоугольный дизайн*, рисунок 1.3). Такой N -модовый интерферометр может быть реализован с использованием сетки из $\frac{N(N-1)}{2}$ светоделителей с изменяемыми параметрами φ и θ . Светоделители изготавливают в виде интерферометров Маха-Цандера (рисунок 1.1), так как в такой конфигурации коэффициент деления можно варьировать путём изменения фазы θ . Такой контролируемый светоделитель осуществляет следующее унитарное преобразование 2×2 :

$$U_{BS} = \begin{pmatrix} e^{i\varphi} \cos \theta & -\sin \theta \\ e^{i\varphi} \sin \theta & \cos \theta \end{pmatrix}. \quad (1.7)$$

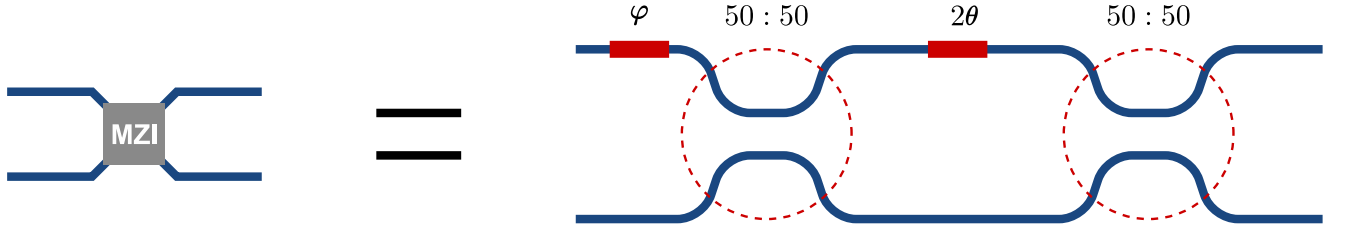


Рисунок 1.1 — Строительный блок линейно-оптической схемы в стандартном дизайне — интерферометр Маха-Цандера. Фазы φ и θ являются варьируемыми параметрами. Пунктиром обведены места, в которых излучение может перераспределяться между каналами.

Унитарная матрица размера $N \times N$ для всех N мод для контролируемого светоделителя между i -й и j -й модой ($j = i + 1$) записывается в виде следующей блочной матрицы (верхний правый элемент U_{BS} располагается в общей матрице на позиции с индексами (i, i) , то есть матрица U_{BS} находится на диагонали общей матрицы):

$$U_{i,j} = \begin{pmatrix} I & \dots & O & \dots & O \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ O & \dots & U_{BS} & \dots & O \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ O & \dots & O & \dots & I \end{pmatrix}, \quad (1.8)$$

где I — единичная матрица размера 2×2 , O — нулевая матрица размера 2×2 , а U_{BS} определяется формулой (1.7).

Общая унитарная матрица U размера $N \times N$, которую реализует линейно-оптическая схема, записывается в виде произведения [2]:

$$U = D \left(\prod_{(i,j) \in S} U_{i,j} \right), \quad (1.9)$$

где S определяет конкретную упорядоченную последовательность двухмодовых преобразований (1.7), которая в случае треугольного и прямоугольного дизайнов различаются, и где D — диагональная матрица с комплексными элементами с модулем, равным единице. Матрица

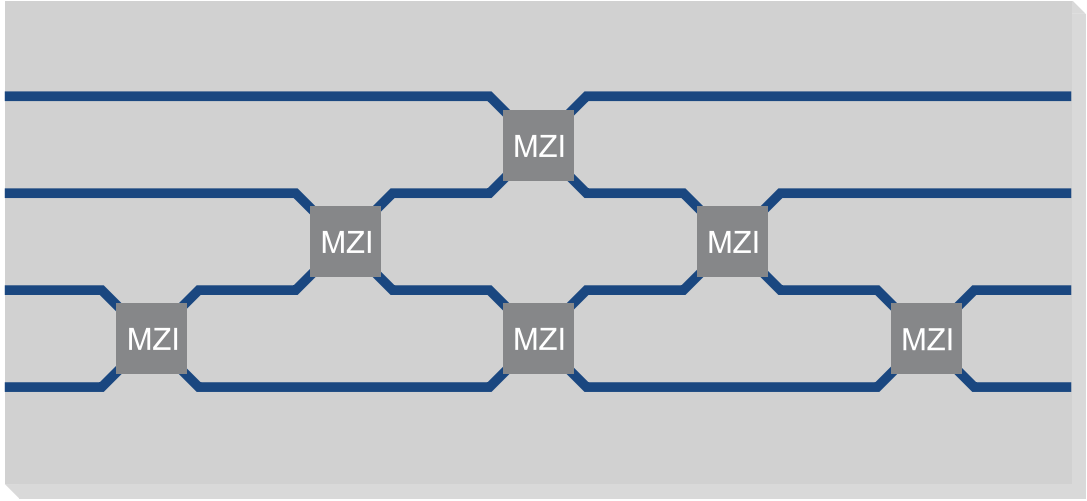


Рисунок 1.2 — Треугольный дизайн линейно-оптических схем, предложенный в работе [1]. Пересечения линий соответствуют интерферометрам Маха-Цандера, изображённым на рисунке 1.1 с двумя управляемыми фазами. Для того, чтобы иметь возможность реализовывать произвольную унитарную матрицу из $U(N)$ нужно добавить на выход схемы ещё N фазовых задержек (тогда общее число фаз будет N^2).

$D = \text{diag}(e^{i\varphi_1}, \dots, e^{i\varphi_N})$ определяет фазовые задержки на выходе линейно-оптической схемы во всех пространственных модах. Стоит отметить, что во многих физических приложениях матрица D опускается, так как схемы с различными D при обычных измерениях, проводимых с такими устройствами, эквивалентны друг другу (нет никаких проблем измерить и эти фазы, тогда D опускать нельзя). Отметим, что с учётом матрицы D у нас как раз получается $2\frac{N(N-1)}{2} + N = N^2$ действительных параметров. Физический интерферометр состоит из светоделителей и фазовых задержек в определённой конфигурации, определенной S , со значениями, определенными θ и φ в каждой матрице U_{BS} поэтому он будет реализовывать унитарное преобразование U , определённое выражением (1.9).

В работах [1; 2] описывается аналитический алгоритм для обоих дизайнов, позволяющий по заданному требуемому унитарному преобразованию $\bar{U}$ настроить фазовые задержки θ и φ во всех элементарных блоках, так чтобы линейно-оптическая схема рассматриваемого дизайна определяла преобразование $\bar{U}$.

Отметим, что хотя общее количество светоделителей в дизайнах равны $\frac{N(N-1)}{2}$, прямоугольный дизайн [2] явно имеет гораздо меньшую оптическую глубину, чем треугольный [1] и, следовательно, имеет меньшие потери при

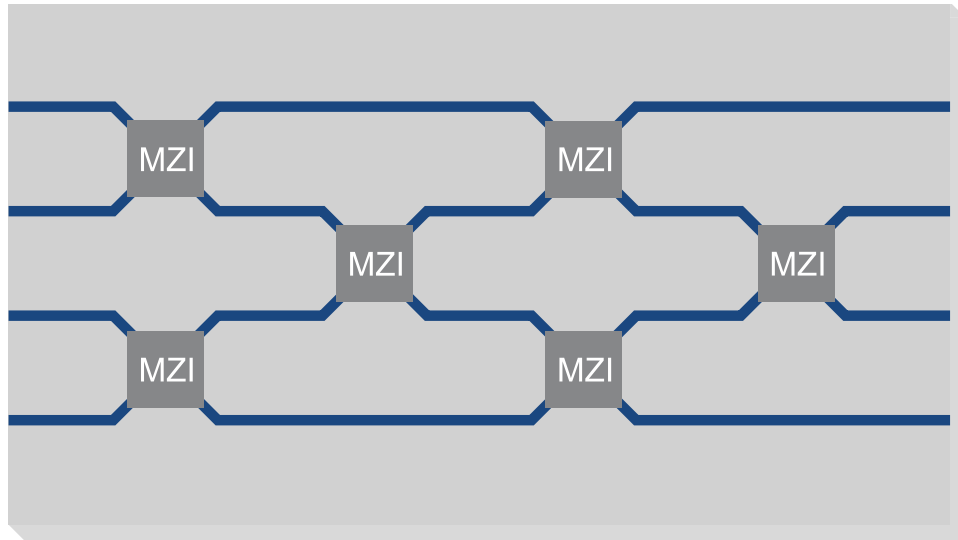


Рисунок 1.3 — Прямоугольный дизайн линейно-оптических схем, предложенный в работе [2]. Пересечения линий соответствуют интерферометрам Маха-Цандера, изображённым на рисунке 1.1 с двумя управляемыми фазами. Для того, чтобы иметь возможность реализовывать произвольную унитарную матрицу из $U(N)$ нужно добавить на выход схемы ещё N фазовых задержек (тогда общее число фаз будет N^2).

распространении в нём света. Это уменьшение оптической глубины происходит из того факта, что каждая мода пересекает своего ближайшего соседа при первом возможном случае в отличие от треугольной схемы, где верхние моды должны распространяться на некоторое расстояние, прежде чем взаимодействовать с другими модами. Кроме того, прямоугольная конструкция обладает большей симметрией, что повышает устойчивость к потерям в интерферометре [21; 22].

1.2.2 Блочный дизайн

Чтобы схемы, с треугольным или прямоугольным дизайном были универсальными, необходимо, чтобы светоделители, которые составляют двухмодовые интерферометры Маха-Цандера, были идеально сбалансированы (коэффициент деления должен быть строго 50:50), поэтому предъявляются строгие требования на их изготовление. Однако на практике это условие не выполняется полностью, так что многомодовые схемы универсальны только

с определенной степенью точности [21]. Этот отрицательный эффект становится все более значимым при масштабировании схемы интерферометра [22]. Поэтому реализация многомодовых линейно-оптических схем требует разработки новой архитектуры для многомодовой интерференции, более устойчивой к ошибкам реализации.

В работе [23] предлагается новый метод декомпозиции произвольного унитарного преобразования на основе многомодовых блоков — будем называть их *базисными матрицами*. В работе показано, что схема универсальна, даже когда базисные матрицы выбираются случайным образом из пространства унитарных матриц, что делает её универсальной, практически независимо от ошибок изготовления. Естественно, такая схема не всегда будет универсальной, например, если взять в качестве базисных матриц — единичные, то моды совсем не будут перемешиваться, то есть схема точно не будет в этом случае универсальной (то есть необходимо, чтобы базисные матрицы перемешивали моды, но к ним уже не выдвигается более жёстких требований). Метод может быть полезен для высокоразмерных унитарных преобразований. В указанной работе рассматривается несколько похожих дизайнов, которые несколько отличаются друг от друга количеством базисных матриц и расположением фаз. Мы выберем для изучения дизайн линейно-оптического интерферометра, изображённый на рисунке 1.4 (N базисных матриц, и N^2 фазовых задержек с целью возможности реализации произвольного преобразования из унитарной группы $U(N)$). Пример реализации такого дизайна для $N = 4$ показан на рисунке 1.5.

Для данного дизайна справедлива связь итоговой унитарной матрицы U с базисными матрицами U_ℓ и фазовыми задержками $\varphi_{\ell k}$:

$$U = \Phi_{N+1} U_N \Phi_N U_{N-1} \dots \Phi_{\ell+1} U_\ell \Phi_\ell \dots \Phi_2 U_1 \Phi_1, \quad (1.10)$$

где $\Phi_\ell = \text{diag}(e^{i\varphi_{\ell 1}}, \dots, e^{i\varphi_{\ell(N-1)}}, 1)$ при $\ell = 1, \dots, N$ и для выходного слоя $\Phi_{N+1} = \text{diag}(e^{i\varphi_{(N+1)1}}, \dots, e^{i\varphi_{(N+1)(N-1)}}, e^{i\varphi_{(N+1)N}})$.

Итак, данный дизайн оказывается практически нечувствительным к ошибкам изготовления базисных матриц, а потому является перспективным при создании крупномасштабных линейно-оптических преобразований.

Наряду с треугольным дизайном Река [1], прямоугольным дизайном Клементса [2] и блочными робастными схемами [23], в литературе рассматриваются и другие варианты построения программируемых многопортовых

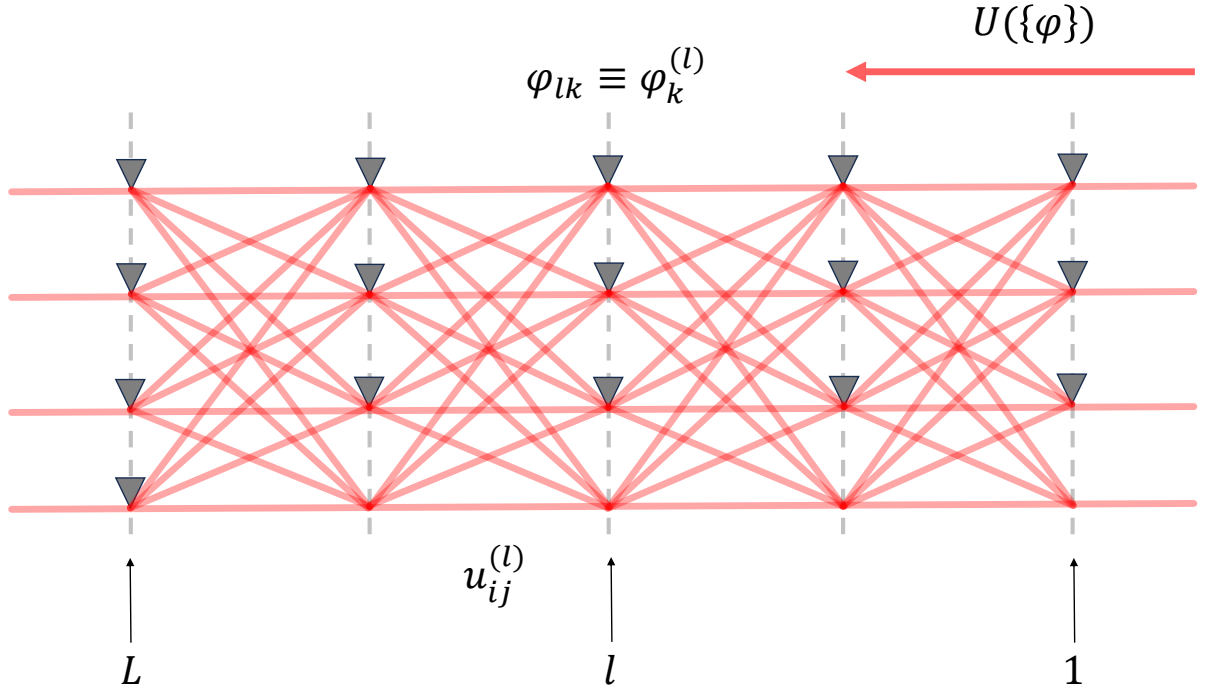


Рисунок 1.4 — Схема структуры блочного дизайна многомодового интерферометра. Базисные матрицы U_ℓ с элементами $u_{ij}^{(\ell)}$ описывают перераспределение излучения между каналами интерферометра. Соответствующие интегральные элементы могут быть реализованы в виде решеток связанных волноводов, что позволяет передавать энергию между различными волноводами [24].

интерферометров. Эти варианты мотивированы главным образом необходимостью уменьшения оптической глубины схемы, повышения устойчивости к ошибкам изготовления и упрощения интегральной реализации. Так, в работе [25] предложен error-tolerant дизайн, в котором вместо интерферометра Маха-Цандера используется элементарный блок, состоящий из одного статического светоделителя и перестраиваемого фазового сдвига. Такая архитектура сохраняет возможность реализации широкого класса преобразований и при этом демонстрирует повышенную устойчивость к технологическим отклонениям и потерям.

Кроме того, в более поздних работах [26; 27] предложены компактные архитектуры малой глубины, ориентированные на реализацию матрично-векторного умножения для неунитарных матриц. В работе [26] используются компактные малоглубинные сетки светоделителей, более подходящие для планарной интегральной фотоники, чем полносвязные многомодовые блоки, а в статье [27] рассматривается two-unitary дизайн,

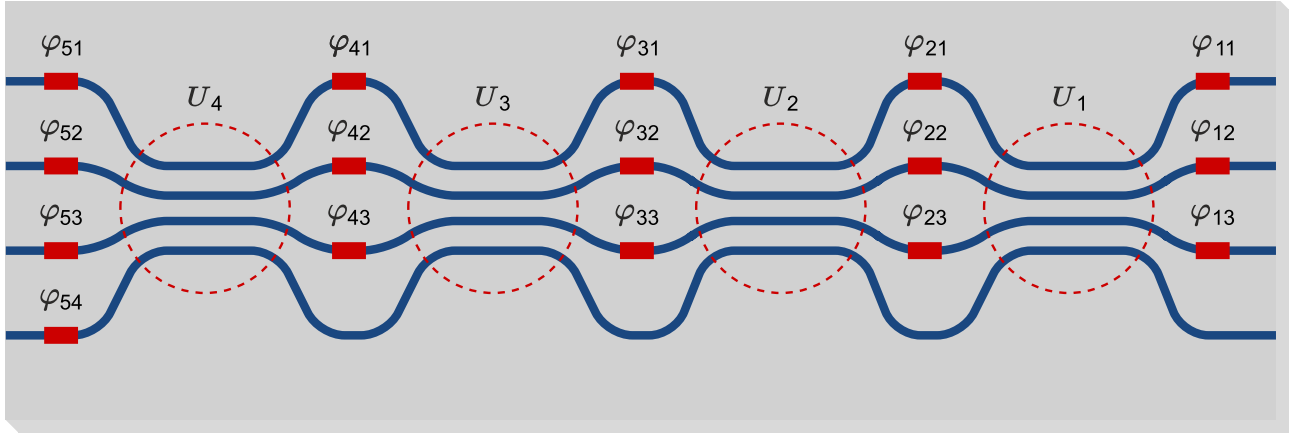


Рисунок 1.5 — Пример четырёхканального (четырёхмодового) интерферометра. Красным помечены места контроля фазовых задержек $\varphi_{\ell k}$. Пунктиром обведены места, в которых излучение может переходить между каналами.

основанный на представлении общей неунитарной матрицы в виде суммы двух унитарных преобразований. Тем самым, наряду с универсальностью, важными критериями качества современных дизайнов становятся также глубина схемы, компактность, устойчивость к аппаратным ошибкам и аналитическая программируемость.

Следует отметить, что обобщённая форма разложения (1.10), рассмотренная в [23], носит универсальный характер и может быть применена к произвольной конструкции линейно-оптического интерферометра. Нетрудно показать, что любой оптический интерферометр, построенный из независимых компонентов — фиксированных линейно-оптических элементов произвольной топологии и фазовых модуляторов, — допускает представление в виде произведения унитарных преобразований, соответствующих фиксированным преобразованиям и фазовым сдвигам. Например, схема Клементса, представленная на рисунке 1.6, может быть приведена к блочному виду, показанному на рисунке 1.5, а её математическая форма разложения задаётся выражением (1.10).

Структурная гибкость элементов смешивания мод, а также отсутствие ограничений на количество, конфигурацию и порядок следования модовых смесителей и фазовых сдвигов обеспечивают архитектурную независимость алгоритмов, предложенных в главах 2 и 3 настоящей диссертации.

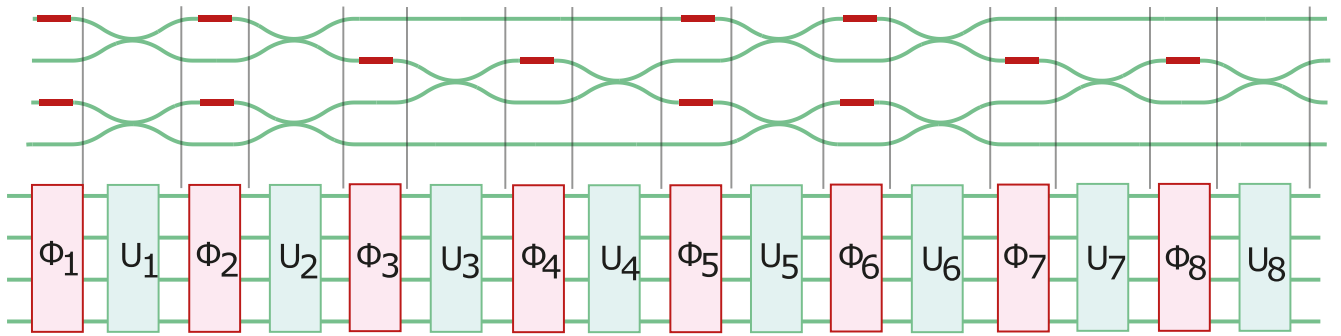


Рисунок 1.6 — Представление линейно-оптической схемы Клемента в виде схемы с блочным дизайном, представленным на рисунке 1.5.

1.3 Изготовление

1.3.1 Литография

Наиболее популярной технологией изготовления линейно-оптических схем в виде интегрально-оптических интерферометров является *литография* [28—30]. Она предоставляет все необходимые инструменты для создания достаточно качественных устройств. В основном при создании линейно-оптических схем с помощью данной технологии используются многослойные структуры на кремниевой подложке, в которых формируются волноводные и вспомогательные слои из кремния, нитрида кремния и диоксида кремния. Также с помощью литографии можно наносить на поверхность структур металлические плёнки, предназначенные для осуществления реконфигурации преобразований схемы [9; 31] с помощью термооптического эффекта. Большим преимуществом метода литографии является возможность создания структур малого размера, высокая точность и быстрый отклик на контролирующее термооптическое воздействие.

1.3.2 Фемтосекундная лазерная печать

Альтернативный подход к изготовлению — *фемтосекундная лазерная печать* (ФСЛП)[32] — модификация материала с помощью жёстко сфоку-

сированного импульсного лазерного излучения, которая используется для изготовления подобных линейно-оптических схем в Центре квантовых технологий Физического факультета МГУ [33]. Основными преимуществами технологии ФСЛП является относительная простота и доступность. ФСЛП не требует от специалиста, занимающегося квантовой оптикой, почти никаких дополнительных навыков, в отличие от литографии. Недостатком ФСЛП является относительно большой размер структур по сравнению с их размером при изготовлении с помощью литографии. Это является следствием невысокого контраста показателя преломления, что вынуждает делать более плавные изгибы волноводов.

1.4 Методы томографии преобразования

В данной части рассматриваются экспериментальные методы восстановления (*томографии*) матрицы линейно-оптического преобразования, реализуемого интерферометром. В простейшей постановке предполагается, что рассматриваемое преобразование описывается унитарной матрицей. Для восстановления матрицы преобразования могут использоваться как когерентные состояния света, так и фоковские состояния. Вначале рассмотрим метод томографии с использованием когерентных состояний.

1.4.1 Когерентные состояния

В работе [4] предложен метод прямого восстановления неизвестной матрицы линейно-оптического преобразования с использованием когерентных состояний света. Экспериментально такой подход требует только лазерного источника, фазового модулятора и детекторов интенсивности.

Элементы матрицы U (не обязательно унитарной) являются комплексными числами, которые удобно представлять в виде $U_{ij} = r_{ij}e^{i\varphi_{ij}}$, где $r_{ij} = |U_{ij}|$ — модуль комплексного числа U_{ij} , а φ_{ij} — его фаза.

Матрица U связывает когерентное состояние света $|\beta_j\rangle$ в j -й моде, с когерентными состояниями $|\alpha_i\rangle$ в i -ых модах на входе:

$$\beta_j = \sum_{i=1}^N U_{ij} \alpha_i. \quad (1.11)$$

Следовательно, задача томографии сводится к определению модулей r_{ij} и фаз φ_{ij} всех элементов матрицы U .

1. *Определение модулей $r_{ij} = |U_{ij}|$:*

Если подать когерентное состояние с интенсивностью I только в i -ю входную моду, оставив остальные входы в вакуумном состоянии, то интенсивность в j -й выходной моде равна

$$I_j = \left| \sum_{i=1}^N U_{ij} \alpha_i \right|^2 = |U_{ij}|^2 I_i = r_{ij}^2 I. \quad (1.12)$$

Отсюда получаем

$$r_{ij} = \sqrt{\frac{I_j}{I}}. \quad (1.13)$$

2. *Определение фаз φ_{ij} :*

Для того, чтобы определить φ_{ij} направим сначала когерентное состояние лазера на светоделитель с коэффициентом деления 50:50, а затем отправим полученные когерентные состояния на вход линейно-оптической схемы в 1-ую и i -ую моды (рисунок 1.7). К когерентному состоянию, направленному в i -ую моду применим фазовую задержку φ .

Измерим интенсивность на выходе в j -й моде:

$$I_j = |U_{1j} + U_{ij} e^{i\varphi}|^2. \quad (1.14)$$

Поскольку фазы можно определить лишь с точностью до выбора опорных фаз, без ограничения общности фазы элементов первой входной моды можно принять равными нулю $\varphi_{1j} = 0$:

$$I_1 = I (r_{11}^2 + r_{i1}^2 + 2r_{11}r_{i1} \cos(\varphi)), \quad (1.15)$$

$$I_j = I (r_{1j}^2 + r_{ij}^2 + 2r_{1j}r_{ij} \cos(\varphi + \varphi_{ij})). \quad (1.16)$$

Из зависимости выходной интенсивности от фазового сдвига φ восстанавливается фаза φ_{ij} .

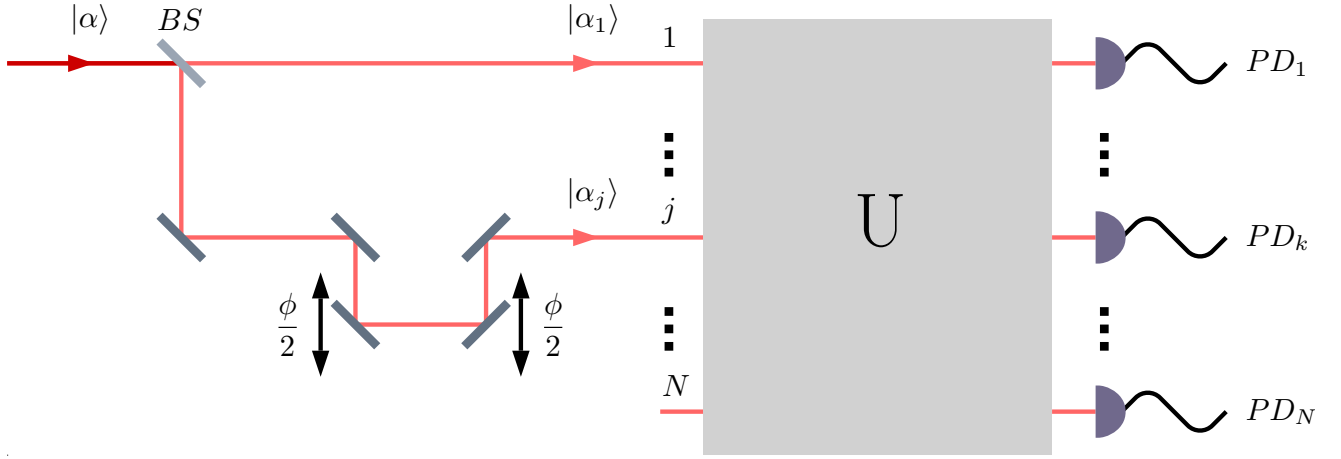


Рисунок 1.7 — Томография когерентными состояниями.

Таким образом, измерения выходных интенсивностей при одно- и двухмодовых когерентных входных состояниях позволяют восстановить модули и фазы всех элементов матрицы линейно-оптического преобразования U .

1.4.2 Фоковские состояния

Метод *Super Stable Tomography* (SST) был представлен в работе [3]. В нём используется статистика измерений прохождения пар неразличимых фотонов через линейно-оптическую схему. Преимущество метода — нечувствительность к потерям на входах в моды в линейно-оптической сети и к эффективности детектирования.

Рассмотрим часть линейно-оптической сети, которая описывается матрицей размера 2×2 :

$$\begin{pmatrix} \tau_{j,k} e^{i\alpha_{j,k}} & \tau_{j,h} e^{i\alpha_{j,h}} \\ \tau_{g,k} e^{i\alpha_{g,k}} & \tau_{g,h} e^{i\alpha_{g,h}} \end{pmatrix}, \quad (1.17)$$

где k, h — входные моды, а j, g — выходные. Пусть коэффициенты передачи на входах k и h равны, соответственно, s_k и s_h , а эффективности детектирования в выходных модах j и g равны, соответственно, r_j и r_g .

Тогда преобразование с учётом входных потерь и эффективности детектирования имеет вид $P_2 = D_d S_2 D_c$:

$$P_2 = \begin{pmatrix} \sqrt{r_j} & 0 \\ 0 & \sqrt{r_g} \end{pmatrix} \begin{pmatrix} \tau_{j,k} e^{i\alpha_{j,k}} & \tau_{j,h} e^{i\alpha_{j,h}} \\ \tau_{g,k} e^{i\alpha_{g,k}} & \tau_{g,h} e^{i\alpha_{g,h}} \end{pmatrix} \begin{pmatrix} \sqrt{s_k} & 0 \\ 0 & \sqrt{s_h} \end{pmatrix}. \quad (1.18)$$

Инжектируем фоковское состояние с S_k фотонами в k -ую моду схемы, а в другие моды — вакуумные состояния (рисунок 1.8). Тогда количество задетектированных фотонов в j -й моде будет равно $R_{j,k} = s_k \tau_{j,k}^2 r_j S_k$.

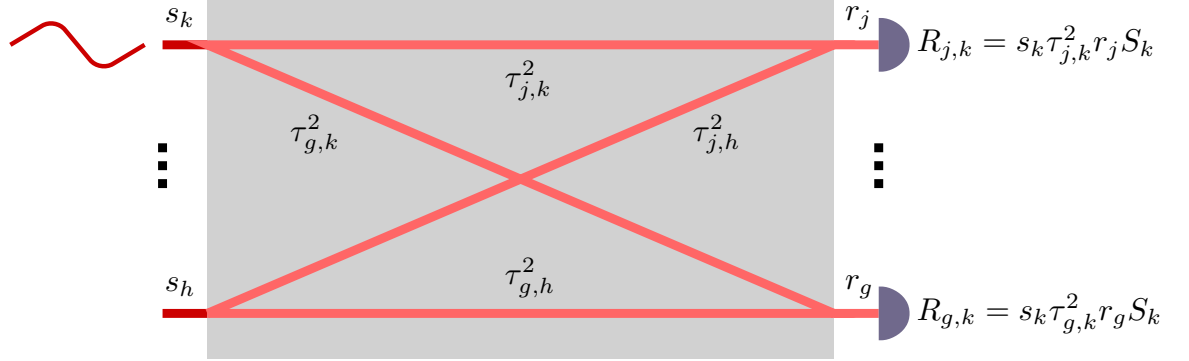


Рисунок 1.8 — Томография фоковскими состояниями для определения модулей.

Введём вспомогательные величины $X_{g,h,j,k}$ и $x_{g,h,j,k}$, которые выражаются через экспериментально измеряемые фотоотсчёты и не зависят от коэффициентов передачи на входе, эффективностей детектирования и числа инжектированных фотонов:

$$X_{g,h,j,k} \equiv \frac{R_{j,k} R_{g,h}}{R_{j,h} R_{g,k}} = \left(\frac{\tau_{j,k} \tau_{g,h}}{\tau_{j,h} \tau_{g,k}} \right)^2, \quad (1.19)$$

$$x_{g,h,j,k} \equiv \frac{\tau_{j,k} \tau_{g,h}}{\tau_{j,h} \tau_{g,k}} = \sqrt{\frac{R_{j,k} R_{g,h}}{R_{j,h} R_{g,k}}}, \quad (1.20)$$

где $R_{j,k}$ — отсчёты фотодетекторов, непосредственно измеряемые в эксперименте.

Для того чтобы извлечь информацию о фазах элементов матрицы преобразования, необходимо измерить результат прохождения пары неразличимых фотонов через линейно-оптическую схему.

Инжектируем в k -ую и h -ую входные моды два неразличимых фотона, и будем фиксировать количество совпадений в фотоотсчётах детекторов в выходных модах j и g (рисунок 1.9). Тогда теоретически будет следующее количество совпадений:

$$Q_{g,h,j,k} = |\text{per } P_2|^2 = s_h s_k r_g r_j (\tau_{j,k}^2 \tau_{g,h}^2 + \tau_{g,k}^2 \tau_{j,h}^2) + 2s_h s_k r_g r_j \tau_{j,k} \tau_{g,h} \tau_{g,k} \tau_{j,h} \cos(\alpha_{j,k} - \alpha_{j,h} - \alpha_{g,k} + \alpha_{g,h}) . \quad (1.21)$$

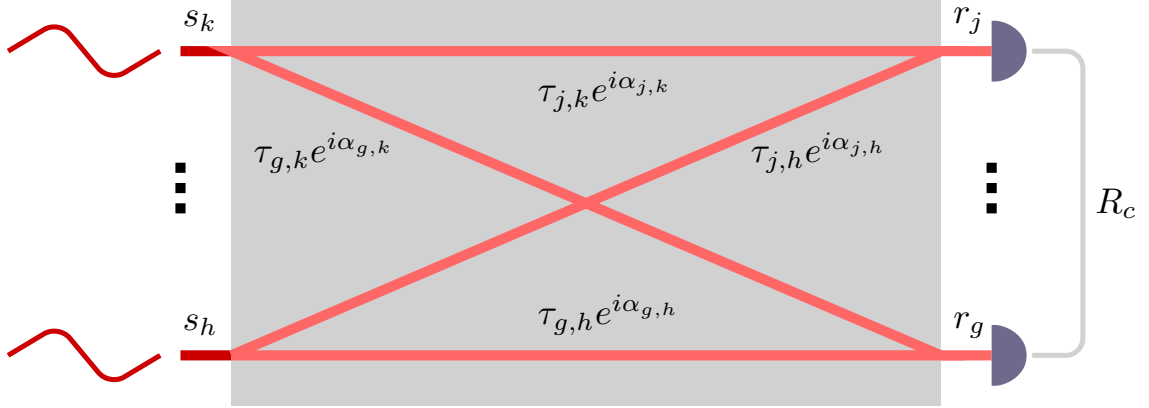


Рисунок 1.9 — Томография фоковскими состояниями для определения фаз.

Теперь введём фазовую задержку между фотонами, чтобы они стали различимыми и интерференция между ними исчезла. Тогда количество совпадений будет равно:

$$C_{g,h,j,k} = \text{per } |P_2|^2 = s_h s_k r_g r_j (\tau_{j,k}^2 \tau_{g,h}^2 + \tau_{g,k}^2 \tau_{j,h}^2) = R_{j,k} R_{g,h} + R_{g,k} R_{j,h} . \quad (1.22)$$

Далее введём глубину провала Хонга–У–Мандела [34]:

$$V_{g,h,j,k} \equiv \frac{C_{g,h,j,k} - Q_{g,h,j,k}}{C_{g,h,j,k}} = -2 \cos(\alpha_{j,k} - \alpha_{j,h} - \alpha_{g,k} + \alpha_{g,h}) \frac{\tau_{j,k} \tau_{g,h} \tau_{g,k} \tau_{j,h}}{(\tau_{j,k}^2 \tau_{g,h}^2 + \tau_{g,k}^2 \tau_{j,h}^2)} . \quad (1.23)$$

Данная видность — экспериментально измеряемая величина. Для того, чтобы её измерить, необходимо сдвигать фотоны во времени друг относительно друга и наблюдать, как они перестают перекрываться во времени, при этом будет изменяться количество совпадений отсчётов фотодетекторов.

Введём вспомогательные величины:

$$y_{g,h,j,k} = x_{g,h,j,k} + x_{g,h,j,k}^{-1} . \quad (1.24)$$

Тогда фазы α_{jk} выражаются следующим образом:

$$\cos(\alpha_{j,k} - \alpha_{j,h} - \alpha_{g,k} + \alpha_{g,h}) = -\frac{V_{g,h,j,k} y_{g,h,j,k}}{2} . \quad (1.25)$$

Надо отметить, что формула (1.25) неустойчива в случае определения фаз элементов матрицы с очень маленьким модулем. Произведение $V_{g,h,j,k}$ и $y_{g,h,j,k}$ должно быть ограничено двойкой. В эксперименте же, в случае маленького элемента, $y_{g,h,j,k}$ — очень большое число. Небольшая неточность или недостаточная статистика могут привести к тому, что мы выйдем из области значений функции $\cos$.

Итак, экспериментально измеряемыми данными в данном случае являются отсчёты фотодетекторов $R_{j,k}$ и видности провалов Хонга–У–Манделя $V_{g,h,j,k}$.

В работе [3] показано, что экспериментальные данные не зависят от фазовых задержек, внесённых в каждый из каналов на входе или выходе линейно-оптической схемы, следовательно, восстанавливаемую матрицу U можно искать в виде:

$$U = \begin{pmatrix} \tau_{1,1} & \tau_{1,2} & \cdots & \tau_{1,N} \\ \tau_{2,1} & \tau_{2,2}e^{i\alpha_{2,2}} & \cdots & \tau_{2,N}e^{i\alpha_{2,N}} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{N,1} & \tau_{N,2}e^{i\alpha_{N,2}} & \cdots & \tau_{N,N}e^{i\alpha_{N,N}} \end{pmatrix}. \quad (1.26)$$

Далее, используя формулу (1.25), можно определить фазы с учётом того, что в первой строке и первом столбце они полагаются равными нулю. После этого с использованием формулы (1.20) и дополнительных ограничений, налагаемых на матрицу преобразования, можно восстановить модули её элементов.

Глава 2. Обучение модели машинного обучения линейно-оптического интерферометра

В главе 1 было показано, что программирование реконфигурируемого линейно-оптического интерферометра сводится к подбору управляющих параметров, обеспечивающих реализацию требуемого преобразования. Для стандартных универсальных архитектур, таких как схемы Река и Клементса [1; 2], эта задача существенно упрощается благодаря наличию аналитических или полуаналитических процедур декомпозиции целевой унитарной матрицы. Однако в реальных устройствах отклонения параметров светоделителей от проектных значений приводят к ухудшению точности программирования, а использование численной оптимизации вместо аналитической настройки существенно увеличивает вычислительные и экспериментальные затраты [21; 35].

Как обсуждалось ранее, развитие блочных (робастных) архитектур многопортовых интерферометров [23; 25] позволило существенно ослабить требования к точности изготовления фиксированных смешивающих моды элементов. Однако такая устойчивость достигается ценой потери явного аналитического соответствия между требуемым унитарным преобразованием и набором управляющих фаз. В результате программирование устройства становится задачей многопараметрической оптимизации.

Прямая оптимизация фазовых сдвигов на физическом устройстве имеет ряд существенных недостатков. На каждой итерации необходимо экспериментально восстанавливать реализуемую устройством матрицу преобразования [5; 36], что делает процедуру медленной и ресурсоёмкой. Кроме того, высокая размерность пространства параметров, наличие локальных минимумов и чувствительность градиентных методов к шумам измерений затрудняют надёжное применение таких подходов на практике. Поэтому для сложных линейно-оптических архитектур актуальна разработка методов программирования, в которых дорогостоящая работа с физическим устройством выполняется заранее, а дальнейшая настройка осуществляется уже с использованием его численной модели.

В данной главе предлагается архитектурно-независимый алгоритм программирования линейно-оптического интерферометра, основанный на

идеях *обучения с учителем*, широко применяемых для обучения нейронных сетей [37—39]. Подход состоит в построении параметризованной модели интерферометра по обучающей выборке, полученной с исследуемого устройства. После этого поиск фазовых сдвигов для реализации нового унитарного преобразования осуществляется не на самом устройстве, а на его обученной модели. Тем самым задача разделяется на два этапа: обучение модели по набору экспериментально восстановленных преобразований и последующее программирование устройства с использованием найденной модели. Отметим, что в данной диссертации рассмотрен только первый этап обучения, так как этап программирования уже хорошо изучен и описан, например, в работе [23].

2.1 Формулировка проблемы

Рассмотрим многомодовый линейно-оптический интерферометр, состоящий из чередующихся фазосдвигающих и модосмешивающих слоёв. Как было показано в разделе 1.2.2, широкий класс реконфигурируемых линейно-оптических архитектур допускает представление в виде произведения фиксированных унитарных преобразований и диагональных фазовых матриц. Для выбранной архитектуры матрица преобразования интерферометра описывается выражением

$$U = \Phi_{N+1} U_N \Phi_N U_{N-1} \dots \Phi_{\ell+1} U_\ell \Phi_\ell \dots \Phi_2 U_1 \Phi_1, \quad (2.1)$$

где $\Phi_\ell = \text{diag}(e^{i\varphi_{\ell 1}}, \dots, e^{i\varphi_{\ell N}})$, $\ell = 1, \dots, N + 1$. Мы будем называть U_ℓ *базисными матрицами*, поскольку они полностью определяют работу интерферометра. Если U_ℓ известны, простая численная оптимизация находит соответствующие фазовые сдвиги $\varphi_{\ell k}$, тем самым завершая задачу программирования устройства.

Проблема, обуславливающая сложность программирования требуемого унитарного преобразования в многомодовой архитектуре, заключается в том, что базисные матрицы U_ℓ изготовленного устройства не соответствуют матрицам, заложенным в оптическую схему. Универсальность интерферометра не снижается, но эффективная оценка фазовых сдвигов

$\varphi_{\ell k}$ становится невозможной, поскольку U_ℓ больше не известны. Это подводит нас к первому этапу нашего алгоритма обучения — реконструкции параметров базисных матриц с использованием информации из обучающего набора $\mathcal{T}$, собранной с интересующего нас устройства (см. рисунок 2.1).

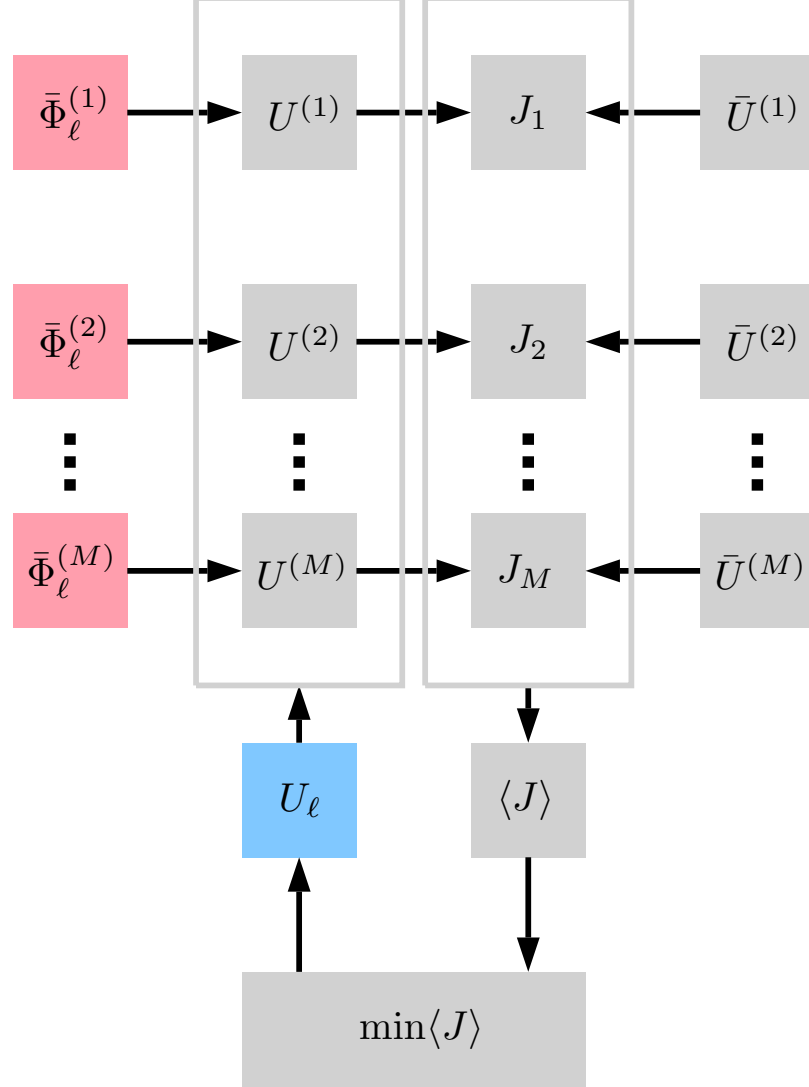


Рисунок 2.1 — Задача обучения модели линейно-оптического интерферометра. Обучаемыми параметрами являются комплексные элементы базисных матриц (выделены синим), обучающей выборкой является множество пар $(\bar{\Phi}_\ell^{(k)}, \bar{U}^{(k)})$. Обучение заключается в минимизации среднего функционала $\langle J \rangle$ между унитарными матрицами из обучающей выборки и унитарными матрицами $U^{(k)}$, которые выдаёт модель при тех же фазах (выделены красным).

Множество $\mathcal{T}$ включает M пар $(\bar{\Phi}^{(i)}, \bar{U}^{(i)})$, полученных путём заполнения устройства случайными фазовыми сдвигами $\bar{\Phi}^{(i)}$ и применения процедуры унитарной реконструкции по выбору [3; 4] (см. разделы 1.4.1 и

1.4.2) для получения $\bar{U}^{(i)}$. Затем базисные матрицы U_ℓ определяются как решение задачи оптимизации

$$\{U_\ell\} = \underset{\{U_\ell\}}{\operatorname{argmin}} \langle J(\bar{U}, U(\{U_\ell\}, \bar{\Phi})) \rangle_{\mathcal{T}}, \quad (2.2)$$

где показатель качества усредняется по обучающему набору $\mathcal{T}$. После определения базисных матриц мы переходим ко второму этапу алгоритма (см. рисунок 2.2) — находим соответствующие фазовые сдвиги $\phi_{\ell k}$, которые подстраивают модель интерферометра $\mathcal{M}$ к унитарной матрице $\bar{U} \notin \mathcal{T}$.

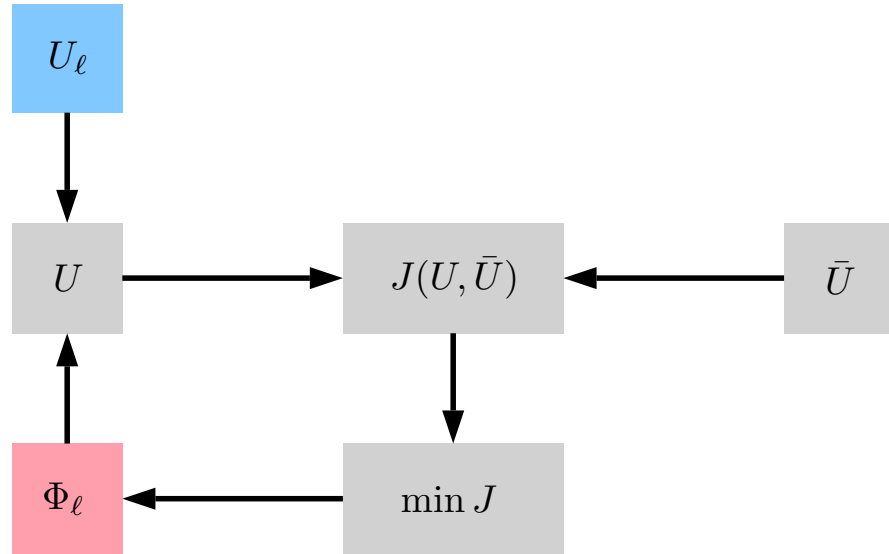


Рисунок 2.2 — Задача настройки модели линейно-оптического интерферометра. В данном случае базисные матрицы (выделены синим) фиксированы, меняются только фазы (выделены красным). В процессе настройки фазы меняются так, чтобы минимизировать функционал невязки J между U и $\bar{U}$.

2.2 Алгоритм обучения

В этом разделе представлен алгоритм обучения модели интерферометра $\mathcal{M}$ на основе исходных данных, содержащихся в обучающей выборке $\mathcal{T}$. Мы сводим задачу обучения к многопараметрической оптимизации нелинейного функционала J . В этом разделе представлена математическая структура алгоритма обучения и продемонстрирована его эффективность на многомодовом интерферометре.

2.2.1 Вычисление метрики

Показатель качества J , который будет изучаться в нашей работе, — это норма Фробениуса J_{FR} :

$$J_{FR}(U, \bar{U}) \equiv \frac{1}{N} \sum_{i,j=1}^N |u_{ij} - \bar{u}_{ij}|^2. \quad (2.3)$$

Он инвариантен только относительно тождественного преобразования, то есть $J_{FR}(U, \bar{U}) = 0$, только если амплитуда и фаза матричных элементов U и $\bar{U}$ одинаковы. Выражение 2.3 можно переписать, используя операцию произведения Адамара $(A \odot B)_{i,j} = (A)_{i,j} \cdot (B)_{i,j}$, и оно принимает следующий вид:

$$J_{FR}(U, \bar{U}) = \frac{1}{N} \sum_{i,j=1}^N ((U - \bar{U}) \odot (U - \bar{U})^*)_{i,j}. \quad (2.4)$$

Градиент J_{FR} относительно набора параметров $\{\bar{\alpha}\}$ определяется по формуле:

$$\partial_{\bar{\alpha}} J_{FR}(U(\bar{\alpha}), \bar{U}) = \frac{2}{N} \text{Re} \sum_{i,j=1}^N ((U - \bar{U})^* \odot \partial_{\bar{\alpha}} U)_{i,j}. \quad (2.5)$$

2.2.2 Вычисление градиента метрики

Разложение многомодового интерферометра (2.1) допускает простые аналитические формы градиентов по элементам u_{ij} базисных матриц U_ℓ и по фазовым сдвигам $\varphi_{\ell k}$. Выведем аналитические выражения градиентов $\partial_{x_{ij}^{(\ell)}} J_{FR}$, $\partial_{y_{ij}^{(\ell)}} J_{FR}$ и $\partial_{\varphi_{\ell k}} J_{FR}$, необходимых на этапах обучения и настройки алгоритма соответственно. Из уравнения (2.5) следует, что вычисление градиентов $\partial_{x_{ij}^{(\ell)}} J_{FR}$, $\partial_{y_{ij}^{(\ell)}} J_{FR}$ и $\partial_{\varphi_{\ell k}} J_{FR}$ сводится к нахождению выражений для $\partial_{x_{ij}^{(\ell)}} U$, $\partial_{y_{ij}^{(\ell)}} U$ и $\partial_{\varphi_{\ell k}} U$ соответственно.

Сначала мы рассмотрим градиенты $\partial_{x_{ij}^{(\ell)}} J_{FR}$ и $\partial_{y_{ij}^{(\ell)}} J_{FR}$. Для упрощения вычислений введем N вспомогательных матриц A_ℓ и еще N вспомогательных матриц B_ℓ как частичные произведения, взятые из уравнения разложения (2.1):

$$U = A_\ell U_\ell B_\ell, \quad (2.6)$$

где A_ℓ и B_ℓ можно вычислить итеративно:

$$\begin{cases} A_N = \Phi_{N+1}, \\ A_\ell = A_{\ell+1}(U_{\ell+1}\Phi_{\ell+1}), \quad \ell = N-1, \dots, 1, \\ B_1 = \Phi_1, \\ B_\ell = (\Phi_\ell U_{\ell-1})B_{\ell-1}, \quad \ell = 2, \dots, N. \end{cases} \quad (2.7)$$

Следовательно, имеем:

$$u_{ij} = \sum_{k=1}^N \sum_{m=1}^N a_{ik}^{(\ell)} u_{km}^{(\ell)} b_{mj}^{(\ell)}. \quad (2.8)$$

Вычислим производную:

$$\frac{\partial u_{ij}}{\partial u_{pt}^{(\ell)}} = \sum_{k=1}^N \sum_{m=1}^N a_{ik}^{(\ell)} \frac{\partial u_{km}^{(\ell)}}{\partial u_{pt}^{(\ell)}} b_{mj}^{(\ell)} = \sum_{k=1}^N \sum_{m=1}^N a_{ik}^{(\ell)} \delta_{kp} \delta_{mt} b_{mj}^{(\ell)} = a_{ip}^{(\ell)} b_{tj}^{(\ell)}. \quad (2.9)$$

Пусть унитарные элементы будут $u_{ij}^{(\ell)} = x_{ij}^{(\ell)} + iy_{ij}^{(\ell)}$, и мы хотим вычислить производные комплексных элементов u_{ij} матрицы U по действительной $x_{pt}^{(\ell)}$ и мнимой $y_{pt}^{(\ell)}$ частям комплексного числа u_{ij} :

$$\frac{\partial u_{ij}}{\partial x_{pt}^{(\ell)}} = \frac{\partial u_{ij}}{\partial u_{pt}^{(\ell)}} \frac{\partial u_{ij}^{(\ell)}}{\partial x_{pt}^{(\ell)}} = \frac{\partial u_{ij}}{\partial u_{pt}^{(\ell)}} = a_{ip}^{(\ell)} b_{tj}^{(\ell)}, \quad (2.10)$$

$$\frac{\partial u_{ij}}{\partial y_{pt}^{(\ell)}} = \frac{\partial u_{ij}}{\partial u_{pt}^{(\ell)}} \frac{\partial u_{ij}^{(\ell)}}{\partial y_{pt}^{(\ell)}} = i \frac{\partial u_{ij}}{\partial u_{pt}^{(\ell)}} = ia_{ip}^{(\ell)} b_{tj}^{(\ell)}. \quad (2.11)$$

Преобразуем полученные выражения в матричную форму. Используя вспомогательные матрицы $\Delta^{(mn)}$, элементы которых — нули, за исключением $\Delta_{mn}^{(mn)} = 1$, преобразуем формулу (2.9) к следующему виду:

$$\frac{\partial u_{ij}}{\partial u_{pt}^{(\ell)}} = a_{ip}^{(\ell)} b_{tj}^{(\ell)} = \sum_{k=1}^N \sum_{m=1}^N a_{ik}^{(\ell)} \Delta_{km}^{(pt)} b_{mj}^{(\ell)} \Rightarrow \frac{\partial U}{\partial u_{ij}^{(\ell)}} = A_\ell \Delta^{(ij)} B_\ell. \quad (2.12)$$

Наконец мы приходим к выражению:

$$\begin{aligned} \frac{\partial U}{\partial x_{ij}^{(\ell)}} &= A_\ell \Delta^{(ij)} B_\ell, \\ \frac{\partial U}{\partial y_{ij}^{(\ell)}} &= i A_\ell \Delta^{(ij)} B_\ell. \end{aligned} \quad (2.13)$$

После того, как базисные матрицы модели $\mathcal{M}$ определены, мы можем использовать их для вычисления градиентов $\partial_{\varphi_{\ell k}} J_{FR}$. Вывод $\partial_{\varphi_{\ell k}} J_{FR}$ также требует введения $N + 1$ вспомогательных матриц C_ℓ и $N + 1$ матриц D_ℓ :

$$U = C_\ell \Phi_\ell D_\ell. \quad (2.14)$$

Следовательно, имеем:

$$u_{ij} = \sum_{k=1}^N c_{ik}^{(\ell)} \left(\sum_{m=1}^N e^{i\varphi_{\ell k}} \delta_{km} d_{mj}^{(\ell)} \right) = \sum_{k=1}^N c_{ik}^{(\ell)} e^{i\varphi_{\ell k}} d_{kj}^{(\ell)}. \quad (2.15)$$

Вычислим производную:

$$\begin{aligned} \frac{\partial u_{ij}}{\partial \varphi_{pt}} &= \frac{\partial}{\partial \varphi_{pt}} \left(\sum_{k=1}^N c_{ik}^{(\ell)} e^{i\varphi_{\ell k}} d_{kj}^{(\ell)} \right) = \sum_{k=1}^N c_{ik}^{(\ell)} \left(\frac{\partial}{\partial \varphi_{pt}} e^{i\varphi_{\ell k}} \right) d_{kj}^{(\ell)} = \\ &= \sum_{k=1}^N c_{ik}^{(\ell)} i e^{i\varphi_{\ell k}} \frac{\partial \varphi_{\ell k}}{\partial \varphi_{pt}} d_{kj}^{(\ell)} = \sum_{k=1}^N i c_{ik}^{(\ell)} e^{i\varphi_{\ell k}} \delta_{\ell p} \delta_{kt} d_{kj}^{(\ell)} = i e^{i\varphi_{pt}} c_{it}^{(\ell)} d_{tj}^{(\ell)}. \end{aligned} \quad (2.16)$$

Получаем выражение для $\frac{\partial u_{ij}}{\partial \varphi_{\ell k}}$:

$$\frac{\partial u_{ij}}{\partial \varphi_{\ell k}} = i e^{i\varphi_{\ell k}} c_{ik}^{(\ell)} d_{kj}^{(\ell)}. \quad (2.17)$$

Теперь выразим (2.17) в матричной форме. Введем вспомогательные матрицы $\Delta^{(kk)}$, все элементы которых равны нулям, за исключением $\Delta_{kk}^{(kk)} = 1$. Тогда формула (2.17) преобразуется к виду:

$$\frac{\partial u_{ij}}{\partial \varphi_{\ell k}} = i e^{i\varphi_{\ell k}} c_{ik}^{(\ell)} d_{kj}^{(\ell)} = i e^{i\varphi_{\ell k}} \sum_{j=1}^N \sum_{m=1}^N c_{ij}^{(\ell)} \Delta_{jm}^{(kk)} d_{mj}^{(\ell)}. \quad (2.18)$$

Тогда имеем:

$$\frac{\partial U}{\partial \varphi_{\ell k}} = i e^{i\varphi_{\ell k}} C_\ell \Delta^{(kk)} D_\ell. \quad (2.19)$$

2.3 Численные симуляции

В этом разделе приводятся ключевые показатели эффективности алгоритма обучения модели интерферометра. Исследуются масштабируемые

свойства алгоритма в зависимости от размера обучающей выборки M и числа мод интерферометра N . Качество модели $\mathcal{M}$ оценивается на тестовой выборке, не входящей в обучающий набор $\mathcal{T}$.

Моделирование алгоритма обучения выполняется в несколько этапов. Сначала мы генерируем обучающий набор $(\bar{\Phi}^{(i)}, \bar{U}^{(i)})$, используя уравнение разложения многомодового интерферометра (2.1), при этом фазы выбираются случайным образом из равномерного распределения от 0 до 2π . Базисные матрицы U_ℓ выбираются случайным образом из равномерного распределения Хаара с использованием QR-разложения [40]. В реальной ситуации элементы $\mathcal{T}$ являются результатами $\bar{U}_{rec}^{(i)}$ алгоритмов унитарной реконструкции [5; 36], применяемых к реконфигурируемому интерферометру, запрограммированному с фазами $\bar{\Phi}^{(i)}$. Тонкости сбора подходящего обучающего набора экспериментальным путем обсуждаются в разделе 3.4.

Адекватность модели интерферометра определяется её способностью с высокой точностью воспроизводить унитарное преобразование реального устройства при заданной конфигурации фазовых сдвигов Φ . Проверка на тестовой выборке направлена на оценку качества такого воспроизведения на данных, не использованных при обучении модели $\mathcal{M}$.

Для этой цели формируется тестовый набор $(\hat{\Phi}^{(i)}, \hat{U}^{(i)})$, состоящий из случайно выбранных конфигураций фазовых сдвигов $\hat{\Phi}^{(i)}$ и соответствующих им унитарных матриц $\hat{U}^{(i)}$, полученных для реального устройства. Для каждого элемента тестовой выборки сравниваются матрица $\hat{U}^{(i)}$ и предсказание модели $U(\hat{\Phi}^{(i)})$.

Модель считается успешно прошедшей тестирование, если для соответствующих пар матриц выполняется условие на среднее по тестовому множеству функционала невязки $J(U(\hat{\Phi}^{(i)}), \hat{U}^{(i)}) \leq 10^{-2}$. Данное пороговое значение было выбрано эмпирически на основе анализа поведения функционала невязки J_{FR} на тестовой выборке: для адекватно обученной модели значения J_{FR} быстро уменьшаются и достигают уровня ниже 10^{-2} .

Модель инициализируется базисными матрицами, выбранными либо случайным образом, либо на основе априорных знаний, доступных на основе конструкции физических элементов, реализующих базисные матрицы. Мы рассмотрим оба случая и будем называть случайную инициализацию моделью *чёрного ящика*. В каждой эпохе процесса обучения мы оцениваем средний градиент по набору примеров из $\mathcal{T}$ и обновляем базисные матри-

цы в соответствии с алгоритмом оптимизации (стохастический алгоритм L-BFGS-B, пакет SciPy). Вместо усреднения градиента по всему обучающему набору мы случайным образом выбираем подмножество из $m = 5$ пар $(\bar{\Phi}, \bar{U})$ в каждой эпохе и используем это подмножество для усреднения. Значение $m = 5$ было определено эмпирически, поскольку оно обеспечивало существенное ускорение вычислений при сохранении высокой точности обучения. Мы не используем унитарную параметризацию базисных матриц во время процедуры обучения, и эти матрицы просто являются комплекснозначными квадратными матрицами. Поскольку параметрами модели тогда являются действительная $x_{ij}^{(\ell)}$ и мнимая $y_{ij}^{(\ell)}$ части каждой базисной матрицы U_ℓ , обновлённые базисные матрицы не принадлежат унитарному пространству. Мы используем полярное разложение $A = HV$, где H — эрмитова, а V — унитарная матрица, для проецирования обновлённой комплексной базисной матрицы C_ℓ на ближайшую унитарную U_ℓ на каждом шаге оптимизации [41]. Этот метод помогает избежать проблемы локального минимума, которая может возникнуть из-за сложной параметризации унитарной матрицы.

Код симуляции написан на Python с использованием пакетов Numpy и Scipy. Он доступен публично на GitLab [42].

2.3.1 Модель — чёрный ящик

Начнём с рассмотрения модели *чёрного ящика*. Этот сценарий подразумевает отсутствие априорной информации о базисных матрицах U_ℓ , а интерферометр представлен системой типа *чёрный ящик* с N^2 переменными фазовыми параметрами $\varphi_{\ell k}$. Следовательно, модель должна быть инициализирована полностью случайным приближением. Начальные базисные матрицы выбираются из унитарного случайного распределения Хаара [40]. На рисунке 2.3 показана сходимость среднего значения функционала J_{FR} в процессе обучения при различных значениях размерности интерферометра N . Перекрёстная проверка модели (см. рисунок 2.4) определяет размер обучающей выборки M для каждого значения N .

Следует отметить, что с ростом N в процессе обучения возникает плато, существенно замедляющее сходимость алгоритма. Плато становится

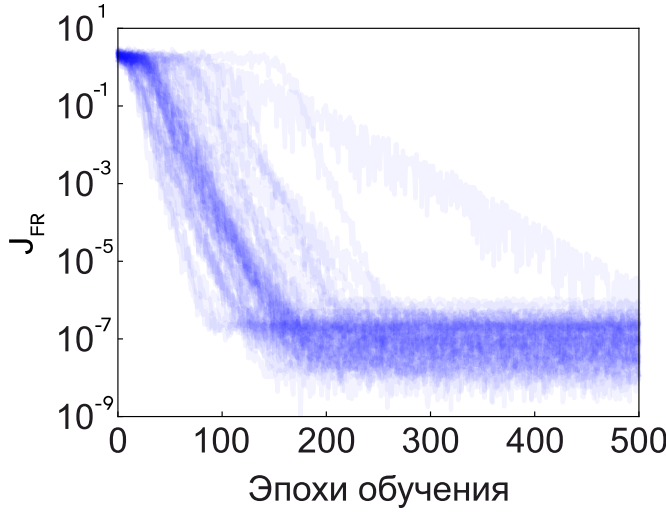
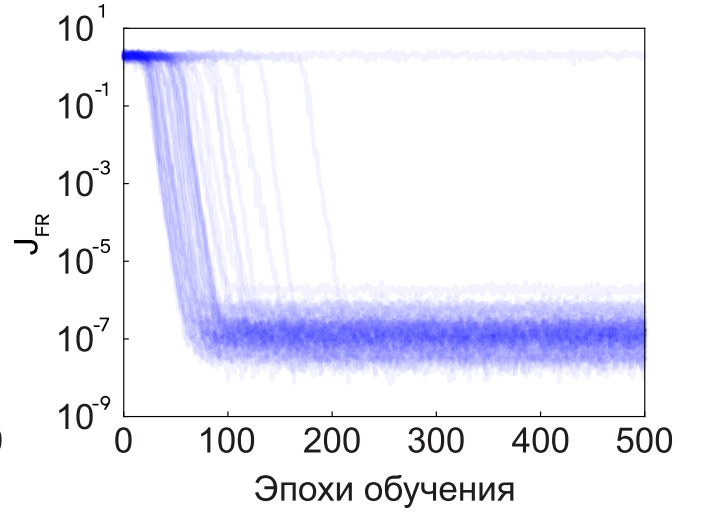
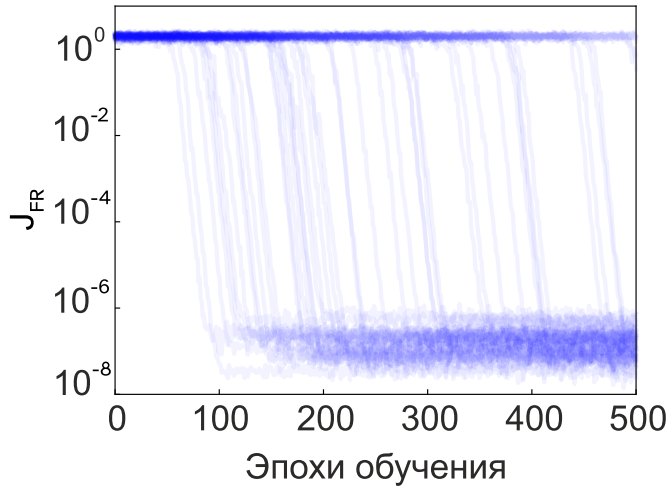
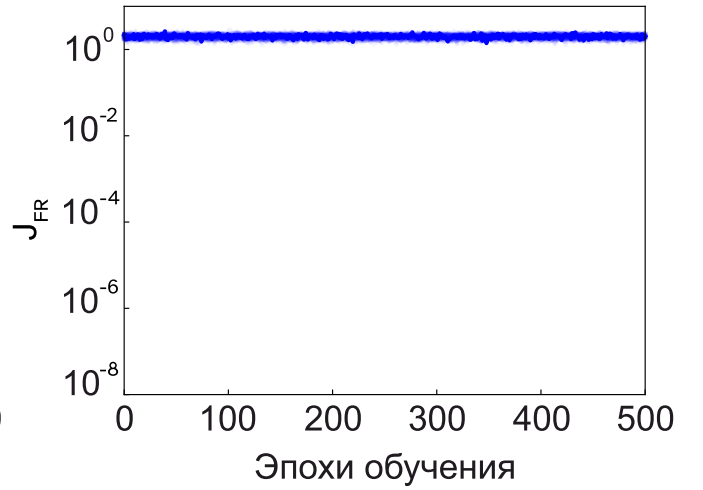
а) $N = 4$.б) $N = 5$.в) $N = 6$.г) $N = 7$.

Рисунок 2.3 — Примеры обучения для различного количества оптических мод N . Показаны графики сходимости функционала Фробениуса на тестовом наборе данных в зависимости от количества эпох для 50 различных экземпляров алгоритма оптимизации (для каждого экземпляра мы случайным образом выбирали начальные базисные матрицы U_ℓ). Случай а) — количество оптических мод $N = 4$, размер обучающего набора данных $M = 7$, б) — $N = 5$, $M = 30$, в) — $N = 6$, $M = 230$, г) — $N = 7$, $M = 2000$.

значительным уже при $N = 6$ (рисунок 2.3в). При $N > 6$ нам не удалось наблюдать обучение для модели чёрного ящика — среднее значение показателя качества остается на плато все время. Работа [43] предполагает, что причиной плато в задачах оптимизации высокой размерности является наличие большого количества седловых точек на ландшафте оптимизированной функции, а не локальных минимумов. Было разработано несколько

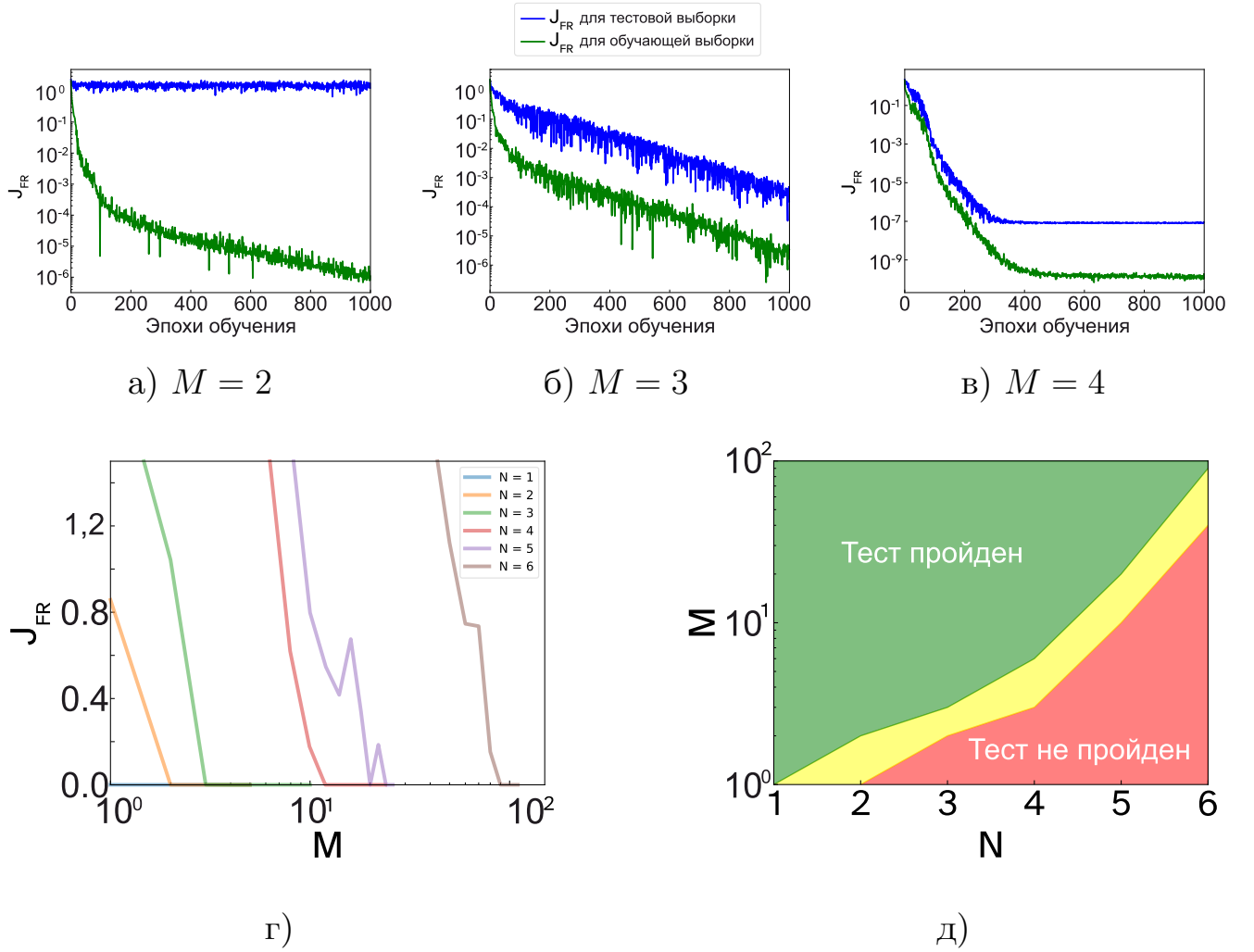


Рисунок 2.4 — Верхний ряд иллюстрирует поведение функционала невязки на обучающей и тестовой выборке для интерферометра с $N = 3$ при обучении на наборе $\mathcal{T}$ разного размера M . Случай а) — модель не проходит тестовую проверку, б) — промежуточный случай, в) — модель проходит тестовую проверку. На панели г) в нижнем ряду показана зависимость оптимизированного J_{FR} , усредненного по 10 случаям, для разного количества мод N и разных размеров обучающего набора данных M . Изображение на панели д) отображает область, где модель проходит критерии проверки на тестовой выборке.

алгоритмов, использующих идею адаптивного градиента, для решения проблемы выхода из плато [44; 45].

До этого момента элементы $\mathcal{T}$ включали матрицы $\bar{U}^{(i)}$, которые были искусственно сгенерированы с помощью уравнения (2.1), инициализированного набором фаз $\bar{\Phi}^{(i)}$. Сбор того же набора с помощью реального устройства означает, что реконструкция матриц $\bar{U}_{exp}^{(i)}$ должна быть выполнена с абсолютной точностью, что никогда не достигается в эксперименте.

Алгоритм обучения должен быть спроектирован так, чтобы допускать определённое расхождение между идеальными $\bar{U}_{ideal}$ и реконструированными матрицами $\bar{U}_{exp}$. Эти отклонения являются неизбежным следствием несовершенства измерительных приборов, используемых в процессе реконструкции. Мы смоделировали поведение алгоритма обучения с использованием обучающего набора, включающего фазовые сдвиги $\bar{\Phi}^{(i)}$ и унитарные значения $\bar{U}_{exp}^{(i)}$, немного отклоняющиеся от их теоретического аналога $\bar{U}_{ideal}^{(i)}$.

Отклонение вводится между $\bar{U}_{exp}$ и $\bar{U}_{ideal}$ как полярная проекция [41] возмущённого $\bar{U}_{ideal}$ на унитарное пространство:

$$A = \bar{U}_{ideal} + \alpha(X + iY), \quad A = H\bar{U}_{exp}, \quad (2.20)$$

где X и Y — случайные вещественные матрицы размера $N \times N$, элементы которых выбираются из нормального распределения $\mathcal{N}(0, 1)$. Степень отклонения контролируется вещественным параметром α . Калибровочные кривые, представленные на рисунке 2.5а, сопоставляют более распространённую меру матричного расстояния — фиделити F , вычисляемую как скалярное произведение Гильберта-Шмидта, с величинами отклонения $F(\alpha)$. Эти кривые должны развить интуицию для интерпретации значений J_{FR} . Рисунок 2.5 иллюстрирует сходимость модели для простейшего случая $N = 2$, обучаемой на наборе данных, выбранным с заданным отклонением α . $\alpha \approx 0.04$ указывает порог, при котором модель не проходит критерий перекрестной проверки $J_{FR} \leq 10^{-2}$.

2.3.2 Модель с априорным знанием

Результаты модели чёрного ящика, изложенные в разделе 2.3.1, свидетельствуют о том, что сложность оптимизации модели с произвольными начальными базисными матрицами быстро растёт при объёме обучающей выборки M . В данном разделе мы исследуем выбор начального приближения для базисных матриц U_ℓ , что позволяет проводить обучение для большей размерности N . Случай, когда базисные матрицы U_ℓ полностью неизвестны, неадекватно отражает реальную ситуацию. На практике

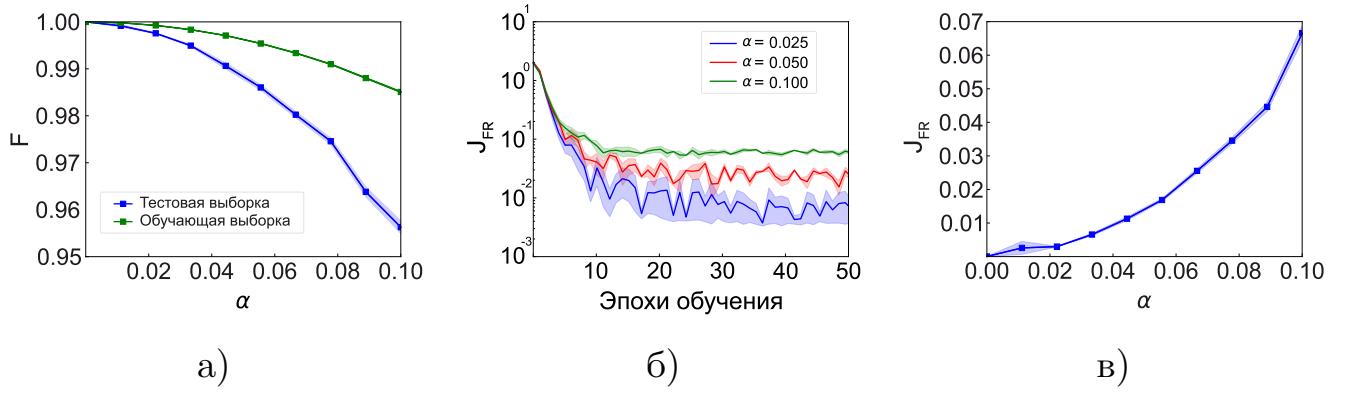
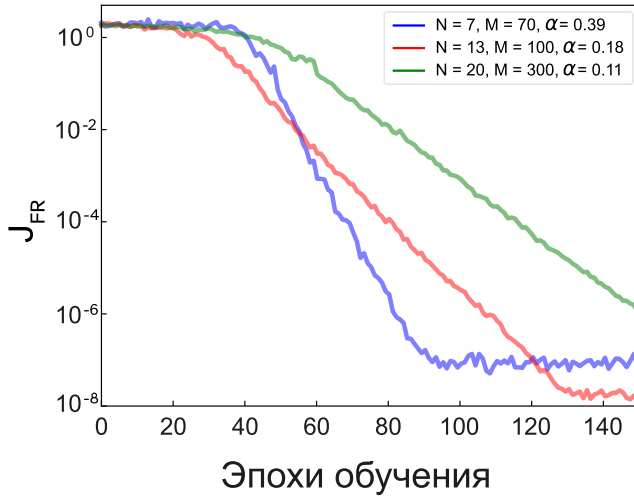


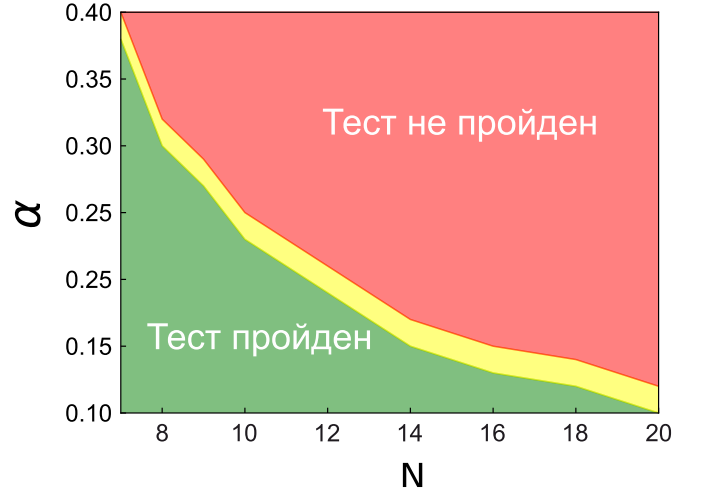
Рисунок 2.5 — Влияние несовершенства обучающего набора на обучение модели интерферометра. а) — Зависимость среднего значения фиделити (усреднение по 1000 унитарным матрицам), с которой обученная модель может производить унитарные матрицы для заданных фаз (синий), и среднего значения фиделити (усреднение по 10000 унитарным матрицам), с которой известны матрицы из обучающего набора (зелёный), от параметра α для матрицы размера $N = 2$. б) — Графики сходимости для моделей интерферометра $N = 2$, снабжённых обучающим набором $M = 5$, для которых $\alpha = 0.025, 0.05, 0.1$. в) — Среднее значение функционала Фробениуса J_{FR} , вычисленное в ходе проверки на тестовой группе. Усреднение проводилось по 1000 моделям, обученным на обучающих наборах, соответствующих различным базисным матрицам $U_\ell^{(0)}$. Для каждой модели был выполнен тест с 1000 наборами фазовых сдвигов.

базисные матрицы реализуются с помощью оптических схем с чётко определённой геометрией и оптическими свойствами. Прототипы этих схем могут быть предварительно протестированы для проверки их производительности, включая генерируемое ими преобразование мод. Современные технологии производства интегральных фотонных устройств гарантируют воспроизводимость до определённого уровня точности, специфичного для каждой технологической линии. Базисная матрица преобразования U_ℓ^{est} может быть оценена заранее. Эта оценка служит начальным приближением для алгоритма оптимизации на этапе обучения модели. В этом разделе мы покажем, как эти знания существенно упрощают оптимизацию и позволяют изучать модели интерферометров с N по крайней мере до нескольких десятков мод.

Мы используем оценённые матрицы U_ℓ^{est} в качестве начального приближения для нашей процедуры оптимизации. Эти матрицы выбираются



а)



б)

Рисунок 2.6 — Исследование процесса обучения с использованием априорных знаний о базисных матрицах U_ℓ . а) — Примеры обучения с использованием априорных знаний о базисных матрицах при больших значениях числа оптических мод N интерферометра, б) — Зависимость максимального требуемого значения α , при котором модель может быть успешно обучена, от числа оптических мод N .

так, чтобы быть близкими к идеальным базисным матрицам U_ℓ , используемым для генерации обучающего набора. Мы получаем начальное приближение, используя процедуру, описанную в уравнении (2.20). На рисунке 2.6 показана сходимость алгоритма обучения, использующего знание унитарной базисной матрицы с определённой точностью. Априорная информация о базисных матрицах позволила обучить модель интерферометра до $N = 20$, используя то же вычислительное оборудование. Чем больше размерность интерферометра N , тем выше должна быть точность оценки U_ℓ^{est} . На рисунке 2.6б показаны области значения α в зависимости от размера интерферометра, где модель $\mathcal{M}$ может успешно обучаться.

2.4 Эксперимент

2.4.1 Дизайн эксперимента

Мы демонстрируем наш программный алгоритм на фотонном чипе, изготовленном с использованием фемтосекундной лазерной записи (FSLW) в образце из плавленого кварцевого стекла [46]. Одномодовая волноводная структура чипа содержала целевой многопортовый интерферометр 4×4 и подготовительный интерферометр. Подготовительный интерферометр представляет собой сетку из трёх интерферометров Маха-Цендера, которая позволяет задавать произвольное распределение амплитуды на выходе подготовительной части. Мы используем подготовительный интерферометр в качестве инструмента для калибровки целевого интерферометра и исследования его матрицы преобразования. Как подготовительный, так и целевой интерферометры оснащены перестраиваемыми термооптическими фазовыми сдвигами, которые также изготовлены с использованием установки фемтосекундной лазерной записи.

Технология FSLW позволяет создавать волноводы внутри трёхмерного пространства. Эта особенность позволяет нам записывать четыре вспомогательных волновода на глубину 35 мкм ниже плоскости основного волновода. Эти волноводы эванесцентно связаны с выходной секцией каждого волновода подготовительного интерферометра. Мы использовали эти волноводы для мониторинга амплитуд, подаваемых на вход целевого интерферометра, что является важнейшим требованием для восстановления цифровой модели целевого интерферометра.

Свет с выходных портов чипов собирался с помощью волоконно-оптического линейного массива, а соответствующая оптическая мощность измерялась с помощью четырех фотодетекторов. Входное волокно и выходная волоконно-оптическая матрица были установлены на высокоточных шестиосевых позиционерах. С помощью выходного позиционера можно было переключаться между выходом целевого интерферометра и вспомогательными волноводами. Структура чипа и экспериментальная установка схематически показаны на рисунке 2.7.

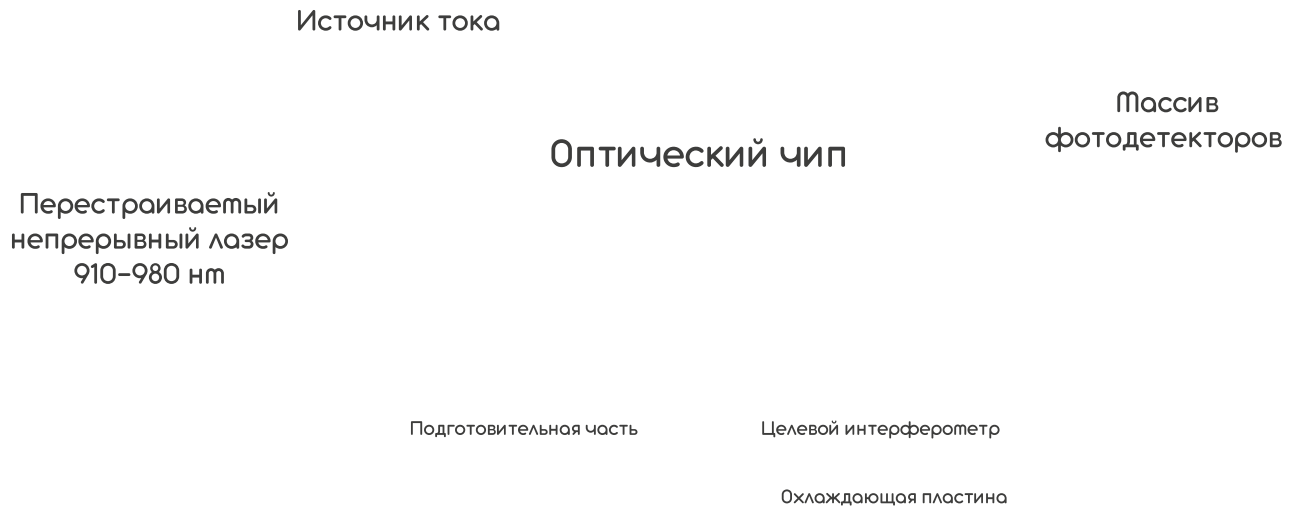


Рисунок 2.7 — Схема структуры фотонного чипа. Подготовительная часть, состоящая из трех каскадно соединенных интерферометров Маха-Цендера (MZI), и целевой комплексный интерферометр, состоящий из двух многопортовых соединителей 4×4 , между которыми расположены три перестраиваемых фазовращателя, находятся в одном и том же образце из плавленого кварца. Лазерный свет подается на второй входной порт чипа. Это позволяет подготовить произвольное оптическое распределение света между четырьмя входными портами целевого интерферометра. Интерферометры Маха-Цендера в интерферометре могут быть откалиброваны индивидуально с помощью вспомогательных волноводов, слабо связанных с выходной секцией подготовительной части интерферометра. Вспомогательные волноводы расположены на 35 мкм ниже плоскости основного волновода с подготовительным и целевым интерферометрами. Каждый вспомогательный волновод связан только с соответствующим волноводом подготовительного интерферометра и только в его выходной области.

Калибровка интерферометра на основе интерферометров Маха-Цендера — известная процедура, которая выполняется путём индивидуальной калибровки каждого из составляющих его интерферометров Маха-Цендера [9; 47–49]. Мы использовали те же принципы для калибровки подготовительного интерферометра.

Целевой интерферометр состоит из двух многопортовых соединителей размером 4×4 , между которыми расположены три перестраиваемых фазовращателя. Оптические интерферометры с такой архитектурой были исследованы в работе [50]. Унитарное преобразование целевого интерферо-

метра можно выразить как

$$U(\vec{\varphi}) = M_2 \times P(\vec{\varphi}) \times M_1, \quad (2.21)$$

где $M_{1,2}$ — унитарные матрицы, соответствующие каждому из 4×4 многопортовых соединителей, которые мы также называем базисными матрицами, $\vec{\varphi} = \{\varphi_1, \varphi_2, \varphi_3\}$ — вектор из трех фазовых сдвигов, а $P(\vec{\varphi}) = \text{diag}(e^{i\varphi_1}, e^{i\varphi_2}, e^{i\varphi_3}, 1)$ — диагональная матрица слоя фазового сдвига. Ключевой задачей для успешной реализации процедуры программирования является восстановление двух матриц $M_{1,2}$ и $P(\vec{\varphi}(\vec{X}))$, где $\vec{X} = \{x_1, x_2, x_3\}$ — электрические токи, подаваемые на термооптические фазовращатели.

2.4.2 Калибровка целевого интерферометра

Калибровка перестраиваемых фазовращателей в N -портовом реконфигурируемом интерферометре начинается с ввода когерентного лазерного света в образец. Для каждого входного порта электрический ток x_j через j -й фазовращатель изменяется в диапазоне значений, при этом измеряется оптическая мощность P_m с каждого выходного порта m интерферометра $P_m = P_m(x_j)$.

Идеальный сценарий возникает, когда j -й модулятор воздействует только на соответствующий сдвиг на интерферометре $\varphi_j = \varphi(x_j)$. В этом случае выходная мощность P_m подчиняется периодическому закону

$$P_m(x_j) = A_m + B_m \cos \varphi_j(x_j), \quad (2.22)$$

где A_m и B_m — действительные константы (не зависящие от x_j), зависящие от внутренней структуры интерферометра, в нашем случае — матрицы $M_{1,2}$, описывающие многопортовые разветвители. $\varphi = \varphi_j(x_j)$ — это специфическая зависимость j -го фазового сдвига от соответствующего управляющего сигнала x_j , и обычно предполагается, что она является полиномиальной функцией от x_j . Например, в случае термооптического фазового сдвигателя $\varphi_j(x_j) = \varphi_j^{(0)} + \alpha_j x_j^2$, поскольку индуцированный фазовый сдвиг прямо пропорционален тепловой мощности, рассеиваемой резистивным электродом.

После измерения $P_m(x_j)$, аппроксимация экспериментальных данных с помощью (2.22) позволяет получить калибровочные параметры φ_j^0 и α_j для конкретного перестраиваемого фазовращателя. Следует отметить, что конкретные значения констант A_m и B_m в этой процедуре не важны, а наиболее важными параметрами для каждого фазовращателя являются соответствующие φ_j^0 и α_j . После получения всех калибровочных параметров φ_j^0 и α_j для каждого j -го перестраиваемого фазовращателя, можно использовать алгоритмы из раздела 2.2 или [51] для восстановления $M_{1,2}$. После определения $M_{1,2}$ становится доступна полная цифровая модель фотонного чипа для нахождения фазовой конфигурации, соответствующей требуемой унитарной матрице переноса.

Однако в большинстве реальных оптических чипов, даже изготовленных с использованием самых передовых технологий [48], часто наблюдается эффект взаимного перекрестного влияния между различными фазовыми сдвигами. Это может происходить из-за тепловых или электрических перекрестных помех. Из-за эффекта перекрестных помех оптическая мощность P_m m -го выходного порта чипа больше не может быть описана уравнением (2.22), поскольку аппроксимационные константы A_m и B_m являются функциями x_j , а также индексов входного и выходного портов интерферометра. Это приводит к сложной и непериодической зависимости $P_m(x_j)$.

В нашем оптическом чипе мы использовали термооптические фазовращатели, для которых наблюдался значительный эффект перекрестных помех между соседними нагревателями. Из-за наличия взаимной связи между различными фазовращателями фазовый сдвиг φ_i , индуцированный в i -м нагревателе, зависит не только от соответствующего тока x_i , но и от всех токов x_j , подаваемых на другие нагреватели (которые могут влиять на i -й фазовый сдвиг):

$$\varphi_i(\vec{x}) = \varphi_j^{(0)} + \sum_j \alpha_{ij} x_j^2, \quad (2.23)$$

где x_j — ток, подаваемый на j -й нагреватель, $\varphi_j^{(0)}$ — постоянный фазовый сдвиг смещения, существующий в конкретном плече интерферометра, а α_{ij} — константа, определяющая силу воздействия j -го нагревателя на i -й фазовращатель. Соотношение между фазами и токами, протекающими через нагреватели, можно переписать в матричной форме следующим

образом:

$$\vec{\Phi}(\vec{x}) = \vec{\Phi}_0 + A \times \vec{X}^2, \quad (2.24)$$

где $\vec{\Phi} = \{\varphi_1, \varphi_2, \varphi_3\}^T$ — вектор-столбец фазовых сдвигов, которые необходимо установить, $\vec{\Phi}_0 = \{\varphi_1^{(0)}, \varphi_2^{(0)}, \varphi_3^{(0)}\}^T$ — вектор-столбец фазовых сдвигов смещения, $A = \{\alpha_{ij}\}$ — матрица перекрестных помех, содержащая все девять констант перекрестных помех α_{ij} , $i, j = \overline{1,3}$, и $\vec{X}^2 = \{x_1^2, x_2^2, x_3^2\}^T$ — вектор-столбец квадратов значений тока.

Следовательно, согласно (2.23), если изменяется только один параметр x_j , это одновременно повлияет на все три фазовых сдвига φ_i , что приведет к сложной и непериодической зависимости выходной оптической мощности от приложенного тока x_j . Тем не менее, если когерентный свет подается на k -й входной порт фотонного чипа, выходная оптическая мощность на m -м выходном порту может быть описана через m, k -й элемент унитарной матрицы интерферометра:

$$P_{km}(x_j) \sim |M_2 \times \Phi(x_j) \times M_1|_{mk}^2, \quad (2.25)$$

который можно использовать для подгонки соответствующих экспериментальных данных, используя в качестве параметров оптимизации как M_1 , M_2 , так и $\vec{\Phi}_0$, а также матрицу констант перекрестных помех A .

Более того, любая экспериментально измеренная оптическая мощность P_{km} от приложенного тока для любого входного k и выходного m портов интерферометра и для любого нагревателя j должна удовлетворять уравнению (2.25) с теми же параметрами интерферометра $M_{1,2}$, $\vec{\Phi}_0$ и A , поскольку все они генерируются одним и тем же фотонным чипом, описываемым его унитарной матрицей U . После того, как все возможные оптические мощности от зависимостей приложенного тока, $P_{km}(x_j)$, будут экспериментально измерены для каждого k , m и j , их можно одновременно аппроксимировать с помощью единой модели интерферометра, определяющей составляющие оптических чипов $M_{1,2}$ и необходимые калибровочные соотношения между всеми токами и фазовыми сдвигами $\vec{\Phi}_0$ и A .

2.4.3 Целевая модель интерферометра

Аппроксимация всего набора данных $P_{km}(x_j)$ в форме (2.25) требует соответствующей параметризации матрицы передачи оптического чипа $U(\vec{x})$ как функции приложенных токов:

$$U(\vec{x}, \vec{t}) = M_2(\vec{t}_2) \times P(\vec{\varphi}(\vec{x}, \vec{\varphi}_0, A)) \times M_1(\vec{t}_1), \quad (2.26)$$

где $\vec{t}_{1,2} = \{t_1, \dots, t_N\}$ — два набора действительных параметров, определяющих $M_{1,2}$, $P(\vec{\varphi}(\vec{x}, \vec{\varphi}^{(0)}, A)) = \text{diag}(e^{i\varphi_1}, e^{i\varphi_2}, e^{i\varphi_3}, 1)$ — диагональная матрица фазовых сдвигов. В свою очередь, фазовые сдвиги зависят от токов x_j и параметров калибровки $\varphi_j^{(0)}$ и α_{ij} :

$$\begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{pmatrix} = \begin{pmatrix} \varphi_1^{(0)} \\ \varphi_2^{(0)} \\ \varphi_3^{(0)} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \times \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \end{pmatrix}, \quad (2.27)$$

где два последних будут определены экспериментально.

Для блоков смешивания мод $M_{1,2}$ мы использовали хорошо известную унитарную параметризацию [1], которая для интерферометра с 4×4 портами требует девять действительных параметров, изменяющихся от 0 до 2π каждый. Общее число действительных параметров для нашего фотонного чипа составляет 30, включая $9 \times 2 = 18$ параметров для $M_{1,2}$, 3 фаз смещения $\varphi_j^{(0)}$ и 9 констант перекрестных помех α_{ij} .

Адекватное начальное приближение для параметров оптимизации имеет важное значение, поскольку глобальный процесс оптимизации испытывает трудности со сходимостью к глобальному минимуму при использовании случайного начального значения. При установлении предельных значений параметров следует учитывать специфические физические свойства экспериментальной системы.

2.4.4 Детали обучения и результаты эксперимента

Мы провели экспериментальную проверку концепции по реконструкции цифровой модели интерферометра и выполнили программирование

матрицы переноса в соответствии с разделом 2.2, основанной на различных наборах данных.

Этот метод использует обучающие и тестовые наборы данных, состоящие из предварительно реконструированных унитарных матриц, соответствующих значениям управляющих сигналов $\vec{x}$ (в нашем случае токам через нагреватели) или значениям фазового сдвига $\vec{\varphi}$, вызванным управляющими сигналами. Мы использовали калибровку $\varphi(x)$ и унитарные матрицы и подготовили два набора данных. Первый набор данных $(\vec{\varphi}, U)$ включает фазовые сдвиги $\vec{\varphi}$ и соответствующие матрицы U , а второй набор данных $(\vec{x}, U)$ включает значения тока $\vec{x}$ и соответствующие матрицы U . Затем мы использовали оба набора данных для оценки эффективности метода и сравнения ее с результатами, полученными в результате глобальной оптимизации.

Датасет «Фазы — Матрицы преобразования». Набор данных содержал 27 пар унитарных матриц и соответствующие фазовые сдвиги $\vec{\varphi}$, полученные в результате экспериментальной калибровки, описанной в разделе 2.4.3. Это означает, что в этом наборе данных нам известны фазовые сдвиги $\vec{\varphi}$, установленные на чипе, и соответствующее измеренное унитарное преобразование $U_{\text{data}}(\vec{\varphi})$ чипа.

Обученная модель была следующей:

$$U_{\text{trained}}(\vec{p}) = f_{\text{rb}}(A_2(\vec{p}_2) \times P(\vec{\varphi}) \times A_1(\vec{p}_1)), \quad (2.28)$$

где $f_{\text{rb}}(\cdot)$ — функция над унитарной матрицей, которая преобразует её в унитарную матрицу с вещественной первой строкой и вещественным первым столбцом путём умножения на диагональные фазовые матрицы слева и справа, $A_{1,2}$ — две произвольные комплексные матрицы (не обязательно унитарные), а $P(\vec{\varphi}) = \text{diag}(e^{i\varphi_1}, e^{i\varphi_2}, e^{i\varphi_3}, 1)$ — диагональная матрица фазового сдвига.

Матрицы $A_{1,2}$ рассматривались как общие комплексные матрицы размером 4×4 и параметризовались $2 \times 4^2 = 32$ действительными параметрами $\vec{p}$ каждая, представляющими параметризацию действительной и мнимой частей $A_{1,2}$. Начальные значения $\vec{p}$ выбирались случайным образом для каждого запуска оптимизации.

Модель (2.28) была обучена на 27 унитарных матрицах, для которых соответствие между ними и калибровочной моделью превышает 0.95, и протестирована на остальных 65 измеренных унитарных матрицах.

Мы использовали алгоритм оптимизации Adaptive Moment Estimation (Adam) с 5000 эпохами, каждая из которых состояла из мини-пакетов, содержащих по 5 случайных элементов. В качестве функции потерь была выбрана метрика Фробениуса, как в работе [52].

Результаты представлены на рисунке 2.8а. Как видно из результатов, фиделити между измеренными унитарными матрицами и матрицами, предсказанными моделью машинного обучения, составляет не менее 0.90 с пиком около 0.99, что указывает на точность представленной модели оптического чипа на основе машинного обучения. Однако набор данных матриц фазовых сдвигов требует точного знания информации о фазовых сдвигах, которая получается в результате процедуры калибровки, описанной в разделе 2.4.3, обеспечивающей нахождение самих элементов матриц смешивания мод $M_{1,2}$. Поэтому выполнение еще одной процедуры оптимизации для определения матриц $A_{1,2}$ при уже имеющихся матрицах $M_{1,2}$ может показаться несколько избыточным. В этом отношении существует высокая потребность в более фундаментальном наборе данных токов и соответствующих унитарных матриц для преобразования мод.

Датасет «Токи — Матрицы преобразования». Набор данных содержал 60 пар унитарных матриц и соответствующие им токи, заданные на чипе. Этот сценарий имеет большее практическое значение, чем набор данных $\vec{\varphi} — U$, поскольку он содержит только исходные данные без какой-либо предварительной обработки, например, процедуры калибровки.

Модель линейно-оптического интерферометра была обучена на 60 случайных унитарных матрицах из измеренных матриц на оптическом чипе и протестирована на остальных 36 измеренных унитарных матрицах. В отличие от набора данных $\vec{\varphi} — U$ в наборе данных $\vec{x} — U$, мы предполагаем, что отсутствует информация о калибровке $\varphi(x)$, и поэтому не можем выбрать 60 унитарных матриц с точностью калибровочной модели более 0.95, как в предыдущем подразделе.

Обученная модель была следующей:

$$U_{\text{trained}}(\vec{p}) = f_{\text{rb}}(A_2(\vec{p}_2) \times P(\vec{x}) \times A_1(\vec{p}_1)), \quad (2.29)$$

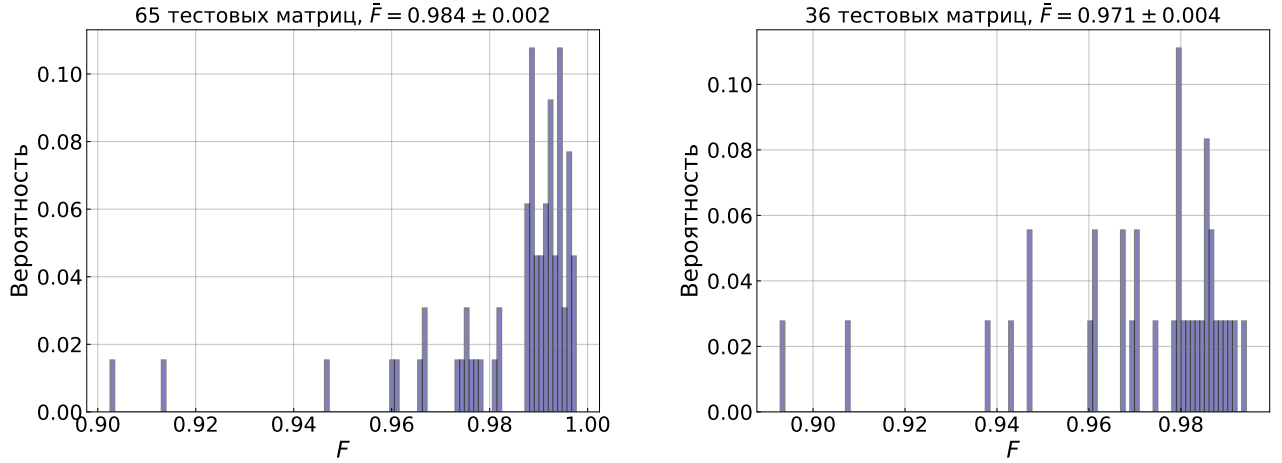
а) Датасет $\vec{\varphi} \rightarrow U$ б) Датасет $\vec{x} \rightarrow U$

Рисунок 2.8 — Гистограммы распределения фиделити для тестирования модели на основе машинного обучения. а) Тестирование обучения модели на наборе данных фазовых сдвигов и соответствующих матриц ($\vec{\varphi} - U$). б) Тестирование обучения модели на наборе данных токов и соответствующих матриц ($\vec{x} - U$).

где $f_{\text{rb}}(\cdot)$ — та же самая функция приведения к форме с вещественными первой строкой и первым столбцом, используемая в уравнении (2.28), $A_{1,2}$ — две произвольные комплексные матрицы (не обязательно унитарные), а $P(\vec{x}) = \text{diag}(e^{i\varphi_1(\vec{x})}, e^{i\varphi_2(\vec{x})}, e^{i\varphi_3(\vec{x})}, 1)$ — диагональная матрица фазового сдвига с фазовыми сдвигами, обусловленными зависимостью от тока (2.24), в результате чего получается 12 вещественных параметров: 3 фазовых сдвига смещения $\vec{\varphi}_0$ и 9 элементов матрицы перекрестных помех A_{ij} . Матрицы $A_{1,2}$ снова рассматривались как общие комплексные матрицы 4×4 и параметризовались 32 действительными параметрами. Начальные значения для $\vec{r}$ выбирались случайным образом для каждого запуска оптимизации, в то время как начальное предположение для элементов матрицы перекрестных помех $\{\alpha_{ij}\}$ бралось в соответствии с теоретической оценкой элементов этой матрицы исходя из дизайна интерферометра.

Для обучения сети использовался алгоритм оптимизации Adaptive Moment Estimation с 6000 эпохами, каждая из которых состояла из мини-пакетов, содержащих по 5 случайных элементов. В качестве функции потерь была выбрана метрика Фробениуса, как в работе [52].

Результаты представлены на рисунке 2.8б. Как видно из результатов, фиделити между измеренными унитарными матрицами и матрицами, пред-

сказанными моделью машинного обучения, составляет не менее 0.90 с пиком около 0.99, что указывает на точность представленной модели оптического чипа на основе машинного обучения.

Стоит отметить, что в этом эксперименте обучение модели линейно-оптического интерферометра на наборе данных матриц токов ($\vec{x} - U$) для определения модели оптического чипа на основе машинного обучения направлено на решение той же задачи, что и калибровка слоя фазового сдвига. Оба метода возвращают цифровую модель чипа как функцию экспериментально настраиваемых параметров: электрических токов $\vec{x}$. Эта модель позволяет предсказывать унитарное преобразование чипа при заданном токе. Однако, несмотря на общую цель, эти два метода различаются используемыми экспериментальными данными. Метод калибровки требует всей выходной мощности из зависимостей тока, тогда как метод машинного обучения требует набора унитарных матриц, измеренных на чипе. В реалистичных экспериментальных условиях не всегда легко выполнить полную томографию оптического чипа, в отличие от простых измерений калибровки выходной мощности. По крайней мере, потому что, если мы используем одни и те же экспериментальные ресурсы (когерентное лазерное излучение), полная томография чипа потребует одновременного ввода излучения в два порта чипа и точного изменения их относительного фазового сдвига. Однако при калибровочных измерениях использовался только один входной порт. Затем оба метода используют оптимизацию по набору вещественных параметров для минимизации соответствующей функции потерь.

2.5 Выводы по главе

В данной главе предложен архитектурно-независимый подход к обучению цифровой модели реконфигурируемого линейно-оптического интерферометра по выборке вида «конфигурация фазовых сдвигов — реализуемое преобразование». Такой подход позволяет перенести процедуру программирования устройства на его численную модель, восстановленную по экспериментальным данным, и тем самым сократить число обращений к реальному интерферометру в процессе настройки.

Показано, что задача обучения модели сводится к задаче многопараметрической оптимизации по базисным матрицам, описывающим фиксированные модосмешивающие элементы схемы. Для выбранного функционала невязки получены аналитические выражения для градиентов как по элементам этих матриц, так и по управляемым фазовым сдвигам, что обеспечивает возможность эффективного применения градиентных методов как при обучении модели, так и при последующем программировании устройства.

Численное моделирование показало, что в сценарии *чёрного ящика*, когда априорная информация о структуре интерферометра отсутствует, сложность обучения быстро возрастает с увеличением числа мод. При этом использование даже приближённой информации о базисных матрицах существенно улучшает сходимость и делает возможным обучение моделей большей размерности. Также установлено, что алгоритм сохраняет работоспособность при умеренных ошибках в обучающей выборке, связанных с экспериментальной реконструкцией унитарных матриц.

Экспериментальная апробация метода выполнена на фотонном чипе 4×4 , изготовленном по технологии фемтосекундной лазерной записи. Показано, что как в случае обучения по данным с известными фазами $(\vec{\phi}, U)$, так и в случае обучения по данным с известными токами $(\vec{x}, U)$, где $\vec{x}$ задаёт непосредственно управляющие электрические токи, предложенный подход обеспечивает высокую точность предсказания реализуемых унитарных преобразований.

Таким образом, разработанный метод является универсальным средством программирования линейно-оптических интерферометров. Он позволяет строить цифровые модели реальных устройств, пригодные не только для настройки, но и для анализа их фактических ограничений, обусловленных технологическими отклонениями и особенностями физической реализации.

Результаты данной главы опубликованы в работах [52] и [53].

Глава 3. Переход в новое признаковое пространство фаз для управления линейным оптическим интерферометром

В предыдущей главе задача программирования линейного оптического интерферометра рассматривалась как задача обучения модели, отображающей конфигурацию фазовых сдвигов в соответствующее преобразование устройства. Однако при увеличении размерности системы и глубины схемы такое представление приводит к нелинейной задаче оптимизации с большим числом параметров, что затрудняет как обучение модели, так и последующий поиск фазовой конфигурации, реализующей требуемую матрицу преобразования.

В настоящей главе предлагается альтернативный подход, основанный на переходе от исходного пространства фазовых параметров к специальному признаковому пространству. Основная идея состоит в том, чтобы использовать структуру линейно-оптической схемы для построения признаков, в которых зависимость элементов матрицы преобразования от фазовых сдвигов принимает линейный вид. Это позволяет свести задачу обучения модели интерферометра к решению линейной задачи наименьших квадратов и тем самым избежать невыпуклой оптимизации, характерной для прямого обучения в исходных параметрах.

Такой подход особенно эффективен для схем малой глубины, где полная модель интерферометра может быть восстановлена по обучающей выборке целиком. Для более глубоких схем рассматривается послойная стратегия, в которой обучение и настройка выполняются последовательно для отдельных фазовых слоёв. Оба подхода сохраняют архитектурную независимость и могут применяться к широкому классу линейно-оптических интерферометров.

3.1 Математические основы

В данной главе рассматривается многомодовый интерферометр, состоящий из чередующихся слоёв фазовых сдвигов и преобразований,

смешивающих моды. Такая структура охватывает широкий класс линейно-оптических схем и удобна для построения признакового описания модели. Топология интерферометра приведена на рисунке 3.1.

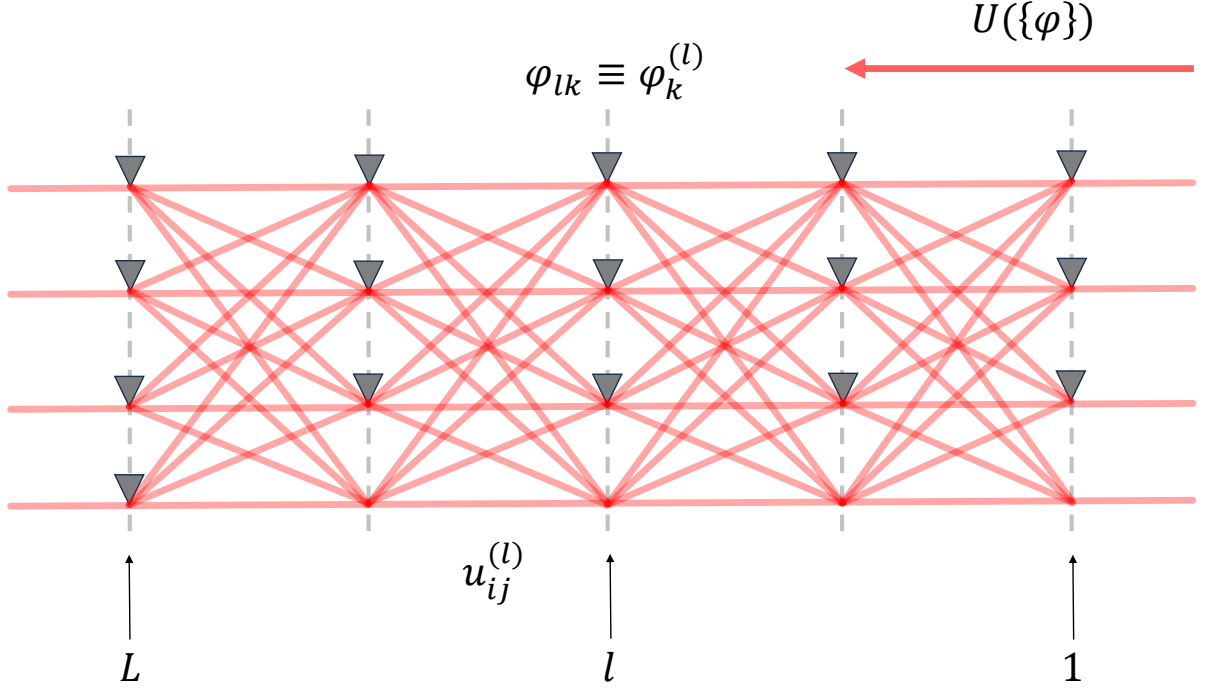


Рисунок 3.1 — Схема структуры многомодового интерферометра. Базисные матрицы U_ℓ описывают перераспределение излучения между каналами интерферометра. Соответствующие интегральные элементы могут быть реализованы в виде решеток связанных волноводов, что позволяет передавать энергию между различными волноводами [24].

Унитарная матрица U представляется в виде:

$$U = \Phi_L U_{L-1} \dots U_\ell \Phi_\ell \dots \Phi_2 U_1 \Phi_1, \quad (3.1)$$

где $\Phi_\ell = \text{diag}\{e^{i\varphi_{\ell k}}\} = \text{diag}\{e^{i\varphi_k^{(\ell)}}\}$ и $\{U_\ell\}_{ij} = u_{ij}^{(\ell)}$. Для дальнейших целей мы вводим обозначение $e^{i\varphi_k^{(\ell)}} \stackrel{\text{def}}{=} \psi_{\ell k} \stackrel{\text{def}}{=} \psi_k^{(\ell)}$. Таким образом, $\{\Phi_\ell\}_{ij} = \delta_{ij} e^{i\varphi_{\ell j}} = \delta_{ij} e^{i\varphi_j^{(\ell)}} = \delta_{ij} \psi_j^{(\ell)}$.

Мы называем U_ℓ *базисными* матрицами, поскольку они однозначно характеризуют матрицу передачи интерферометра U . Как только U_ℓ известны, численная оптимизация позволяет определить требуемые фазовые сдвиги $\varphi_{\ell k}$, завершая программирование устройства.

3.2 Алгоритмы обучения

3.2.1 Обучение на всех слоях

Рассмотрим уравнение (3.1), выраженное поэлементно, в частности, элемент матрицы интерферометра u_{ij} :

$$u_{ij} = \sum_{k_1, \dots, k_{L-2}} \psi_i^{(L)} u_{ik_1}^{(L-1)} \psi_{k_1}^{(L-1)} \dots u_{k_{L-2}j}^{(1)} \psi_j^{(1)}. \quad (3.2)$$

Далее мы отделяем множители, связанные с базисными матрицами $u_{pr}^{(\ell)}$, от множителей, связанных с фазовыми сдвигами $e^{i\varphi_k^{(\ell)}} \stackrel{\text{def}}{=} \psi_k^{(\ell)}$:

$$u_{ij} = \sum_{k_1, \dots, k_{L-2}} \underbrace{u_{ik_1}^{(L-1)} \dots u_{k_{L-2}j}^{(1)}}_{w_{i\bar{k}j}} \underbrace{\psi_i^{(L)} \psi_{k_1}^{(L-1)} \dots \psi_j^{(1)}}_{\theta_{i\bar{k}j}}. \quad (3.3)$$

Графическая иллюстрация вычисления матричного элемента u_{ij} по формуле (3.3) представлена на рисунке 3.2.

Следовательно, задача сводится к линейной модели, при этом каждый элемент матрицы интерферометра u_{ij} представлен в виде скалярного произведения вектора весов $w_{i\bar{k}j}$ и вектора признаков $\theta_{i\bar{k}j}$, который зависит только от фазовых сдвигов:

$$u_{ij} = \sum_{\bar{k}} w_{i\bar{k}j} \theta_{i\bar{k}j}, \quad \bar{k} = (k_1, \dots, k_{L-2}). \quad (3.4)$$

Таким образом, зависимость матрицы линейно-оптического интерферометра U от экспонент фазовых сдвигов $\varphi_k^{(\ell)}$ может быть охарактеризована как линейная модель, причём тензор третьего ранга $w_{i\bar{k}j}$ комплексно описывает эту модель.

Пусть i, j фиксированы, тогда $u_{ij} \stackrel{\text{def}}{=} u$, $w_{i\bar{k}j} \stackrel{\text{def}}{=} w_{\bar{k}}$, и $\theta_{i\bar{k}j} \stackrel{\text{def}}{=} \theta_{\bar{k}}$:

$$u = \sum_{\bar{k}} w_{\bar{k}} \theta_{\bar{k}}. \quad (3.5)$$

Предположим, что выполнено M измерений $u^{(m)}$, соответствующих M различным фазовым настройкам $\varphi_k^{(l)}$ и, следовательно, M значениям $\theta_{\bar{k}}^{(m)}$. Это приводит к системе линейных уравнений:

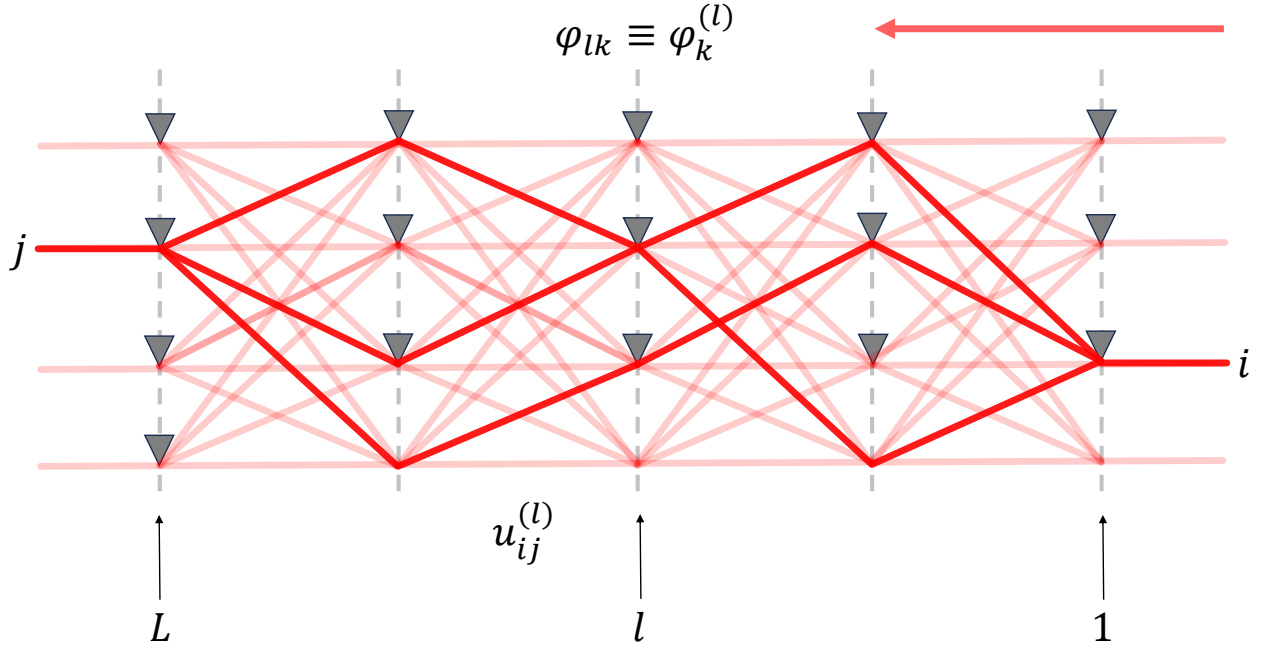


Рисунок 3.2 — Иллюстрация расчёта матричного элемента u_{ij} по уравнению (3.3). Матричный элемент u_{ij} формируется путем суммирования произведений весов $w_{i\bar{k}j}$ и новых признаков $\theta_{i\bar{k}j}$ по всем путям, соединяющим входную моду i и выходную моду j . На рисунке для примера показано суммирование только по трём путям для фиксированных индексов i и j , но в общем случае суммирование производится по всем путям.

$$\Theta W = U. \quad (3.6)$$

Матрица Θ представляет собой матрицу «объекты–признаки», содержащую M строк, соответствующих отдельным измерениям, и столбцы, соответствующие признакам $\theta_{\bar{k}}$. Вектор W состоит из весов $w_{\bar{k}}$, расположенных в формате вектора-столбца, а U — вектор-столбец, состоящий из значений $u^{(m)}$.

Система линейных уравнений (3.6), построенная по экспериментальным данным томографии линейно-оптического интерферометра U , в общем случае содержит ошибку томографической реконструкции ε :

$$\Theta W = U + \varepsilon. \quad (3.7)$$

Для применения аналитического решения предполагается линейная независимость столбцов матрицы «объекты–признаки». Кроме того, число строк M , соответствующих различным наборам фазовых сдвигов $\{\varphi_k^{(\ell)}\}$,

используемым при томографии матриц преобразования $U^{(m)}$ линейно-оптического интерферометра, должно быть не меньше числа признаков. Такая постановка позволяет сформулировать задачу нахождения весового вектора (при фиксированных индексах i и j) как задачу наименьших квадратов:

$$\|\Theta W - U\|_2^2 \rightarrow \min_W. \quad (3.8)$$

Такая задача минимизации может быть решена аналитически с помощью псевдоинверсии Мура-Пенроуза:

$$W = (\Theta^\dagger \Theta)^{-1} \Theta^\dagger U. \quad (3.9)$$

Теперь рассмотрим общий функционал (3.8), включающий индексы i и j :

$$\mathcal{L} = \frac{1}{M} \sum_{m=1}^M \sum_{i,j} \left| \sum_{\bar{k}} w_{i\bar{k}j} \theta_{i\bar{k}j}^{(m)} - u_{ij}^{(m)} \right|^2 \rightarrow \min_{w_{i\bar{k}j}}. \quad (3.10)$$

Матрица $F = (\Theta^\dagger \Theta)^{-1} \Theta^\dagger$, как показано в уравнении (3.9), не зависит от индексов i и j элемента матрицы u_{ij} . Поэтому F нужно вычислять только один раз, после чего можно непосредственно сформулировать решение для тензора $w_{i\bar{k}j}$, минимизирующее функционал $\mathcal{L}$, заданный уравнением (3.10):

$$w_{i\bar{k}j} = \operatorname{argmin}_{w_{i\bar{k}j}} \mathcal{L} = \sum_m F_{\bar{k}m} u_{ij}^{(m)}. \quad (3.11)$$

Задача минимизации функционала (3.10) также может быть решена численно, например, с помощью методов градиентного спуска (gradient descent, GD) или стохастического градиентного спуска (stochastic gradient descent, SGD).

Теперь обсудим требования к обучающим данным. Для эффективной работы алгоритма число строк M матрицы Θ в уравнении 3.9, то есть число различных наборов фазовых сдвигов $\{\varphi_k^{(\ell)}\}$, при которых выполняется томография матрицы преобразования интерферометра $U^{(m)}$, должно быть не меньше числа столбцов. Это число определяется числом признаков $\theta_{i\bar{k}j} = \psi_i^{(L)} \psi_{k_1}^{(L-1)} \dots \psi_j^{(1)} = e^{i\varphi_i^{(L)}} e^{i\varphi_{k_1}^{(L-1)}} \dots e^{i\varphi_j^{(1)}}$. В случае полной ($L = N + 1$) линейной оптической сети, показанной на рисунках 3.1 и 3.2, количество признаков составляет $O(N^L)$, что приводит к неэффективному масштабированию с увеличением числа мод N . Поэтому данный алгоритм наиболее эффективен, когда сеть имеет небольшую глубину $L \ll N$.

До сих пор мы рассматривали сценарий, в котором цифровая модель (3.4) линейно-оптического интерферометра строится на основе томографических измерений матриц преобразования $U^{(m)}$, выполненных при различных фазовых сдвигах. По результатам этой томографии определяется весовой тензор $w_{i\bar{k}j}$, после чего в цифровой модели оптимизируются фазовые задержки $\varphi_k^{(\ell)}$ для реализации заданного целевого преобразования U_{target} . Однако возможен и другой подход — непосредственная послойная оптимизация физического интерферометра.

3.2.2 Послойное обучение

В этом алгоритме для реализации целевого преобразования U_{target} параметры интерферометра последовательно оптимизируются по слоям, при необходимости с несколькими проходами по всем слоям. На каждом шаге, соответствующем слою ℓ , изменяются только фазы этого слоя, тогда как параметры остальных слоёв сохраняются фиксированными. Таким образом, в рассматриваемом сценарии реализуется метод попеременных наименьших квадратов (alternating least squares, ALS) [54].

Если в качестве оптимизируемых переменных рассматривать только фазы текущего слоя ℓ , то интерферометр на данном шаге обучения можно локально представить в следующем виде:

$$U = U_L^{(\ell)} \Phi^{(\ell)} U_R^{(\ell)}, \quad (3.12)$$

то есть как интерферометр с одним фазовым слоем. Соответствующая схема приведена на рисунке 3.3. Важно отметить, что этот подход также эквивалентен модели, описанной выражением (3.4), хотя в этом случае наши признаки упрощаются до фазовых экспонент $\theta_{i\bar{k}j} = \psi_k^{(\ell)} = e^{i\varphi_k^{(\ell)}}$. Далее мы можем обучить линейную модель для этого одного фазового слоя, используя $O(N)$ томографических образцов, и численно оптимизировать фазы для приближения к целевому преобразованию U_{target} :

$$\Phi^{(\ell)} = \underset{\Phi^{(\ell)}}{\operatorname{argmin}} ||U_{\text{model}}(\Phi) - U_{\text{target}}||_2^2. \quad (3.13)$$

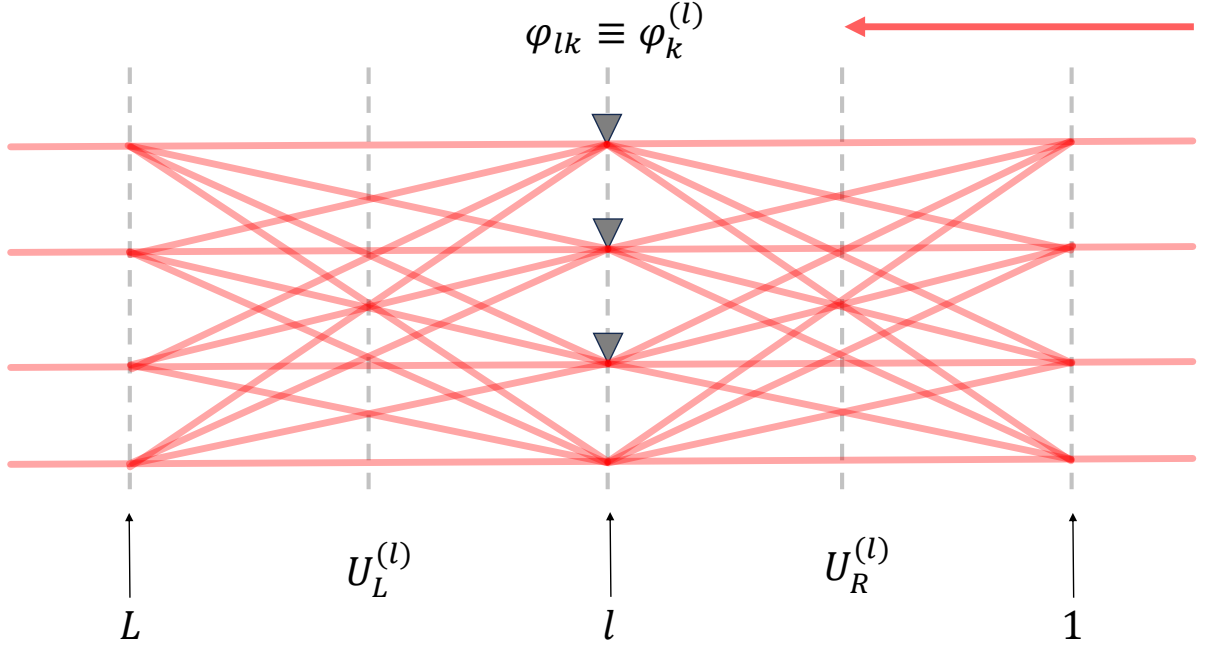


Рисунок 3.3 — Иллюстрируя подход к рассмотрению одного фазового слоя линейно-оптического интерферометра, мы рассматриваем статические базисные матрицы $U_L^{(l)}$ и $U_R^{(l)}$ (3.12) слева и справа от фазового слоя на каждом шаге алгоритма.

Для решения возникающей задачи оптимизации могут использоваться градиентные методы, в частности BFGS [55]. После выполнения этой локальной минимизации мы переходим к следующему слою, проводим томографию и обучаем локальную линейную модель (3.12).

Преимущество данного подхода состоит в том, что требуемый объём обучающей выборки растёт линейно с числом мод, то есть как $O(N)$. Его недостатком является необходимость выполнять томографию физического устройства на каждом слое. В предыдущем алгоритме, напротив, серия томографических измерений, необходимая для восстановления полной модели линейно-оптического интерферометра, проводилась только один раз. При этом в рассматриваемом подходе нет необходимости напрямую измерять на устройстве градиент функции потерь по фазам, поскольку он вычисляется на основе локальной линейной модели интерферометра.

3.3 Численные симуляции

3.3.1 Симуляция обучения на всех слоях

Здесь мы исследуем характеристики алгоритма, представленного в разделе 3.2.1, на примере линейно-оптического интерферометра, показанного на рисунке 3.1. Для моделирования интерферометра генерируются $L - 1$ базисных матриц U_ℓ , случайным образом выбранных в соответствии с равномерным распределением Хаара с использованием QR-разложения [40].

Задача алгоритма обучения на всех слоях состоит в восстановлении полной модели линейно-оптического интерферометра (3.4) по обучающей выборке $(\{\varphi^{(m)}\}, U^{(m)})$ объёма M . Для создания этого набора данных мы равномерно генерируем случайные фазы в диапазоне от 0 до 2π , определяя каждую матрицу $U^{(m)}$ с помощью сгенерированных базисных матриц, этих случайных фаз и разложения (3.1).

В реальных физических системах при получении набора данных $(\{\varphi^{(m)}\}, U^{(m)})$ также присутствуют томографические ошибки ε (3.7), которые мы должны учитывать в нашем моделировании. Для моделирования этих ошибок к каждому элементу $u_{ij}^{(m)}$ матрицы $U^{(m)}$ добавляется комплексная случайная погрешность $\varepsilon_{ij}^{(m)}$:

$$\{U_{\text{noise}}^{(m)}\}_{ij} = u_{ij}^{(m)} + \varepsilon_{ij}^{(m)}, \quad (3.14)$$

причём

$$\varepsilon_{ij}^{(m)} = \frac{\varepsilon}{\sqrt{2}}(x + iy), \quad (3.15)$$

где $x \in \mathcal{N}(0, 1)$, $y \in \mathcal{N}(0, 1)$ — случайные величины, а ε — фиксированный параметр. Таким образом, мы получаем обновлённый набор данных $(\{\varphi^{(m)}\}, U_{\text{noise}}^{(m)})$, который также содержит гауссовский шум. При этом матрица $U_{\text{noise}}^{(m)}$ после добавления ошибки, вообще говоря, перестаёт быть строго унитарной. В рамках рассматриваемой модели это не является проблемой, поскольку $U_{\text{noise}}^{(m)}$ интерпретируется не как точное физически реализуемое преобразование, а как результат его томографической оценки с конечной точностью. Затем, когда обучающий набор данных сгенерирован, мы обучаем веса модели $w_{i\bar{k}j}$ аналитически, используя формулу (3.11).

Чтобы проверить качество обучения, мы выбираем набор случайных фаз $\{\varphi\}$, который не входит в обучающую выборку ($\{\varphi^{(m)}\}, U_{\text{noise}}^{(m)}$). Затем мы оцениваем предсказание модели для этих фаз, $U_{\text{model}}(\{\varphi\})$, и сравниваем его с истинными значениями $U(\{\varphi\})$ (без учёта томографических ошибок). В качестве меры близости между $U_{\text{model}}(\{\varphi\})$ и $U(\{\varphi\})$ используется норма Фробениуса:

$$\mathcal{L}_{\text{Frobenius}} = \frac{1}{N} \|U_{\text{model}}(\{\varphi\}) - U(\{\varphi\})\|_2^2. \quad (3.16)$$

Для исследования свойств алгоритма обучения используются результаты, усреднённые по множеству реализаций линейно-оптических интерферометров, обучающих выборок и тестовых оценок, рассчитанных на основе нормы Фробениуса.

На рисунке 3.4 представлена зависимость метрики Фробениуса, усреднённой по 1000 образцам линейно-оптических интерферометров, от размера обучающей выборки M для различных уровней шума ε при $N = 4$ модах и $L = 5$ слоях. Наблюдается, что успешное обучение, обеспечивающее высокую точность на тестовой выборке, достигается лишь при превышении некоторого минимального порога $M_{\min}$, как обсуждалось в разделе 3.2.1.

Рассмотрим асимптотическое поведение данной зависимости при больших значениях M . Для $M \gg M_{\min}$ ожидается, что метрика Фробениуса будет уменьшаться обратно пропорционально M :

$$\mathcal{L}_{\text{Frobenius}} \sim O(M^{-1}), \quad (3.17)$$

поскольку параметры линейной модели оцениваются по M независимым томографическим измерениям, а ошибка их оценивания убывает с ростом объёма выборки. Такое масштабирование согласуется с общими результатами статистической теории оценивания [56].

Рисунок 3.5 подтверждает это предсказание: зависимость $\mathcal{L}_{\text{Frobenius}}$ от M при значениях, значительно превышающих пороговое значение $M_{\min}$, является линейной в логарифмическом масштабе. Метод наименьших квадратов для аппроксимации численных данных функциональной формой CM^{-k} дает показатель степени $k \approx 1$, что находится в отличном согласии с теоретическим ожиданием.

Аналогично, зависимость ошибки реконструкции от уровня томографических ошибок ε (рисунок 3.6) может быть объяснена с помощью теории

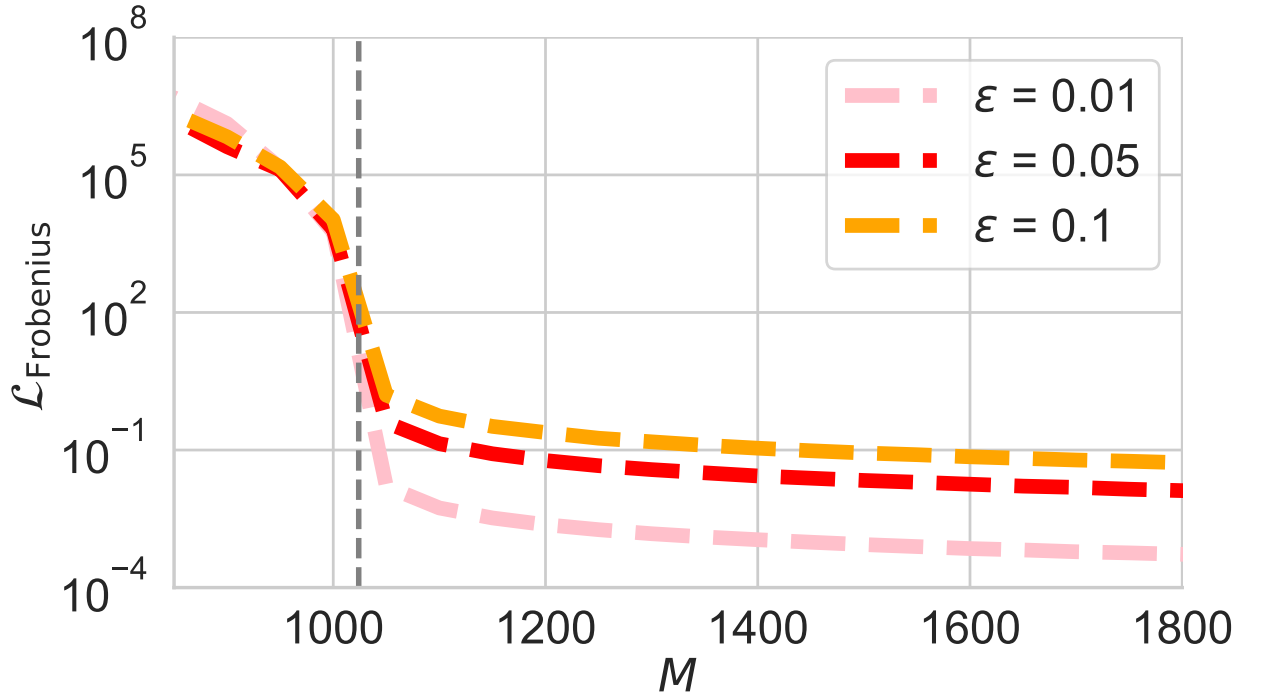


Рисунок 3.4 — Зависимость метрики Фробениуса на тестовой выборке в зависимости от размера обучающей выборки M . Усреднение выполняется по 1000 реализациям линейно-оптических интерферометров при $N = 4$ модах и $L = 5$ слоях. Серая вертикальная пунктирная линия обозначает минимальный размер обучающей выборки, необходимый для успешного прохождения проверки на тестовом множестве: $M_{\min} = 1024$.

информации Фишера: в присутствии аддитивного гауссова шума с дисперсией, пропорциональной ε^2 , достижимая точность оценки уменьшается пропорционально ε^2 , что приводит к наблюдаемой линейной зависимости метрики Фробениуса от ε^2 .

3.3.2 Симуляция послойного обучения

В этом подразделе мы сосредоточимся на исследовании свойств алгоритма послойного обучения, описанного в разделе 3.2.2. Мы оцениваем свойства алгоритма настройки интерферометра на преобразование U_{target} , учитывая число мод интерферометра N и его глубину L .

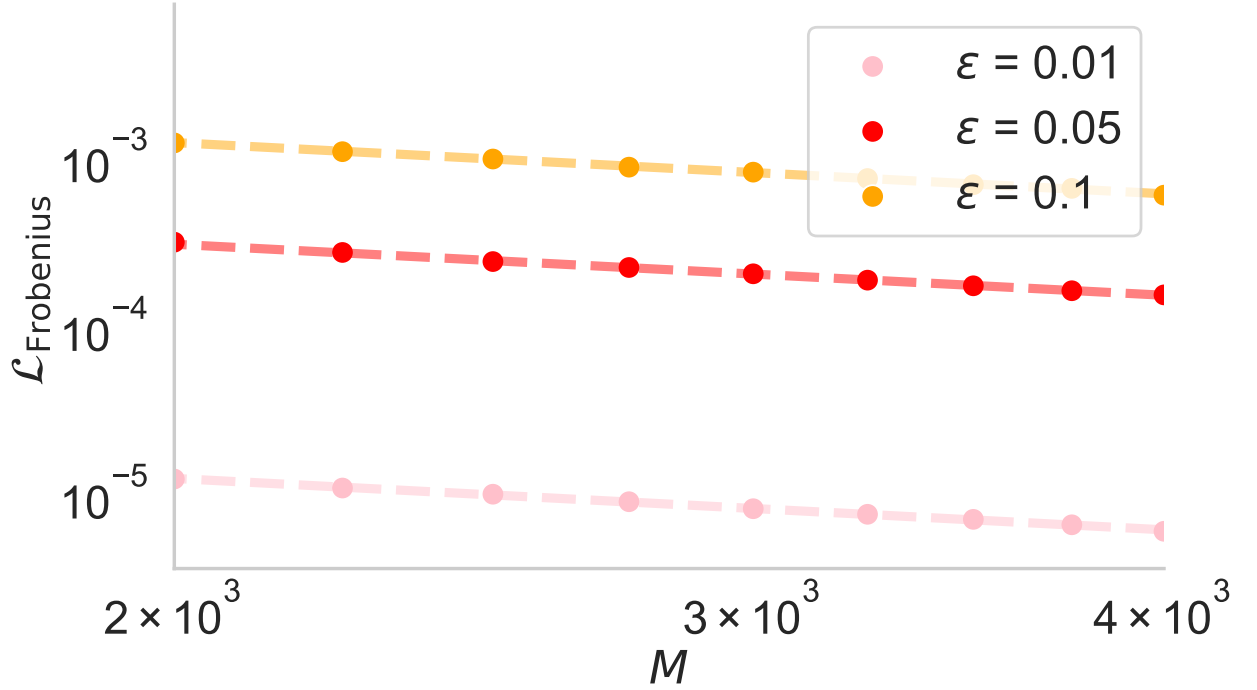


Рисунок 3.5 — Зависимость метрики Фробениуса на тестовой выборке в зависимости от размера обучающей выборки M . Усреднение проводится на 1000 примеров линейно-оптических интерферометров с числом мод $N = 3$ и числом слоев $L = 4$. Рассматриваются значения $M \gg M_{\min} = 81$. Точки моделирования аппроксимируются соотношением $\mathcal{L}_{\text{Frobenius}} = \mathcal{C}M^{-k}$ методом наименьших квадратов, с результатами: $k = (1.017 \pm 0.022)$ для $\varepsilon = 0.01$, $k = (1.010 \pm 0.013)$ для $\varepsilon = 0.05$, $k = (1.014 \pm 0.024)$ для $\varepsilon = 0.1$.

Мы также рассматриваем архитектуру линейно-оптического интерферометра, изображенного на рисунке 3.1. Сначала мы генерируем $L - 1$ базисные матрицы U_ℓ , которые случайным образом выбираются из равномерного распределения Хаара с помощью QR-разложения [40].

Затем мы генерируем случайные фазы, равномерно распределенные между 0 и 2π , и используем их для определения U_{target} . Для этого используются сгенерированные базисные матрицы, случайные фазы и разложение (3.1) для определения U_{target} . На этом этапе моделирования мы фиксируем наш линейно-оптический интерферометр и целевое преобразование U_{target} , которое мы хотим реализовать.

Далее задача настройки интерферометра состоит в подборе фаз $\{\varphi\}$, начиная с некоторого начального приближения. В результате нескольких итераций алгоритма, описанного в разделе 3.2.2, необходимо найти такие

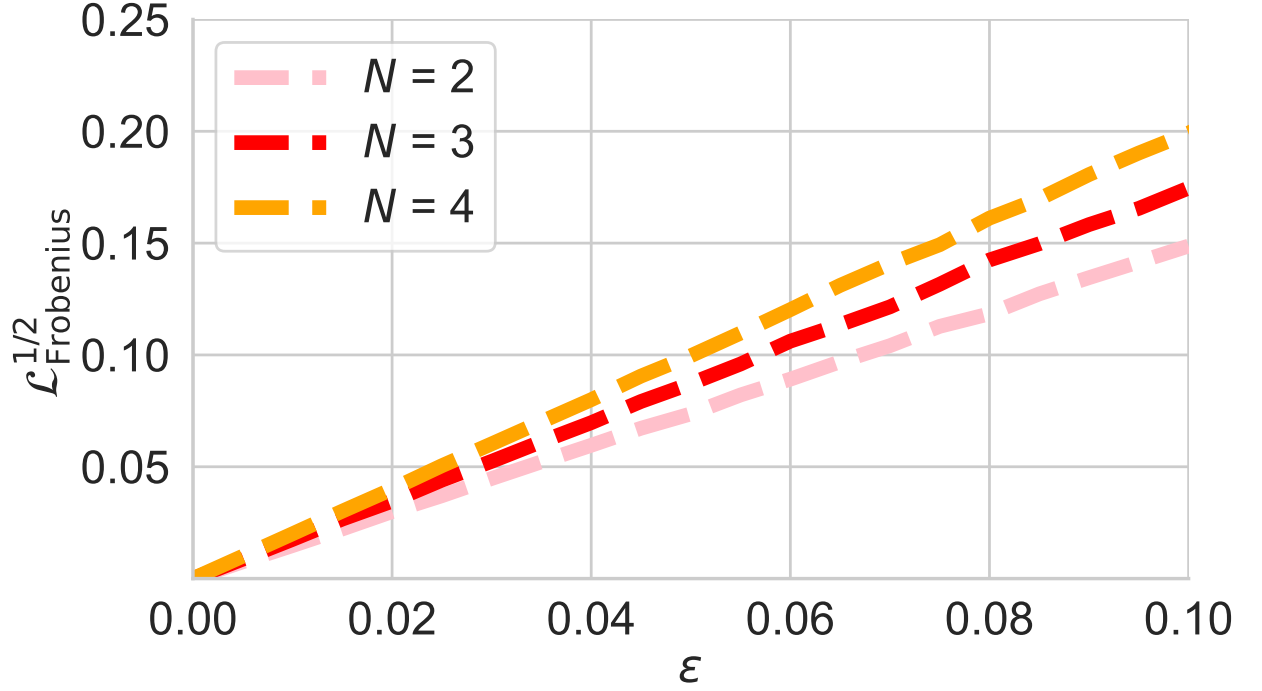


Рисунок 3.6 — Зависимость метрики Фробениуса на тестовой выборке в зависимости от уровня шума ϵ в обучающей выборке. Усреднение производится по 10000 примерам линейно-оптических интерферометров для $N = 2$ и 1000 примерам для $N = 3$ и $N = 4$. Рассматриваются полнослойные интерферометры, где $L = N + 1$. Размер обучающей выборки в два раза больше минимального, $M = 2M_{\min}$, при этом M равно 16, 162 и 2048 соответственно.

значения фаз, при которых преобразование интерферометра $U_{\text{model}}(\{\varphi\})$ наиболее близко к целевому преобразованию U_{target} в смысле нормы Фробениуса: $\|U_{\text{model}}(\{\varphi\}) - U_{\text{target}}\|_2^2 \rightarrow \min_{\{\varphi\}}$.

Для реализации алгоритма ALS выполняется 1000 последовательных проходов по слоям линейно-оптического интерферометра. В рамках каждого прохода слои интерферометра $\ell = 1, \dots, L$ обрабатываются последовательно. Для каждого слоя томография моделируется при $M = N + 1$ значениях фаз, поскольку число обучающих образцов должно быть не меньше N , в то время как фазы остальных слоёв фиксируются. По полученной выборке обучается модель (3.12) согласно процедуре, описанной в разделе 3.2.2. После этого с использованием обученной модели выполняется минимизация соответствующего функционала методом BFGS; на каждом шаге ALS проводится 5 итераций этого метода.

В практических приложениях элементы обучающего набора образцов $(\{\varphi^{(m)}\}, U^{(m)})$ получаются из алгоритмов унитарной реконструкции $U_{rec}^{(m)}$ [5; 36], применяемых к реконфигурируемому интерферометру, запрограммированному с фазами $\{\varphi^{(m)}\}$.

Мы провели эксперимент на 50 случайно сгенерированных линейно-оптических интерферометрах, чтобы проследить, как функция потерь Фробениуса $\mathcal{L}_{\text{Frobenius}} = \frac{1}{N} \|U_{\text{model}}(\{\varphi\}) - U_{\text{target}}\|_2^2$ зависит от числа итераций в случае интерферометра полной глубины ($L = N + 1$). Усреднённые результаты настройки линейно-оптического интерферометра представлены на рисунке 3.7. Отметим, что при экспериментальной настройке необходимое число обучающих примеров M для каждого фиксированного слоя растёт как $O(N)$.

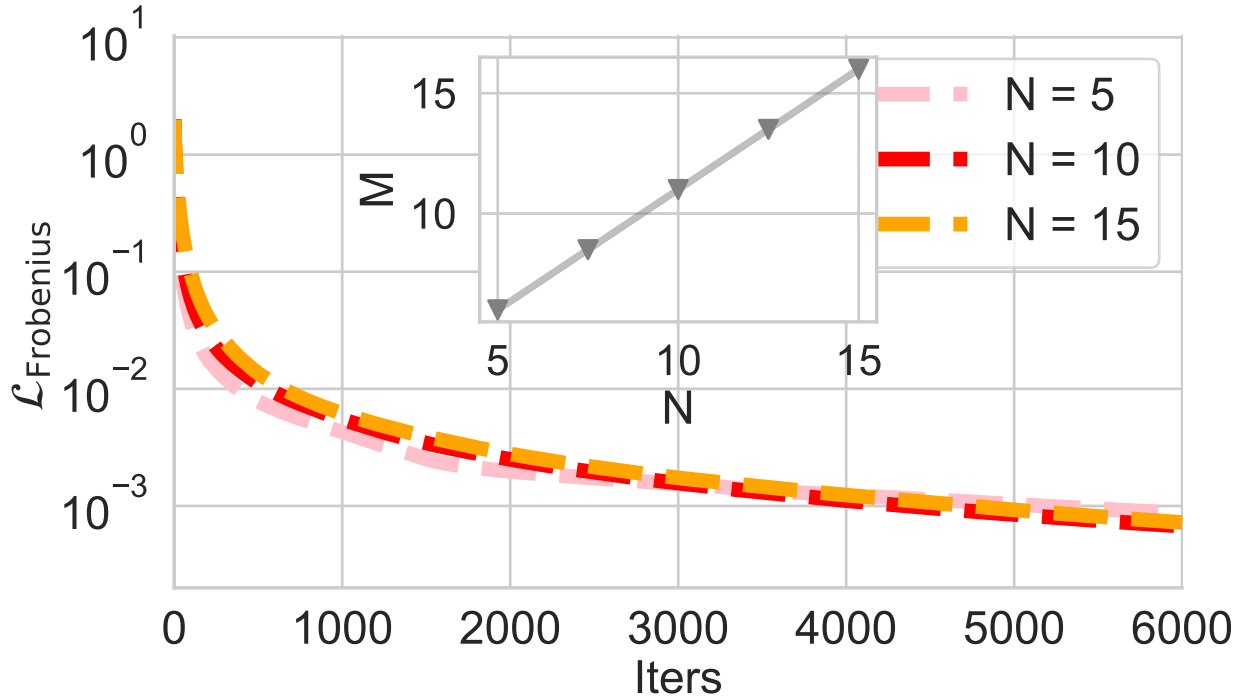


Рисунок 3.7 — Иллюстрация процесса настройки полнослойного линейно-оптического интерферометра ($L = N + 1$). Показана зависимость функции потерь Фробениуса $\mathcal{L}_{\text{Frobenius}} = \frac{1}{N} \|U_{\text{model}}(\{\varphi\}) - U_{\text{target}}\|_2^2$, усреднённой по 50 независимым запускам алгоритма настройки, от числа итераций, равного произведению числа проходов по слоям на общее число слоёв интерферометра. На врезке приведена линейная зависимость размера обучающей выборки M от числа мод N : $M = N + 1$.

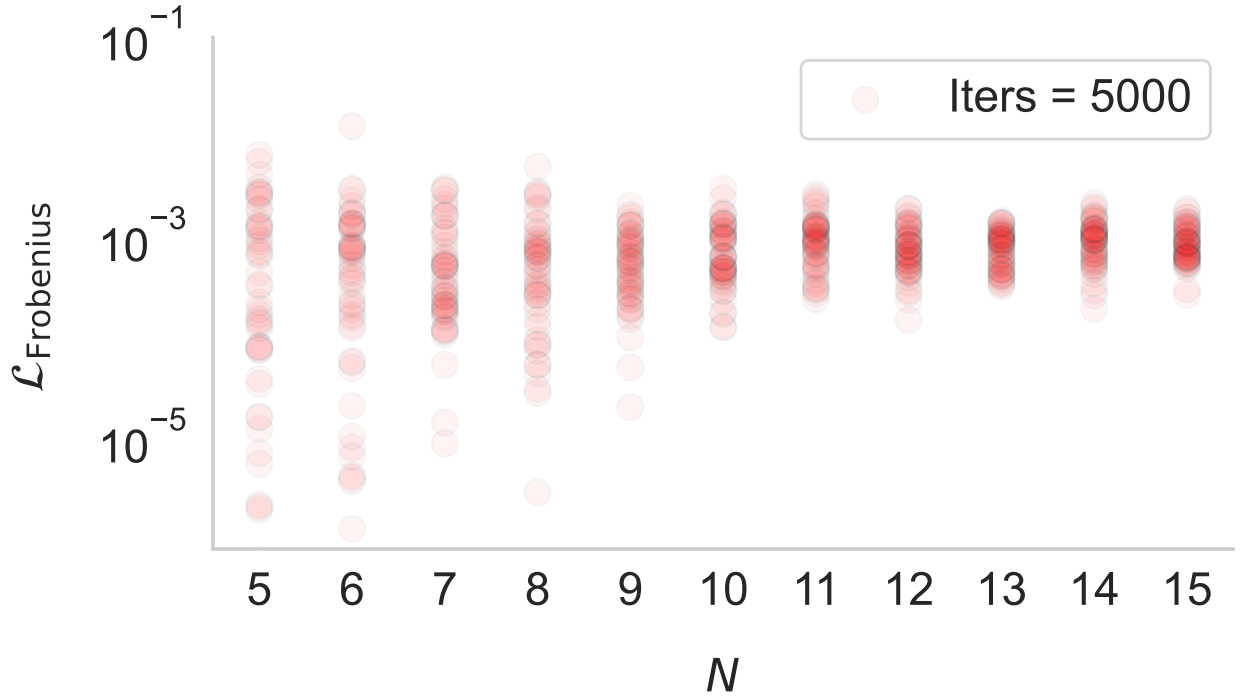


Рисунок 3.8 — Иллюстрация процесса настройки полнослойного линейно-оптического интерферометра ($L = N + 1$). На рисунке представлены значения функции потерь Фробениуса $\mathcal{L}_{\text{Frobenius}} = \frac{1}{N} \|U_{\text{model}}(\{\varphi\}) - U_{\text{target}}\|_2^2$ на 5000-й итерации для 50 независимых запусков алгоритма настройки и различных значений числа мод N .

На рисунке 3.7 показаны результаты настройки линейно-оптического интерферометра без учета шума ε случайных ошибок, возникающих в процессе послойной томографии.

Отдельный интерес представляет вопрос о наличии локальных минимумов, поскольку на каждом шаге оптимизируются фазы только одного слоя, а не всех слоёв одновременно. В принципе такие эффекты локализации могут препятствовать достижению глобального минимума. Однако результаты, представленные на рисунке 3.8, показывают, что во всех 50 запусках оптимизации послойный метод демонстрирует стабильную сходимость без значительных выбросов. Это свидетельствует о хорошей сходимости метода в рассмотренных численных экспериментах.

В реальных устройствах неизбежно присутствует шум, модель которого обсуждалась в разделе 3.3.1. Поэтому далее мы учитываем шум и в процедуре настройки интерферометра. С этой целью проводится тот же эксперимент, что и для рисунка 3.7, однако теперь в каждом слое томогра-

фия моделируется для $M = 10N$ значений фаз, поскольку, как следует из результатов, представленных на рисунках 3.4 и 3.5, ошибка обучения модели уменьшается с ростом M . При этом в моделировании учитывается шум уровня ε , как описано в разделе 3.3.1. Результаты настройки в присутствии шума представлены на рисунке 3.9. Отметим, что при экспериментальной настройке требуемое число обучающих примеров M для каждого фиксированного слоя по-прежнему растёт как $O(N)$.

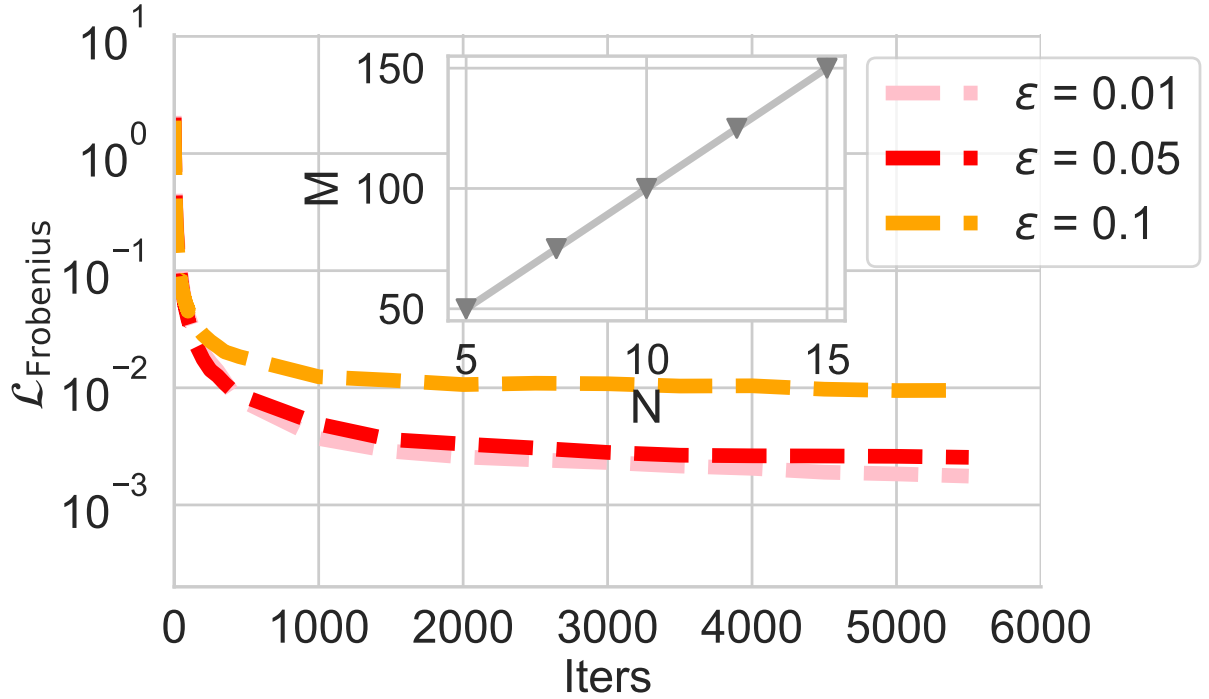


Рисунок 3.9 — Иллюстрация процесса настройки для полнослойного ($L = N + 1$) линейно-оптического интерферометра с послойной томографией, выполненного в условиях шума ε . В данном анализе рассматривается $N = 5$, $L = 6$. На рисунке показано, как функция потерь Фробениуса, усредненная по 50 запускам алгоритма настройки $\mathcal{L}_{\text{Frobenius}} = \frac{1}{N} \|U_{\text{model}}(\{\varphi\}) - U_{\text{target}}\|_2^2$, изменяется с количеством итераций, где каждая итерация определяется как произведение количества проходов слоев и общего количества слоев. График на вставке иллюстрирует линейную зависимость размера обучающей выборки M от количества режимов N ($M = 10N$).

3.4 Практические аспекты экспериментальной реализации

В данном разделе обсуждается масштабируемость методов, представленных в разделах 3.2.1 и 3.2.2, с точки зрения их потенциальной экспериментальной реализации.

Оба подхода требуют полной томографии унитарного преобразования, выполненного с помощью линейно-оптического интерферометра при заданных фазовых настройках $\{\varphi\}$ с точностью до *глобальной фазы*. Это требование накладывает определённые ограничения на методы томографии, которые могут быть использованы на практике.

В работе [4] предложен эффективный интерферометрический протокол для прямой томографии комплексной матрицы переноса U N -модовой линейно-оптической сети, основанный исключительно на стандартных лазерных источниках и фотодетекторах. Этот метод не требует вспомогательного интерферометра или неклассической интерференции. Протокол требует всего лишь $(2N - 1)$ различных входных конфигураций — N одно-модовых и $N - 1$ двухмодовых когерентных состояний — для определения всех относительных фаз и модулей M , при этом все N выходных портов измеряются одновременно в каждой конфигурации. Это обеспечивает общее масштабирование $O(N)$ по числу конфигураций измерения, несмотря на $O(N^2)$ число независимых параметров в U .

Данная схема может быть естественным образом расширена для восстановления U до его *глобальной фазы* путём введения единственного опорного луча, полученного от того же лазерного источника. Путём интерференции одной выбранной выходной моды с этим известным опорным фазовым лучом на 50 : 50 светоделителе можно определить абсолютное фазовое смещение всей матрицы без необходимости создания вспомогательного интерферометра. Эта модификация сохраняет масштабирование $O(N)$ и увеличивает количество необходимых конфигураций только с $2N - 1$ до $2N$.

Таким образом, экспериментальная сложность наших алгоритмов имеет следующий вид: для алгоритма *всех слоев* $O(N^{L+1})$; для алгоритма *последовательного* $O(N^2)$ для одного слоя и $O(PN^2L)$ для P проходов через все L слоев.

В нашей текущей реализации оптимальные веса линейной модели получаются путем минимизации функционала нормы Фробениуса (3.10), что представляет собой выпуклую задачу оптимизации, решаемую аналитически с использованием псевдообратной матрицы Мура-Пенроуза.

Если в реальном эксперименте доступны только модули элементов матриц переноса, функционал потерь необходимо модифицировать, определив его через модули этих элементов:

$$\mathcal{L}' = \frac{1}{M} \sum_{m=1}^M \sum_{i,j=1}^N \left(\left| \sum_{\bar{k}} w_{i\bar{k}j} \theta_{i\bar{k}j} \right|^2 - |u_{ij}|^2 \right)^2. \quad (3.18)$$

Подобная оптимизация, как правило, приводит к неудовлетворительным результатам [57] вследствие потери фазовой информации, которая вызывает неоднозначности и снижает сходимость при сопоставлении только модулей амплитуд.

Одним из возможных способов решения данной проблемы является применение алгоритма PhaseLift [5], в котором задача восстановления фазы переформулируется как задача выпуклого программирования. Это даёт возможность восстанавливать недостающую фазовую информацию по измерениям одной лишь интенсивности. Использование PhaseLift в рамках рассматриваемого метода потенциально делает возможным обучение при отсутствии интерферометрического доступа к фазе. Вместе с тем включение такого подхода выходит за рамки настоящей диссертации и представляет интерес для дальнейших исследований.

В заключение данного раздела обсудим применимость разработанных методов к современным многомодовым линейно-оптическим интерферометрам с большим числом мод, например при $N \approx 100$. Недавние достижения в экспериментальной реализации реконфигурируемых линейных оптических интерферометров сделали возможным создание устройств с десятками и даже сотнями мод [58; 59].

Если такой интерферометр содержит только один фазовый слой (типичным примером является реализация в свободном пространстве с использованием пространственного модулятора света (spatial light modulator, SLM) в качестве единственного фазового слоя), то алгоритм «всех слоёв»

может быть применён особенно эффективно. В этом случае после изготовления устройства достаточно однократно выполнить $O(N)$ томографий при различных фазовых настройках $\{\varphi\}$. Как было показано выше, это требует порядка $O(N^2)$ экспериментальных измерений — примерно 10,000 измерений, что соответствует приблизительно 100 томографическим реконструкциям, — и находится в пределах возможностей современных средств фазовой модуляции.

Для полнослойных интерферометров, где число фазовых слоев составляет порядка $L \sim N$, метод «всех слоев» потребовал бы $O(N^L)$ томографических реконструкций (например, 100^{100} для $N = L = 100$), что заведомо недостижимо в эксперименте. В таких случаях следует использовать алгоритм «послойного обучения». Тогда обучение одного слоя требует около 100 томографий (примерно 10,000 измерений). Полная оптимизация к целевому преобразованию U_{target} включает в себя ряд проходов P (обычно $P \approx 50$, как показано на рисунке 3.7), в результате чего общее масштабирование составляет $O(PNL) \approx 5 \times 10^5$ томографий ($O(PN^2L) \approx 5 \times 10^7$ измерений).

При использовании современных технологий фазового переключения получение такого набора данных не представляет принципиальных экспериментальных трудностей и может быть осуществлено за разумное время. В частности, в интегральных оптических платформах электрооптическое управление допускает очень высокие скорости перенастройки [60]; для отдельных устройств были продемонстрированы полосы пропускания около 100 ГГц.

3.5 Выводы по главе

В данной главе предложен метод программирования линейно-оптического интерферометра, основанный на переходе к специальному признаковому пространству, в котором зависимость матрицы преобразования от фазовых параметров становится линейной. Это позволило свести задачу обучения модели интерферометра к задаче наименьших квадратов и получить аналитическое выражение для параметров модели.

На основе этого подхода рассмотрены два алгоритма настройки. Алгоритм обучения на всех слоях эффективен для интерферометров малой глубины и позволяет после однократной томографии физического устройства выполнять дальнейшую оптимизацию на цифровой модели. Послойный алгоритм применим к схемам большой глубины и основан на последовательной настройке фазовых слоёв с использованием локальных линейных моделей.

Численные симуляции подтвердили работоспособность обоих алгоритмов. Для первого алгоритма установлены требования к минимальному объёму обучающей выборки и исследовано влияние томографического шума на точность восстановления модели. Для второго алгоритма показана устойчивая сходимость процедуры настройки к целевому преобразованию, в том числе в шумовых условиях.

Разработанные методы применимы к широкому классу линейно-оптических интерферометров, однако их практическая реализация требует достаточно точной томографии комплексной матрицы преобразования и контроля экспериментальных факторов, таких как потеря фазовой информации и перекрёстные помехи между фазовыми модуляторами.

Результаты третьей главы опубликованы в работе [61].

Глава 4. Симуляция квантовых систем с использованием формализма тензорных сетей

Данная глава посвящена разработке и исследованию методов классической симуляции квантовых систем с использованием формализма тензорных сетей. Актуальность такого подхода обусловлена тем, что размерность гильбертова пространства квантовой системы растёт экспоненциально с увеличением числа кубитов, вследствие чего прямое представление квантового состояния в виде полного вектора амплитуд становится вычислительно затруднительным уже для систем умеренного размера. В этих условиях особый интерес представляют методы сжатого представления квантовых состояний, позволяющие учитывать структуру квантовых корреляций и существенно снижать требования к памяти и времени вычислений.

Одним из наиболее известных и эффективных подходов к компактному представлению многокубитных квантовых состояний являются тензорные сети. В настоящей диссертации основное внимание уделено формату тензорных поездов (tensor train, TT), который в контексте квантовых вычислений также известен как matrix product state (MPS). Данный формат позволяет представить тензор амплитуд квантового состояния в виде последовательности тензоров меньшей размерности, связанных между собой вспомогательными индексами. Эффективность такого представления определяется TT-рангами, которые отражают степень запутанности между различными подсистемами. Для широкого класса состояний, обладающих ограниченной структурой запутанности, TT-формат обеспечивает компактное хранение и эффективное выполнение базовых операций, возникающих при моделировании квантовых схем.

Целью данной главы является изложение математического аппарата тензорных сетей, необходимого для дальнейшего моделирования квантовых состояний и операций над ними, а также разработка алгоритмов симуляции квантовых схем в формате тензорных поездов. В главе рассматриваются базовые тензорные разложения, приводится обзор основных архитектур тензорных сетей и существующих программных средств, описывается представление квантовых состояний в TT-формате и программная реализация соответствующих алгоритмов в разработанной автором библиотеке QTensor,

а также исследуются алгоритмы применения одно- и двухкубитных гейтов, ортонормализации, вычисления скалярных произведений и локальных матриц плотности. В заключительной части главы представлены результаты численных экспериментов, демонстрирующие возможности разработанного подхода при моделировании случайных квантовых схем и анализе характеристик получаемых состояний.

4.1 Формализм тензорных сетей

4.1.1 Сингулярное разложение

В линейной алгебре сингулярное разложение (singular value decomposition, SVD) представляет собой факторизацию произвольной комплексной матрицы A ранга r и размером $N \times M$ в следующее произведение трёх матриц:

$$A = USV^\dagger, \quad (4.1)$$

где U и V — матрицы размера $N \times r$ и $M \times r$, у которых столбцы ортонормированы, S — диагональная матрица размером $r \times r$, на диагонали которой расположены по убыванию неотрицательные вещественные числа $s_1 \geq s_2 \geq \dots \geq s_r > 0$, называемые сингулярными числами матрицы A .

Сингулярное разложение обладает важным свойством, которое делает важным его использование в тензорных сетях. Это свойство выражается в теореме Эккарта-Янга [62]: усечённое сингулярное разложение матрицы A (truncated SVD) даёт наиболее точное по норме Фробениуса приближение матрицы A ранга r матрицей A_k ранга $k < r$:

$$A_k = \underset{\text{rank}(B)=k}{\text{argmin}} \|A - B\|_F = US_kV^\dagger, \quad (4.2)$$

где S_k — матрица S , в которой заменили нулями все диагональные элементы, кроме k наибольших элементов.

Сингулярное разложение можно переписать в тензорном виде. Пусть матрица A ($\text{rank } A = r$) имеет комплексные элементы $A_{i_1 i_2}$. Тогда сингулярное разложение тензора $A_{i_1 i_2}$ можно записать в виде:

$$A_{i_1 i_2} = \sum_{k_1=1}^r U_{i_1 k_1} s_{k_1} V_{i_2 k_1}^*. \quad (4.3)$$

Далее, введя обозначения $A_{i_1 k_1}^{(1)} = U_{i_1 k_1} s_{k_1}$, $A_{i_2 k_1}^{(2)} = V_{i_2 k_1}^*$, получаем следующую факторизацию матрицы A :

$$A_{i_1 i_2} = \sum_{k_1=1}^r A_{i_1 k_1}^{(1)} A_{i_2 k_1}^{(2)}. \quad (4.4)$$

4.1.2 Тензорные разложения

Тензоры являются естественным многомерным обобщением матриц и в последние годы вызывают огромный интерес. Полилинейная алгебра, тензорный анализ и теория тензорных аппроксимаций играют всё более важную роль в вычислительной математике и численном анализе [63; 64]. Эффективное представление тензора (под тензором мы понимаем массив комплексных чисел с N индексами) по небольшому числу параметров может дать возможность работать с N -мерными задачами, где N достигает 10, 100 и даже 1000 (такие задачи возникают в квантовой молекулярной динамике [65], стохастических уравнениях в частных производных [66] и финансовом моделировании [67]). Задачи таких размеров не могут быть решены стандартными численными методами из-за экспоненциального роста ресурсов (память, количество операций) с ростом размерности N .

Существует эффективный способ представления большого класса важных N -мерных тензоров с помощью *канонического разложения* данного тензора A с элементами $A(i_1, \dots, i_N)$ [68]:

$$A_{i_1, \dots, i_N} = \sum_{\alpha=1}^r A_{i_1 \alpha}^{(1)} \dots A_{i_N \alpha}^{(N)}. \quad (4.5)$$

Минимальное число слагаемых r , необходимое для записи тензора A в форме (4.5), называется *тензорным рангом* (или *каноническим рангом*). Матрицы $A^{(k)} = A_{i_k \alpha}^{(k)}$ называются *каноническими факторами*. При больших N тензор A никогда не формируется явно, а представляется в некотором малопараметрическом формате. Каноническое разложение (4.5) — хороший кандидат на такой формат. Однако он страдает от нескольких недостатков. Вычисление канонического ранга является NP-трудной задачей [69], и аппроксимация с фиксированным каноническим рангом в норме Фробениуса может быть некорректной задачей [70]. Таким образом, численные алгоритмы для вычисления приближенного представления в таких случаях могут дать сбой.

Легко заметить, что каноническое разложение тензора A (4.5) имеет такой же вид, как сингулярное разложение матрицы, записанное в тензорном виде (4.4). Но выражение (4.4) можно также переписать в немного другом виде, транспонировав вторую матрицу и добавив два формальных вспомогательных индекса k_0 и k_2 , которые принимают всего одно значение, и суммирование по ним. В итоге получаем следующее представление:

$$A_{i_1 i_2} = \sum_{k_0=1}^{r_0} \sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} A_{k_0 i_1 k_1}^{(1)} A_{k_1 i_2 k_2}^{(2)}, \quad (4.6)$$

где $r_0 = r_2 = 1$, $r_1 = r$. Записанное таким образом разложение матрицы A можно обобщить на многомерный случай для тензора $A_{i_1, \dots, i_N}$:

$$A_{i_1 \dots i_N} = \sum_{k_0=1}^{r_0} \sum_{k_1=1}^{r_1} \dots \sum_{k_N=1}^{r_N} A_{k_0 i_1 k_1}^{(1)} A_{k_1 i_2 k_2}^{(2)} \dots A_{k_{N-1} i_N k_N}^{(N)}, \quad (4.7)$$

где $r_0 = r_N = 1$. Оказывается, разложение (4.7) существует для любого тензора $A_{i_1 \dots i_N}$ с комплексными элементами. Такое разложение называют разложением тензора в *тензорный поезд* (Tensor Train decomposition, TT-decomposition), а такой формат представления тензора называют тензорным поездом (Tensor Train, TT), дальше мы будем называть его TT-форматом. Числа $r_0, \dots, r_N$ — называются рангами тензорного поезда, или TT-рангами. Тензоры 3-го ранга (то есть обладающие тремя индексами) $A_{k_{n-1} i_n k_n}^{(n)}$ называются ядрами тензорного поезда или вагонами тензорного поезда (TT-cores, TT-ядра). Графическое представление тензора $A_{i_1 \dots i_N}$ в TT-формате показано на рисунке 4.1.

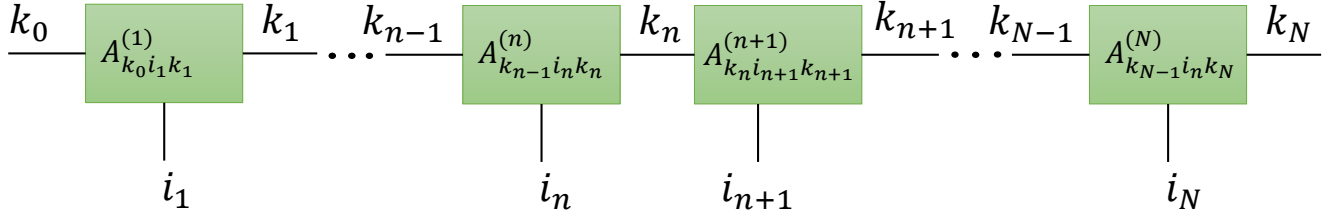


Рисунок 4.1 — Графическое представление разложения тензора $A_{i_1 \dots i_N}$ в тензорный поезд (4.7).

4.1.3 Алгоритм TT-SVD

Прежде чем мы сформулируем теорему о представлении произвольного тензора $A_{i_1 \dots i_N}$ в TT-формате, определим понятие *матрицы развёртки* тензора (unfolding matrix). Для удобства записи будем иногда писать индексы тензора не снизу, а в скобках, то есть $A_{i_1 \dots i_N} \equiv A(i_1, \dots, i_N)$. Пусть тензор $A_{i_1 \dots i_N}$ имеет размеры $n_1 \times n_2 \times \dots \times n_N$. k -ой матрицей развёртки A_k тензора A называется матрица размера $n_1 \dots n_k \times n_{k+1} \dots n_N$, которая записывается следующим образом:

$$A_k = A_k(i_1, \dots, i_k; i_{k+1}, \dots, i_N) = A(i_1, \dots, i_N). \quad (4.8)$$

Таким образом, в матрице A_k индексы $i_1, \dots, i_k$ нумеруют строки матрицы, а индексы $i_{k+1}, \dots, i_N$ — столбцы (в лексикографическом порядке). То есть матрица A_k имеет два индекса: индекс строк $i = (i_1 - 1)(n_2 n_3 \dots n_k) + (i_2 - 1)(n_3 \dots n_k) + \dots + (i_{k-1} - 1)n_k + i_k$ и индекс столбцов $j = (i_{k+1} - 1)(n_{k+2} n_{k+3} \dots n_N) + (i_{k+2} - 1)(n_{k+3} \dots n_N) + \dots + (i_{N-1} - 1)n_N + i_N$ (предполагается, что нумерация индексов начинается с 1). В большинстве фреймворков для научной работы (NumPy, PyTorch, MATLAB) данная операция представлена командой `reshape`:

$$A_k = \text{reshape}(A, \prod_{s=1}^k n_s, \prod_{s=k+1}^N n_s). \quad (4.9)$$

Сформулируем теорему о разложении произвольного тензора $A_{i_1 \dots i_N}$ в тензорный поезд [71]:

Теорема 1 (Оселедец). Пусть дан произвольный тензор $A_{i_1 \dots i_N}$ с комплексными элементами и его матрицы развёртки A_k . Тогда существует TT -разложение вида (4.7) с TT -рангами, удовлетворяющими неравенствам $r_k \leq \text{rank } A_k$, $k = 1, \dots, N-1$.

Приведём описание алгоритма TT -SVD для точного представления тензора $A_{i_1 \dots i_N}$ в TT -формате. Графическое представление данного алгоритма приведено на рисунке 4.2.

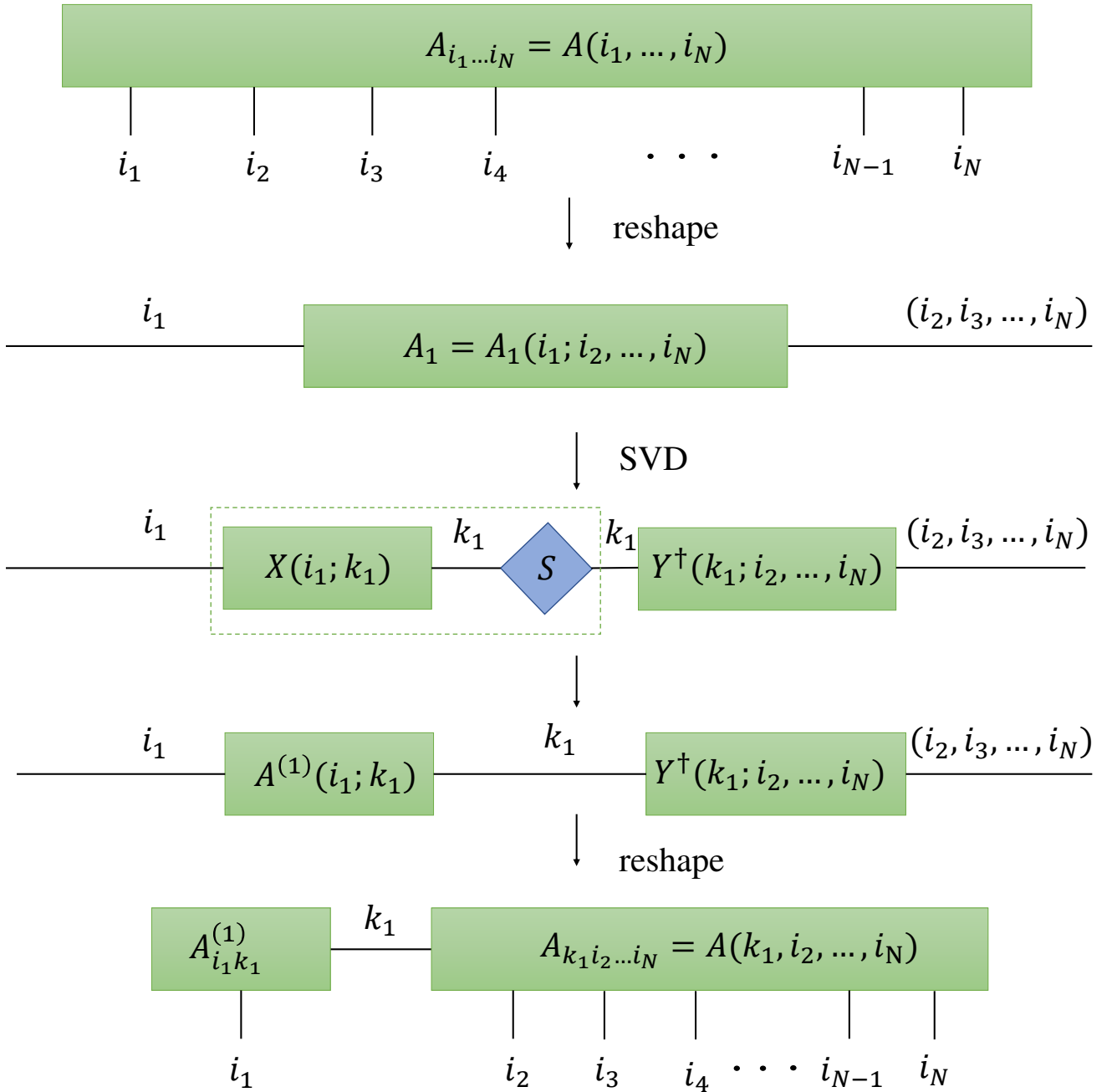


Рисунок 4.2 — Графическое описание первой итерации алгоритма TT -SVD для разложения комплексного тензора $A_{i_1 \dots i_N}$ в тензорный поезд.

Итак, пусть имеется тензор $A_{i_1 \dots i_N} = A(i_1, \dots, i_N)$. Опишем первую итерацию алгоритма TT-SVD. Сначала вычисляем матрицу развёртки тензора $A_1(i_1; i_2, \dots, i_N)$ по формуле (4.9). Затем производим SVD разложение полученной матрицы развёртки:

$$A_1(i_1; i_2, \dots, i_N) = \sum_{k_1=1}^{r_1} X(i_1; k_1) s_{k_1} Y^\dagger(k_1; i_2, \dots, i_N). \quad (4.10)$$

Затем перемножим матрицу $X(i_1; k_1)$ и диагональную матрицу S с элементами s_{k_1} на диагонали и получаем новую матрицу $A^{(1)}(i_1; k_1)$:

$$A^{(1)}(i_1; k_1) = X(i_1; k_1) s_{k_1}. \quad (4.11)$$

Введём переобозначение для матрицы $Y^\dagger(k_1; i_2, \dots, i_N) \equiv A(k_1; i_2, \dots, i_N)$. Тогда выражение (4.10) можно переписать в таком виде:

$$A_1(i_1; i_2, \dots, i_N) = \sum_{k_1=1}^{r_1} A^{(1)}(i_1; k_1) A(k_1; i_2, \dots, i_N). \quad (4.12)$$

Далее мы преобразуем матрицу развёртки $A_1(i_1; i_2, \dots, i_N)$ обратно в исходный тензор $A_{i_1, \dots, i_N}$:

$$A_{i_1, \dots, i_N} = \sum_{k_1=1}^{r_1} A_{i_1 k_1}^{(1)} A_{k_1 i_2 \dots i_N}. \quad (4.13)$$

Таким образом, мы убрали индекс i_1 у исходного тензора A , то есть разделили переменные. Далее аналогичным образом повторяем эту процедуру ещё $N - 2$ (на второй итерации применяем эту процедуру к тензору $A_{k_1 i_2 \dots i_N}$). В конце добавим к этому тензору два дополнительных индекса k_0 и k_N , которые принимают одно значение $k_0 = k_N = 1$).

Здесь мы рассмотрели алгоритм TT-SVD для точного представления заданного тензора в TT-формате. В дополнении к этому, в работе [71] описывается также алгоритм TT-SVD для разложения заданного тензора в тензорный поезд с заданной точностью (это позволяет уменьшить TT-ранги получаемого тензорного поезда).

4.2 Обзор литературы по тензорным сетям

В данном разделе мы сделаем краткий обзор основных архитектур тензорных сетей и алгоритмов работы с ними. Графическое изображение данных архитектур дано на рисунке 4.3.

Самая распространённая архитектура тензорной сети называется тензорным поездом (4.7), в применении к квантовым состояниям эту архитектуру ещё называют matrix product state (MPS) [72; 73] (рисунок 4.3а). Тензорные поезда используют для анализа квантовых систем в виде одномерных цепочек. В формате MPS обычно ищут основное состояние одномерной квантовой системы, у которой имеется взаимодействие только между соседними кубитами. Для этого, согласно вариационному принципу квантовой механики, достаточно решить оптимизационную задачу $\min \langle \psi | \hat{H} | \psi \rangle$, при условии нормировки $\langle \psi | \psi \rangle = 1$. Эту оптимизационную задачу можно решить при помощи алгоритма ренормгруппы матрицы плотности (density matrix renormalization group, DMRG) [72; 74].

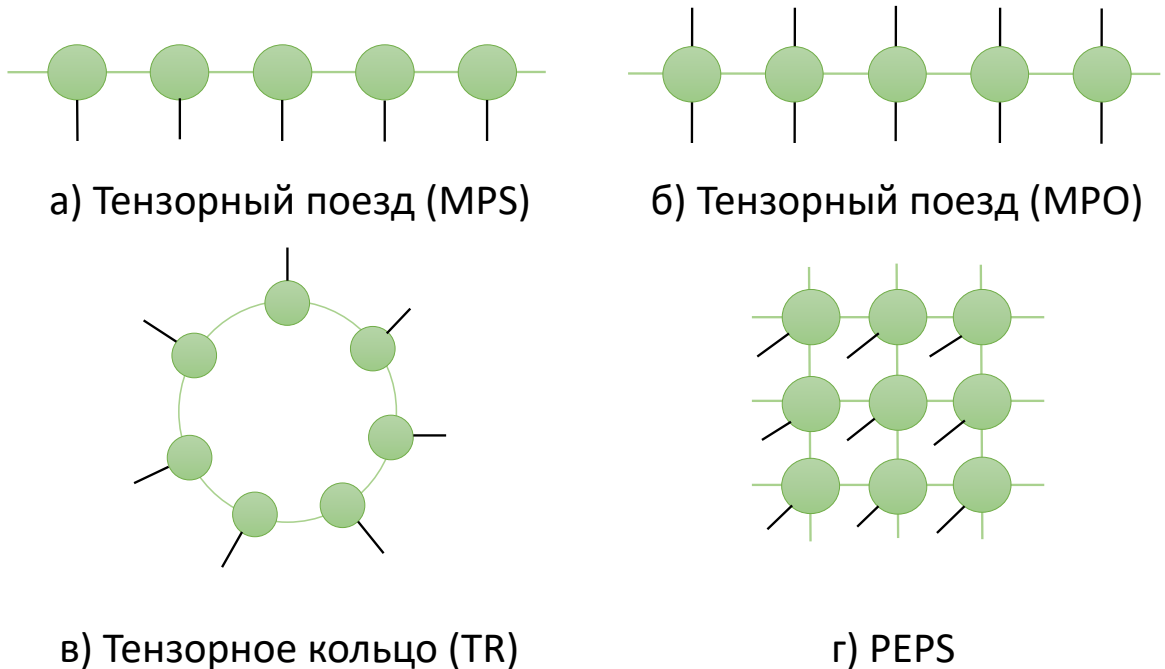


Рисунок 4.3 — Основные архитектуры тензорных сетей.

При расчёте эволюции одномерных цепочек кубитов во времени ТТ-ранги тензорного поезда растут экспоненциально быстро [75]. При помощи

алгоритма разрежения блоков во времени (time-evolving block decimation, TEBD) [75] можно контролировать максимальные ТТ-ранги.

Формат MPS используется для описания квантовых состояний N кубитов, которые описываются тензором с N индексами. Тензор, который описывает матрицу плотности или оператор (например, Гамильтониан), соответствующие такому гильбертову пространству, будет иметь $2N$ индексов. Для их описания подходит формат matrix product operator (MPO) [72] (рисунок 4.3б).

Квантовое машинное обучение (quantum machine learning, QML) представляет собой направление исследований, изучающее применение квантовых вычислений в задачах машинного обучения. Одним из основных инструментов в этой области являются параметризованные вариационные квантовые схемы (variational quantum circuits, VQC). Однако их обучение на классических вычислительных устройствах затруднено тем, что при точной симуляции квантовой схемы затраты памяти и времени, как правило, растут экспоненциально с увеличением числа кубитов и глубины схемы. Для уменьшения этих затрат в работе [76] предложен приближённый метод классического моделирования VQC, основанный на представлении квантового состояния в формате тензорного кольца (tensor ring, TR), структура которого изображена на рисунке 4.3в. Такое представление позволяет хранить состояние в сжатом виде: однокубитные операции сохраняют TR-структуру, а двухкубитные операции после применения аппроксимируются с помощью усечённого сингулярного разложения. В результате при фиксированном ранге тензорного кольца сложность хранения и вычислений существенно уменьшается по сравнению с точной симуляцией, что делает возможным обучение более крупных VQC на классическом компьютере. Эффективность подхода продемонстрирована авторами на задачах классификации для наборов данных Iris и MNIST.

Исторический успех MPS был связан не с точными расчетами, а, напротив, с развитием контролируемых приближённых методов для решения задач квантовой физики многих тел. Обобщениями MPS для более высоких измерений являются проецируемые состояния запутанной пары (projected entangled pair states, PEPS) [73; 77] (рисунок 4.3г) или многомасштабный анзац ренормализации запутанности (multi-scale entanglement renormalization ansatz, MERA) [78; 79]. В основе этих успехов лежит тот факт, что MPS

естественным образом организует состояния в соответствии с величиной энтропии запутанности между различными частями системы. Следовательно, слегка запутанные системы могут быть легко представлены с помощью MPS.

Для расчёта непрерывной динамики квантовых состояний может быть применен симулятор на основе тензорных сетей, основанный на том, что дискретизация уравнения Шрёдингера на временной сетке сводится к решению системы линейных уравнений, записанных в формате тензорных поездов [80; 81]. Для решения таких систем линейных уравнений разработаны специальные алгоритмы — AMEn и tAMEn [82], которые доступны в открытом коде библиотеки `ttpy` [83].

Сделаем краткий обзор основных библиотек, которые применяются в области моделирования квантовых схем с помощью тензорных сетей.

ITensor [84] — это система для программирования расчётов тензорных сетей с интерфейсом, смоделированным на основе нотации тензорных диаграмм, что позволяет пользователям сосредоточиться на связности тензорной сети без ручного учёта тензорных индексов. Интерфейс ITensor исключает распространённые ошибки программирования и позволяет быстро создавать прототипы алгоритмов тензорных сетей. Она включает в себя тензорные факторизации, алгоритмы для MPS и MPO. Эта библиотека используется в работе [85].

Библиотека TensorNetwork [86–88] написана на языке программирования Python и представляет собой обёртку для TensorFlow, JAX, PyTorch, и Numpy.

Библиотека `ttpy` [83] является python-реализацией Tensor Train. Она содержит несколько важных пакетов для работы с ТТ-форматом. Она может выполнять ТТ-округление, решать линейные системы, задачи на собственные значения и динамические задачи. Несколько вычислительных процедур выполняются на Фортране. Библиотека TT-toolbox [89] представляет похожий набор пакетов, написанных на MatLab.

Одной из задач, решаемых в рамках данной диссертационной работы, являлась разработка библиотеки QTensor [90]. Необходимость её создания была связана с тем, что для рассматриваемых в работе задач требовался единый инструмент, объединяющий представление квантовых состояний в формате тензорных поездов, моделирование квантовых схем и градиентную

оптимизацию параметров модели. Библиотека была разработана автором диссертации на основе фреймворка PyTorch, что позволяет использовать автоматическое дифференцирование и выполнять основные тензорные операции на GPU.

4.3 Представление квантовых состояний в формате Tensor Train

4.3.1 Разложение Шмидта

Рассмотрим для начала квантовую систему, состоящую из двух подсистем A и B , пусть размерности этих квантовых систем равны d_A и d_B соответственно. Согласно принципу суперпозиции квантовой механики, произвольное состояние $|\psi\rangle$ такой системы можно разложить по базисным состояниям такой системы, которые являются всевозможными попарными тензорными произведениями базисных состояний каждой из подсистем (для краткости здесь будем обозначать $|a\rangle \otimes |b\rangle \equiv |a\rangle|b\rangle$):

$$|\psi\rangle = \sum_{a=1}^{d_A} \sum_{b=1}^{d_B} \Psi_{ab} |a\rangle|b\rangle, \quad (4.14)$$

где Ψ_{ab} — матрица размера $d_A \times d_B$, а $\{|a\rangle\}$ и $\{|b\rangle\}$ — ортонормированные базисы в гильбертовых пространствах состояний квантовых систем A и B соответственно.

Зная матрицу Ψ с элементами Ψ_{ab} можно вычислить операторы плотности подсистем A и B по формулам $\hat{\rho}_A = \text{Tr}_B |\psi\rangle\langle\psi|$ и $\hat{\rho}_B = \text{Tr}_A |\psi\rangle\langle\psi|$ соответственно. Распишем выражение для оператора плотности всей системы $\hat{\rho}$:

$$\hat{\rho} = |\psi\rangle\langle\psi| = \sum_{a,a'=1}^{d_A} \sum_{b,b'=1}^{d_B} \Psi_{ab} \Psi_{a'b'}^* |a\rangle\langle a'| \otimes |b\rangle\langle b'|. \quad (4.15)$$

Далее вычисляем оператор плотности для подсистемы A :

$$\begin{aligned}\hat{\rho}_A &= \text{Tr}_B |\psi\rangle\langle\psi| = \sum_{a,a'=1}^{d_A} \sum_{b,b'=1}^{d_B} \Psi_{ab} \Psi_{a'b'}^* |a\rangle\langle a'| \langle b'|b\rangle = \\ &= \sum_{a,a'=1}^{d_A} \sum_{b,b'=1}^{d_B} \delta_{b'b} \Psi_{ab} \Psi_{a'b'}^* |a\rangle\langle a'| = \sum_{a,a'=1}^{d_A} \sum_{b=1}^{d_B} \Psi_{ab} \Psi_{ba'}^\dagger |a\rangle\langle a'|.\end{aligned}\quad (4.16)$$

Исходя из этого выражения получаем, что матрица плотности подсистемы A в базисе $\{|a\rangle\}$ равна $\rho_A = \Psi\Psi^\dagger$, аналогично, можно показать, что матрица плотности подсистемы B в базисе $\{|b\rangle\}$ $\rho_B = (\Psi^\dagger\Psi)^T$.

Выполним сингулярное разложение матрицы $\Psi = USV^\dagger$, или в тензорном виде:

$$\Psi_{ab} = \sum_{k=1}^r U_{ak} s_k V_{bk}^*, \quad (4.17)$$

тогда выражение (4.14) можно записать в таком виде:

$$|\psi\rangle = \sum_{a=1}^{d_A} \sum_{b=1}^{d_B} \Psi_{ab} |a\rangle |b\rangle = \sum_{k=1}^r s_k \left(\sum_{a=1}^{d_A} U_{ak} |a\rangle \right) \left(\sum_{b=1}^{d_B} V_{bk}^* |b\rangle \right), \quad (4.18)$$

далее, переходя к новым ортонормированным базисным состояниям в подсистемах A и B по формулам $|k\rangle_A = \sum_{a=1}^{d_A} U_{ak} |a\rangle$ и $|k\rangle_B = \sum_{b=1}^{d_B} V_{bk}^* |b\rangle$ соответственно, получаем *разложение Шмидта* для двухкомпонентной квантовой системы:

$$|\psi\rangle = \sum_{k=1}^r s_k |k\rangle_A |k\rangle_B. \quad (4.19)$$

Разложение Шмидта позволяет достаточно компактно записать выражение для операторов плотности подсистем A и B :

$$\hat{\rho}_A = \sum_{k=1}^r s_k^2 |k\rangle_A \langle k|_A, \quad \hat{\rho}_B = \sum_{k=1}^r s_k^2 |k\rangle_B \langle k|_B. \quad (4.20)$$

Квантовые состояния только с одним ненулевым сингулярным значением $s_1 = 1$ являются простыми незапутанными (факторизованными) состояниями, остальные состояния являются запутанными.

4.3.2 Представление чистых квантовых состояний в ТТ-формате

Обобщим идею разложения Шмидта с двусоставных квантовых систем на многосоставные квантовые системы. Далее мы рассматриваем частный случай таких систем — системы из N кубитов. Размерность таких квантовых систем равна 2^N . Базисом для каждого кубита являются состояния $|0\rangle$ и $|1\rangle$.

Согласно *принципу суперпозиции* квантовой механики, произвольное чистое состояние $|\psi\rangle$ N -кубитной системы можно разложить по базисным состояниям системы:

$$|\psi\rangle = \sum_{i_1=0}^1 \cdots \sum_{i_N=0}^1 \psi_{i_1 \dots i_N} |i_1\rangle \otimes \cdots \otimes |i_N\rangle. \quad (4.21)$$

Возникает проблема хранения и преобразования тензора $\psi_{i_1 \dots i_N}$, который имеет 2^N элементов.

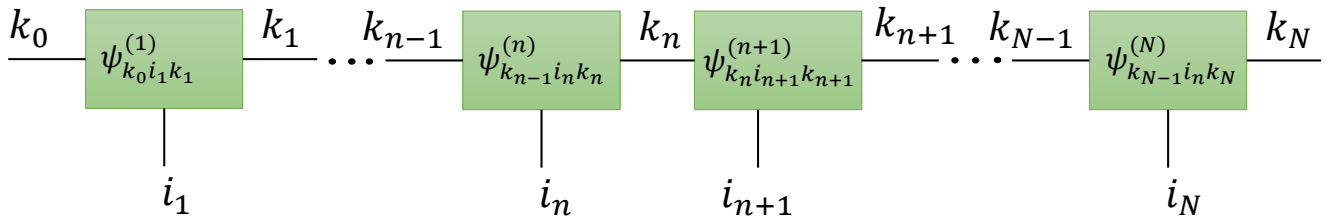


Рисунок 4.4 — Представление тензора $\psi_{i_1 \dots i_N}$, описывающего амплитуды разложения квантового состояния (4.21) по базисным состояниям вычислительного базиса N -кубитной системы в формате тензорного поезда (4.22).

Когда тензор $\psi_{i_1 \dots i_N}$ квантового состояния имеет низкоранговые матрицы развёртки, эффективно представление тензора $\psi_{i_1 \dots i_N}$ в формате тензорного поезда (4.7) (рисунок 4.4):

$$\psi_{i_1 \dots i_N} = \sum_{k_0=1}^{r_0} \sum_{k_1=1}^{r_1} \cdots \sum_{k_N=1}^{r_N} \psi_{k_0 i_1 k_1}^{(1)} \cdots \psi_{k_{N-1} i_N k_N}^{(N)}. \quad (4.22)$$

Разложение тензора $\psi_{i_1 \dots i_N}$ в тензорный поезд можно также записать в матричном виде, поэтому в литературе, посвящённой применению тензорных поездов к квантовым вычислениям квантовые состояния, представленные в ТТ-формате иногда называют состояниями произведения

матриц (matrix product state, MPS):

$$\psi_{i_1 \dots i_N} = \Psi^{(1)}(i_1) \dots \Psi^{(N)}(i_N), \quad (4.23)$$

где $\Psi^{(n)}(i_n)$ — матрицы размером $r_{n-1} \times r_n$, $\Psi_{k_{n-1}k_n}^{(n)}(i_n) = \psi_{k_{n-1}i_nk_n}^{(n)}$, которые зависят от индексов i_n . То есть при каждом фиксированном наборе значений индексов $i_1, \dots, i_N$ мы имеем произведение матриц таких форм, что в результате их перемножения мы получаем число, которое зависит от индексов $i_1, \dots, i_N$. В итоге, мы получаем исходный тензор $\psi_{i_1 \dots i_N}$ (рисунок 4.5). Индексы $i_1, \dots, i_N$ называют также физическими индексами, так как они нумеруют амплитуды вероятностей состояния в определённом базисе.

Рисунок 4.5 — Представление квантового состояния в формате MPS для пяти кубитов. При каждом фиксированном наборе значений индексов $i_1, \dots, i_N$ мы имеем совокупность матриц, произведение которых даёт число, зависящее от индексов $i_1, \dots, i_N$.

4.3.3 Представление оператора плотности квантового состояния в ТТ-формате

В общем случае согласно принципу суперпозиции смешанное состояние $\hat{\rho}$ N -кубитного регистра можно представить в таком виде:

$$\hat{\rho} = \sum_{\substack{i_1, \dots, i_N=0, \\ j_1, \dots, j_N=0}}^1 \rho_{i_1 \dots i_N j_1 \dots j_N} |i_1\rangle \langle j_1| \otimes \dots \otimes |i_N\rangle \langle j_N|, \quad (4.24)$$

где $\rho_{i_1 \dots i_N j_1 \dots j_N}$ — $2N$ -мерный тензор. Для $\rho_{i_1 \dots i_N j_1 \dots j_N}$ разложение в тензорный поезд можно записать в следующем виде:

$$\rho_{i_1 \dots i_N j_1 \dots j_N} = \sum_{k_0, \dots, k_N=1}^{r_0, \dots, r_N} \rho_{k_0 i_1 j_1 k_1}^{(1)} \dots \rho_{k_{N-1} i_N j_N k_N}^{(N)}. \quad (4.25)$$

Это представление также известно как форма matrix product operator (MPO) [91].

4.3.4 Примеры представления состояний

Основная идея использования ТТ-формата для представления квантовых состояний заключается в том, что мы можем хранить в памяти и обрабатывать состояния систем большой размерности. В разделе 4.1.3 мы привели алгоритм ТТ-SVD, который позволяет произвольное состояние представить в ТТ-формате. Однако, для этого нам нужно изначально хранить в памяти большой объём данных — 2^N комплексных чисел, которые описывают амплитуды квантового состояния. Поэтому большой интерес представляют квантовые состояния, которые можно сразу аналитически записать в компактном ТТ-формате.

Приведём конкретные примеры низкоранговых разложений некоторых квантовых состояний — факторизованного состояния и N -кубитного состояния Гринбергера-Хорна-Цайлингера $|\psi\rangle_{\text{GHZ}}$.

Запишем сначала в ТТ-формате состояние $|0\rangle^{\otimes N}$. Состояние одного кубита $|0\rangle$ может быть представлена всего лишь одним вагоном $\psi_{k_0 i_1 k_1}^{(1)}$ — тензором 3-го ранга размером $1 \times 2 \times 1$ с компонентами $\psi_{101}^{(1)} = 1$, $\psi_{111}^{(1)} = 0$. Поскольку общее состояние N -кубитной системы представляет собой тензорное произведение N таких состояний $|0\rangle^{\otimes N} = |0\rangle \otimes \dots \otimes |0\rangle$, то представление такого факторизованного квантового состояния в ТТ-формате можно записать в виде N вагонов $\psi_{k_{n-1} i_n k_n}^{(n)}$ — тензоров 3-го ранга с такими же размерами $1 \times 2 \times 1$ и такими же компонентами $\psi_{101}^{(n)} = 1$, $\psi_{111}^{(n)} = 0$.

Рассмотрим теперь N -кубитное состояние Гринбергера-Хорна-Цайлингера (ГХЦ) $|\psi\rangle_{\text{GHZ}}$:

$$\begin{aligned} |\psi\rangle_{\text{GHZ}} &= \frac{|0\rangle^{\otimes N} + |1\rangle^{\otimes N}}{\sqrt{2}} = \\ &= \frac{1}{\sqrt{2}} \left(|0\rangle \quad |1\rangle \right) \otimes \begin{pmatrix} |0\rangle & 0 \\ 0 & |1\rangle \end{pmatrix} \otimes \dots \otimes \begin{pmatrix} |0\rangle & 0 \\ 0 & |1\rangle \end{pmatrix} \otimes \begin{pmatrix} |0\rangle \\ |1\rangle \end{pmatrix}. \end{aligned} \quad (4.26)$$

Данная нотация весьма удобная, так как каждая из матриц является по сути вагоном искомого тензорного поезда. Действительно, вагон с номером n представляет собой тензор 3-го ранга, то есть обладающий тремя индексами (k_{n-1}, i_n, k_n) и формой $(r_{n-1}, 2, r_n)$, где 2 - размерность гильбертова пространства состояний одного n -го кубита. Состояние ГХЦ (4.26) записан в виде формата MPS (4.23) (рисунок 4.5), так как представляет собой произведение матриц, а эти матрицы содержат объекты $(|0\rangle, |1\rangle, 0)$, которые зависят от одного физического индекса, пробегающего значения от 0 до 1 ($|0\rangle_0 = 1, |0\rangle_1 = 0, |1\rangle_0 = 0, |1\rangle_1 = 1$).

Таким образом, известные простые N -кубитные квантовые состояния могут быть записаны в достаточно простой форме в ТТ-формате и, что важно, с низкими ТТ-рангами — для факторизованного состояния $|0\rangle^{\otimes N}$ все ТТ-ранги равны 1, а для N -кубитного состояния $|\psi\rangle_{\text{GHZ}}$ $r_0 = r_N = 1$, а все остальные ТТ-ранги равны 2.

4.4 Преобразование квантовых состояний в ТТ-формате

Для полноценной симуляции квантовых схем нам необходимо уметь не только эффективно хранить квантовые состояния в формате тензорных поездов, но и преобразовывать их с помощью различных унитарных квантовых гейтов. В настоящей диссертации рассматривается только применение однокубитных и двухкубитных гейтов. При этом двухкубитные гейты наиболее эффективно применять к соседним кубитам. Действие на несоседние кубиты может быть реализовано с помощью последовательности перестановок (SWAP), что, однако, приводит к увеличению вычислительных затрат и может вызывать дополнительный рост ТТ-рангов.

4.4.1 Применение однокубитных гейтов

Рассмотрим сначала процедуру изменения вагонов тензорного поезда, описывающего квантовое состояние N кубитов под действием однокубит-

ного гейта. Начнём с разложения квантового состояния $|\psi\rangle$ по базисным состояниям (4.21), только для большей общности будем рассматривать не кубиты, а произвольные квантовые подсистемы, которые имеют размерности $d_1, \dots, d_N$. Тогда разложение (4.21) для квантового состояния $|\psi\rangle$ будет выглядеть следующим образом:

$$|\psi\rangle = \sum_{i_1=1}^{d_1} \cdots \sum_{i_N=1}^{d_N} \psi_{i_1 \dots i_N} |i_1\rangle \otimes \cdots \otimes |i_N\rangle, \quad (4.27)$$

а разложение тензора $\psi_{i_1 \dots i_N}$ сохраняет вид (4.22), за исключением того, что теперь физические индексы $i_1, \dots, i_N$ пробегают значения не от 0 до 1, а от 1 до $d_1, \dots, d_N$ соответственно:

$$\psi_{i_1 \dots i_N} = \sum_{k_0=1}^{r_0} \sum_{k_1=1}^{r_1} \cdots \sum_{k_N=1}^{r_N} \psi_{k_0 i_1 k_1}^{(1)} \cdots \psi_{k_{N-1} i_N k_N}^{(N)}. \quad (4.28)$$

Рассмотрим действие оператора $U^{(n)}$ на состояние $|\psi\rangle$:

$$\hat{U}^{(n)}|\psi\rangle = \sum_{i_1=1}^{d_1} \cdots \sum_{i_n=1}^{d_n} \cdots \sum_{i_N=1}^{d_N} \psi_{i_1 \dots i_N} |i_1\rangle \otimes \cdots \otimes \hat{U}^{(n)}|i_n\rangle \otimes \cdots \otimes |i_N\rangle, \quad (4.29)$$

где действие оператора $\hat{U}^{(n)}$ на базисное состояние $|i_n\rangle$ описывается тензором второго ранга U_{ji_n} (то есть матрицей, размера $d_n \times d_n$):

$$\hat{U}^{(n)}|i_n\rangle = \sum_{j=1}^{d_n} U_{ji_n}^{(n)} |j\rangle. \quad (4.30)$$

Подставим теперь выражение действия оператора $\hat{U}^{(n)}$ на базисное состояние $|i_n\rangle$ (4.30) в выражение (4.29) для действия оператора $\hat{U}^{(n)}$ на квантовое состояние $|\psi\rangle$:

$$\begin{aligned} \hat{U}^{(n)}|\psi\rangle &= \\ &= \sum_{i_1=1}^{d_1} \cdots \sum_{i_n=1}^{d_n} \cdots \sum_{i_N=1}^{d_N} \psi_{i_1 \dots i_n \dots i_N} |i_1\rangle \otimes \cdots \otimes \sum_{j=1}^{d_n} U_{ji_n}^{(n)} |j\rangle \otimes \cdots \otimes |i_N\rangle = \\ &= \sum_{i_1=1}^{d_1} \cdots \sum_{i_n=1}^{d_n} \cdots \sum_{i_N=1}^{d_N} \sum_{j=1}^{d_n} U_{ji_n}^{(n)} \psi_{i_1 \dots i_n \dots i_N} |i_1\rangle \otimes \cdots \otimes |j\rangle \otimes \cdots \otimes |i_N\rangle = \\ &= \{\text{меняем обозначение } i_n \leftrightarrow j \text{ и переставляем суммирование по } i_n \text{ и } j\} = \\ &= \sum_{i_1=1}^{d_1} \cdots \sum_{i_n=1}^{d_n} \cdots \sum_{i_N=1}^{d_N} \left(\sum_{j=1}^{d_n} U_{i_n j}^{(n)} \psi_{i_1 \dots j \dots i_N} \right) |i_1\rangle \otimes \cdots \otimes |i_n\rangle \otimes \cdots \otimes |i_N\rangle. \end{aligned} \quad (4.31)$$

Итак, мы получаем, что под действием однокубитного гейта, описываемого оператором $\hat{U}^{(n)}$, исходный тензор $\psi_{i_1 \dots i_N}$ амплитуд состояния $|\psi\rangle$ меняется следующим образом:

$$\psi'_{i_1 \dots i_n \dots i_N} = \sum_{j=1}^{d_n} U_{i_n j}^{(n)} \psi_{i_1 \dots j \dots i_N}. \quad (4.32)$$

Учитывая разложение (4.28) тензора $\psi_{i_1 \dots i_n \dots i_N}$ в тензорный поезд, легко понять, что для выполнения преобразования (4.32) достаточно осуществить изменение n -го вагона тензорного поезда следующим образом (остальные вагоны остаются при этом неизменными):

$$\psi'^{(n)}_{k_{n-1} i'_n k_n} = \sum_{i_n=1}^{d_n} U_{i'_n i_n}^{(n)} \psi^{(n)}_{k_{n-1} i_n k_n}. \quad (4.33)$$

Проанализируем выражение (4.33) для преобразование n -го вагона. Видно, что для изменения вагона тензорного поезда нам достаточно сделать свёртку матрицы оператора однокубитного гейта и вагона по физическому индексу (рисунок 4.6), а, значит, данное преобразование не затрагивает индексы k_{n-1} и k_n , по которым производится свёртка с другими вагонами поезда. Следовательно, поскольку остальные вагоны тензорного поезда остаются неизменными, при применении однокубитного гейта ТТ-ранги тензорного поезда не меняются. Значит, можно применять однокубитные гейты к тензорным поездкам точно и без роста ТТ-рангов.

4.4.2 Применение двухкубитных гейтов

Рассмотрим теперь, как изменяются вагоны тензорного поезда, описывающего N -кубитное квантовое состояние после действия двухкубитного гейта на соседние кубиты с номерами n и $n+1$, который описывается унитарным оператором $\hat{U}^{(n, n+1)}$. Аналогично пункту про однокубитные гейты (см. раздел 4.4.1), сначала нам нужно узнать, как меняется тензор $\psi_{i_1 \dots i_N}$ под действием этого двухкубитного оператора. Проводя цепочку рассуждений, аналогичную представленной формулами (4.29), (4.30) и (4.31), легко показать, что этот тензор меняется следующим образом (причём данный

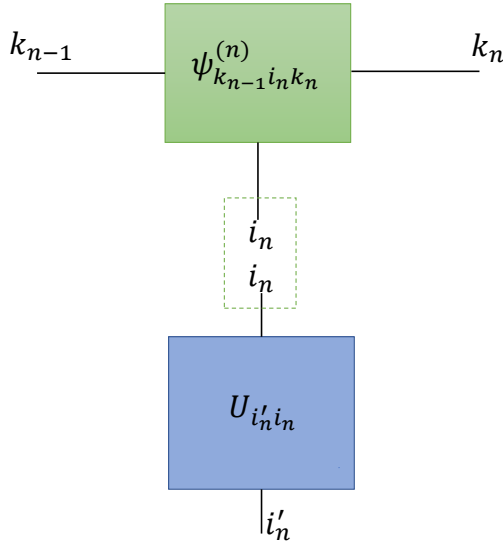


Рисунок 4.6 — Процедура применения однокубитного гейта затрагивает только один вагон тензорного поезда — с тем же номером, как у кубита, к которому этот гейт применяется. При этом все ТТ-ранги остаются теми же.

результат справедлив безотносительно того, применяется ли двухкубитный гейт к соседним кубитам или к произвольным кубитам с номерами n и m):

$$\psi'_{i_1 \dots i_n \dots i_m \dots i_N} = \sum_{j=1}^{d_n} \sum_{k=1}^{d_m} U_{i_n i_m j k}^{(n,m)} \psi_{i_1 \dots j \dots k \dots i_N}, \quad (4.34)$$

где тензор 4-го ранга $U_{i_n i_m j k}^{(n,m)}$ задаёт действие оператора $\hat{U}^{(n,m)}$ на базисные состояния $|i_n\rangle \otimes |i_m\rangle$:

$$\hat{U}^{(n,m)}(|i_n\rangle \otimes |i_m\rangle) = \sum_{j=1}^{d_n} \sum_{k=1}^{d_m} U_{j k i_n i_m}^{(n,m)} |j\rangle \otimes |k\rangle. \quad (4.35)$$

Для дальнейшего рассмотрения действия двухкубитного гейта принципиально важно, чтобы кубиты, для которых мы применяем этот гейт, были соседними. Обозначим их номера n и $n+1$. В разложении тензора $\psi_{i_1 \dots i_N}$ в тензорный поезд (4.28) произведём свёртку по индексу k_n , и из двух соседних вагонов, которые представляют собой тензоры 3-го ранга, мы получаем один тензор 4-го ранга:

$$\Phi_{k_{n-1}i_n i_{n+1}k_{n+1}}^{(n,n+1)} = \sum_{k_n=1}^{r_n} \psi_{k_{n-1}i_n k_n}^{(n)} \psi_{k_n i_{n+1} k_{n+1}}^{(n+1)}. \quad (4.36)$$

Учитывая разложение (4.28) тензора $\psi_{i_1 \dots i_N}$ в тензорный поезд, легко понять, что для выполнения преобразования (4.34) достаточно осуществить изменение свёртки $\Phi_{k_{n-1}i_n i_{n+1}k_{n+1}}^{(n,n+1)}$ n -го и $n+1$ -го вагонов тензорного поезда следующим образом (остальные вагоны тензорного поезда остаются при этом неизменными):

$$\Phi_{k_{n-1}i'_n i'_{n+1}k_{n+1}}'^{(n,n+1)} = \sum_{i_n=1}^{d_n} \sum_{i_{n+1}=1}^{d_{n+1}} U_{i'_n i'_{n+1} i_n i_{n+1}}^{(n,n+1)} \Phi_{k_{n-1}i_n i_{n+1}k_{n+1}}^{(n,n+1)}. \quad (4.37)$$

Проделанные операции свёртки соседних вагонов и применение двухкубитного гейта, который описывается тензором 4-го ранга $U_{i'_n i'_{n+1} i_n i_{n+1}}^{(n,n+1)}$ можно изобразить графически, используя ранее введённую нотацию для тензорных поездов (рисунок 4.7).

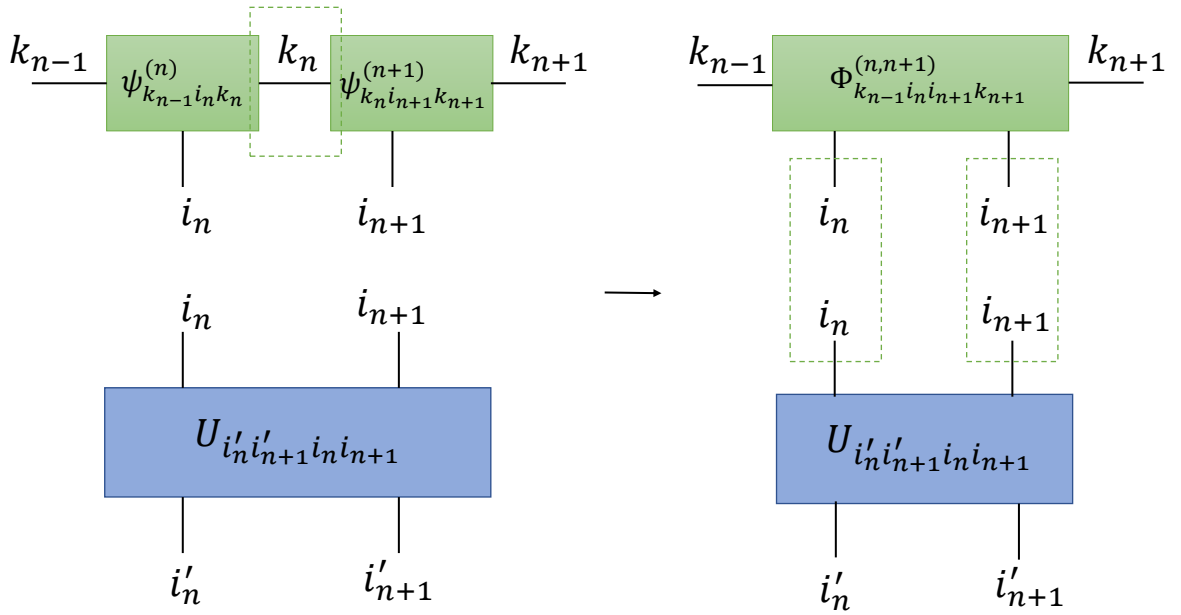


Рисунок 4.7 — Процедура применения двухкубитного гейта к соседним гейтам затрагивает только соседние вагоны тензорного поезда с теми же номерами, как у кубитов, к которым этот гейт применяется. При этом все ТТ-ранги остаются такими же, кроме ТТ-ранга между двумя рассматриваемыми вагонами.

Итак, после применения двухкубитного гейта мы получаем исходную структуру тензорного поезда (4.28), за исключением того, что у нас вместо отдельных вагонов с номерами n и $n+1$ имеется один вагон - тензор 4-го ранга $\Phi_{k_{n-1}i_n i_{n+1}k_{n+1}}^{(n,n+1)}$. Чтобы полностью восстановить структуру

тензорного поезда необходимо произвести сингулярное разложение матрицы развёртки полученного тензора (возможно, усечённое, если мы хотим посчитать действие гейта с ограниченной точностью). С помощью функции `reshape` превращаем сначала этот тензор 4-го ранга в матрицу размером $r_{n-1}d_n \times d_{n+1}r_{n+1}$:

$$\Phi'^{(n,n+1)}(k_{n-1}, i_n, i_{n+1}, k_{n+1}) = \Phi'^{(n,n+1)}(k_{n-1}, i_n; i_{n+1}, k_{n+1}), \quad (4.38)$$

а затем производим сингулярное разложение с учётом ограничения на заданный нами максимальный ТТ-ранг r'_n между вагонами n и $n+1$:

$$\Phi'^{(n,n+1)}(k_{n-1}, i_n; i_{n+1}, k_{n+1}) = X(k_{n-1}, i_n; k'_n) s_{k'_n} Y^\dagger(k'_n; i_{n+1}, k_{n+1}), \quad (4.39)$$

затем перемножаем матрицу $X(k_{n-1}, i_n; k'_n)$ и диагональную матрицу S с элементами $s_{k'_n}$ на диагонали и получаем матрицу $X'(k_{n-1}, i_n; k'_n)$. Далее с помощью функции `reshape` превращаем обратно матрицы X' и $Y^\dagger$ в тензоры 3-го ранга:

$$\begin{aligned} \psi'^{(n)}(k_{n-1}, i_n, k'_n) &= X'(k_{n-1}, i_n; k'_n), \\ \psi'^{(n+1)}(k'_n, i_{n+1}, k_{n+1}) &= Y^\dagger(k'_n; i_{n+1}, k_{n+1}). \end{aligned} \quad (4.40)$$

Описанную процедуру выполнения усечённого сингулярного разложения и восстановления структуры тензорного поезда после применения двухкубитного гейта можно изобразить графически с учётом введённых ранее обозначений (рисунок 4.8).

Точное применение двухкубитного преобразования в общем случае может привести к увеличению ТТ-ранга между соответствующими вагонами не более чем в два раза.

4.4.3 Ортонормализация вагонов тензорного поезда

Для повышения точности вычислений перед применением двухкубитного гейта тензорный поезд приводится к смешанной канонической форме с ортогональным центром на рассматриваемом разрезе между двумя кубитами, к которым применяется данный гейт [72]. Эта процедура называется *ортонормализацией вагонов* тензорного поезда с центром между

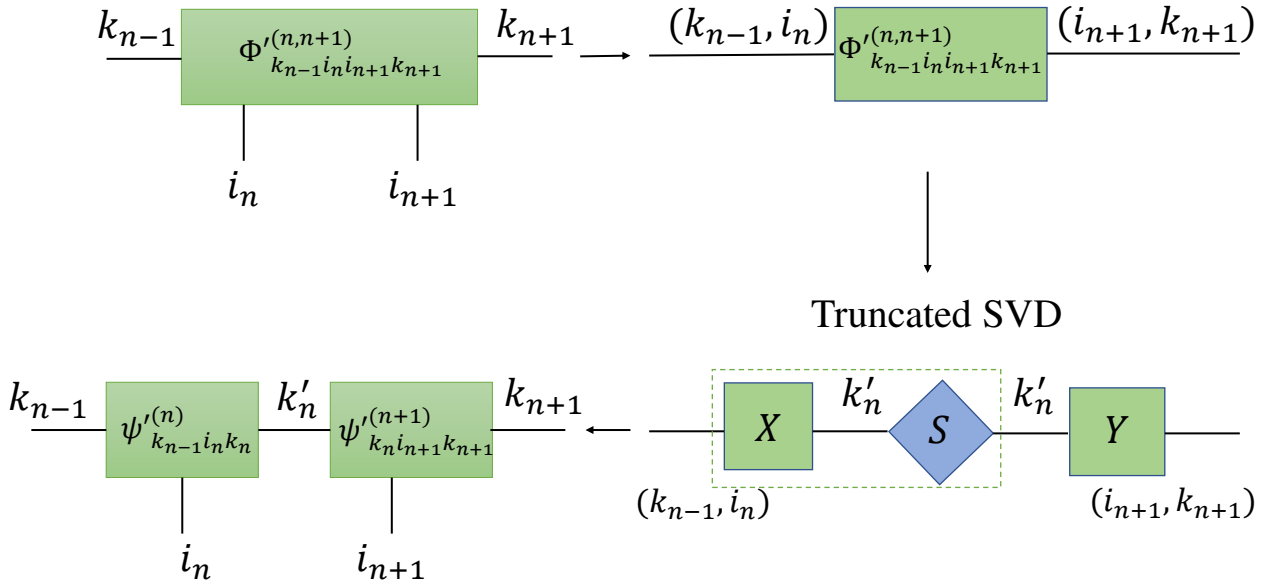


Рисунок 4.8 — Применение алгоритма Truncated SVD позволяет восстановить структуру тензорного поезда после применения оператора двухкубитного гейта, а также ограничить ТТ-ранг между кубитами, к которым применяется двухкубитный гейт.

двумя кубитами. В этой форме части тензорного поезда слева и справа от рассматриваемого разреза задают ортонормированные базисы, поэтому сингулярные значения локальной матрицы совпадают с сингулярными значениями матрицы развёртки всего состояния относительно данного разреза. Следовательно, локальное усечение ранга с помощью truncated SVD даёт лучшее приближение фиксированного ранга для данного разреза в смысле теоремы Эккарта–Янга, а ошибка усечения определяется отброшенными сингулярными значениями.

Дополнительным преимуществом ортонормализации является повышение численной устойчивости алгоритма. Поскольку ТТ-представление обладает калибровочной свободой, один и тот же тензор может быть представлен ядрами с существенно различающимися масштабами, что приводит к ухудшению обусловленности локальных задач и усилению влияния ошибок округления. Приведение к смешанной канонической форме частично устраняет эту неоднозначность и обеспечивает более сбалансированное распределение норм между вагонами тензорного поезда.

Ортонормализация вагонов с центром между двумя кубитами с номерами n и $n + 1$ означает, что все вагоны с номерами $1, \dots, n - 1$ являются *лево-ортонормализованными*, а вагоны с номерами $n + 2, \dots, N$ являются *право-ортонормализованными*.

Рассмотрим для примера процедуру левой ортонормализации вагона тензорного поезда с номером n . Аналогично тому, как мы делали при рассмотрении применения двухкубитного гейта к разложению тензора $\psi_{i_1 \dots i_N}$ в тензорный поезд (4.28), произведём свёртку по индексу k_n и из двух соседних вагонов, которые представляют собой тензоры 3-го ранга, мы получаем один тензор 4-го ранга:

$$\Phi_{k_{n-1}i_ni_{n+1}k_{n+1}}^{(n,n+1)} = \sum_{k_n=1}^{r_n} \psi_{k_{n-1}i_nk_n}^{(n)} \psi_{k_ni_{n+1}k_{n+1}}^{(n+1)}. \quad (4.41)$$

Затем превращаем операцией reshape полученный тензор 4-го ранга в матрицу размером $r_{n-1}d_n \times d_{n+1}r_{n+1}$:

$$\Phi^{(n,n+1)}(k_{n-1}, i_n, i_{n+1}, k_{n+1}) = \Phi^{(n,n+1)}(k_{n-1}, i_n; i_{n+1}, k_{n+1}), \quad (4.42)$$

а затем производим сингулярное разложение сохраняя тот же ТТ-ранг r_n между вагонами n и $n + 1$, какой он был до процедуры ортонормализации:

$$\Phi^{(n,n+1)}(k_{n-1}, i_n; i_{n+1}, k_{n+1}) = X(k_{n-1}, i_n; k'_n) s_{k'_n} Y^\dagger(k'_n; i_{n+1}, k_{n+1}), \quad (4.43)$$

затем перемножаем диагональную матрицу S с элементами $s_{k'_n}$ на диагонали на матрицу $Y^\dagger(k'_n; i_{n+1}, k_{n+1})$ и получаем матрицу $Y'(k'_n; i_{n+1}, k_{n+1})$. Дальше с помощью функции reshape превращаем обратно матрицы X и Y' в тензоры 3-го ранга:

$$\begin{aligned} \psi'^{(n)}(k_{n-1}, i_n, k'_n) &= X(k_{n-1}, i_n; k'_n), \\ \psi'^{(n+1)}(k'_n, i_{n+1}, k_n) &= Y'(k'_n; i_{n+1}, k_{n+1}). \end{aligned} \quad (4.44)$$

Обратим внимание, что в результате применения такой операции состояние квантовой системы, которое описывает данный тензорный поезд остаётся тем же самым. Более того, поскольку данная процедура ортонормализации затрагивает только вагоны с номерами n и $n + 1$ — то и свёртка двух соседних вагонов (4.41) остаётся той же самой.

Поскольку мы сделали SVD-разложение, то матрица $X(k_{n-1}, i_n; k'_n)$ обладает свойством $X^\dagger X = I_{r_n \times r_n}$ (заметим, что в общем случае $XX^\dagger \neq$

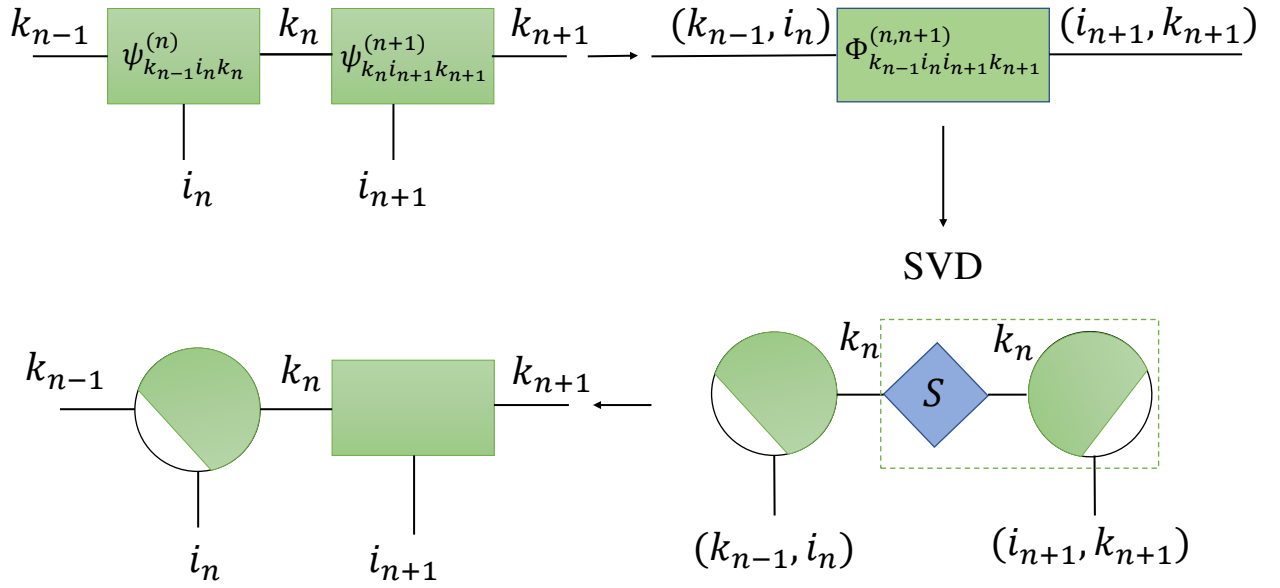


Рисунок 4.9 — Процедура левой ортонормализации n -го вагона тензорного поезда заключается в свёртке по индексу k_n между n -ым и $(n + 1)$ -ым вагоном и дальнейшего применения SVD-разложения. В результате рассмотренной процедуры все ТТ-ранги остаются теми же самыми и процедура сохраняет ту же точность представления квантового состояния.

$I_{r_{n-1}d_n \times r_{n-1}d_n}$), и, следовательно, имеет место следующее выражение для $\psi'^{(n)}(k_{n-1}, i_n, k_n)$:

$$\sum_{i_n=1}^{d_n} \sum_{k_{n-1}=1}^{r_{n-1}} \psi'^{(n)*}(k_n; k_{n-1}, i_n) \psi'^{(n)}(k_{n-1}, i_n; k'_n) = I(k_n; k'_n). \quad (4.45)$$

Будем рассматривать тензор 3-го ранга $\psi'^{(n)}_{k_{n-1}i_n k_n}$ в виде матрицы $\psi'^{(n)}_{k_{n-1}k_n}(i_n)$ размером $r_{n-1} \times r_n$, которая зависит от индекса i_n . тогда выражение (4.45) можно записать в эквивалентном виде:

$$\sum_{i_n=1}^{d_n} \psi'^{(n)\dagger}(i_n) \psi'^{(n)}(i_n) = I_{r_n \times r_n}. \quad (4.46)$$

Выражение (4.46) является свойством лево-ортонормализованных вагонов тензорного поезда. Абсолютно аналогично вышеописанной процедуре рассматривается правая ортонормализация вагонов. Свойство право-ортонормализованных вагонов тензорного поезда записывается следующим

образом:

$$\sum_{i_n=1}^{d_n} \psi'^{(n)}(i_n) \psi'^{(n)\dagger}(i_n) = I_{r_{n-1} \times r_{n-1}}. \quad (4.47)$$

Итак, перед применением двухкубитного гейта между кубитами с номерами n и $n+1$ для повышения точности данной операции при фиксированном максимальном ранге с применением усечённого сингулярного разложения лучше использовать ортонормализацию вагонов тензорного поезда с центром между вагонами n и $n+1$. После процедуры ортонормализации тензорный поезд можно графически представить в виде, показанном на рисунке 4.10 на примере 5-кубитного состояния и ортонормализацией с центром между 3 и 4 вагоном.

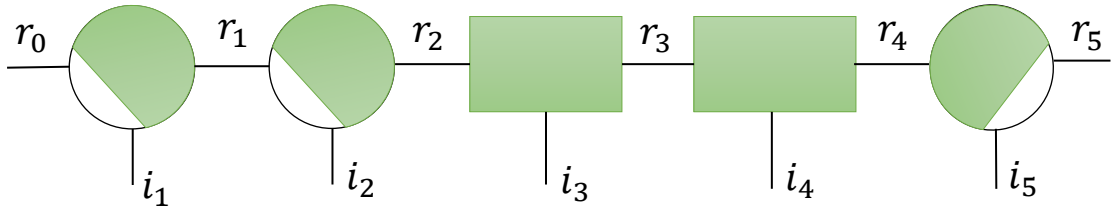


Рисунок 4.10 — Графическое изображение 5-кубитного тензорного поезда, который ортонормализован с центром в точке между 3-им и 4-ым вагоном.

4.4.4 Скалярное произведение квантовых состояний в ТТ-формате

В моделировании квантовых систем важно уметь вычислять скалярное произведение $\langle \varphi | \psi \rangle$ между двумя квантовыми состояниями системы $|\psi\rangle$ и $|\varphi\rangle$. В нашей работе нам необходимо эффективно вычислять данную величину для расчёта *фиделити* между двумя состояниями $F(|\varphi\rangle, |\psi\rangle) = |\langle \varphi | \psi \rangle|^2$, а также для расчёта среднего значения гамильтониана $\hat{H}$ в состоянии $|\psi\rangle$ для определённого класса гамильтонианов $\hat{H}$.

Пусть состояние $|\psi\rangle$ описывается тензором $\psi_{i_1 \dots i_N}$, а состояние $|\varphi\rangle$ описывается тензором $\varphi_{i_1 \dots i_N}$, где каждый индекс i_k пробегает значения от 1 до d_k , и пусть эти тензоры представлены в ТТ-формате (4.28) с вагонами $\psi_{k_{n-1} i_n k_n}^{(n)}$ и $\varphi_{k'_{n-1} i'_n k'_n}^{(n)}$ соответственно.

Алгоритм вычисления скалярного произведения в ТТ-формате представляет собой итеративную процедуру, которая затрагивает сначала первые вагоны тензорных поездов обоих квантовых состояний, потом вторые вагоны и т.д. Понятно, что для вычисления скалярного произведения между двумя состояниями нам необходимо произвести свёртку по всем индексам вагонов, как по физическим, так и по внутренним:

$$\begin{aligned} \langle \varphi | \psi \rangle &= \sum_{i_1=1}^{d_1} \cdots \sum_{i_N=1}^{d_N} \varphi_{i_1 \dots i_N}^* \psi_{i_1 \dots i_N} = \\ &= \sum_{i_1, \dots, i_N} \sum_{k_1, \dots, k_N} \sum_{k'_1, \dots, k'_N} \varphi_{k'_0 i_1 k'_1}^{(1)*} \cdots \varphi_{k'_{N-1} i_N k'_N}^{(N)*} \psi_{k_0 i_1 k_1}^{(1)} \cdots \psi_{k_{N-1} i_N k_N}^{(N)}. \end{aligned} \quad (4.48)$$

Дальше мы можем группировать множители и менять порядок суммирования в этом выражении. Эффективно вычислять эту сумму произведений можно следующим образом. На n -ой итерации нашей процедуры будем обозначать получившийся результат в виде тензора 4-го ранга $c_{k_0 k_n k'_0 k'_n}^{(n)}$. Первая итерация делается в 1 шаг, остальные итерации делаются в 3 шага. На первом шаге первой итерации $c_{k_0 k_1 k'_0 k'_1}^{(1,1)}$ (первый индекс обозначает номер итерации, второй индекс - шаг итерации) равно:

$$c_{k_0 k_1 k'_0 k'_1}^{(1)} = c_{k_0 k_1 k'_0 k'_1}^{(1,1)} = \sum_{i_1=1}^{d_1} \psi_{k_0 i_1 k_1}^{(1)} \varphi_{k'_0 i_1 k'_1}^{(1)*}. \quad (4.49)$$

На первом шаге второй итерации сворачиваем по индексу k_1 тензор $c_{k_0 k_1 k'_0 k'_1}^{(1)}$ со вторым вагоном тензорного поезда квантового состояния $|\psi\rangle$:

$$c_{k_0 k'_0 k'_1 k_2 i_2}^{(2,1)} = \sum_{k_1=1}^{r_1} c_{k_0 k_1 k'_0 k'_1}^{(1)} \psi_{k_1 i_2 k_2}^{(2)}. \quad (4.50)$$

На втором шаге второй итерации сворачиваем по индексу k'_1 тензор $c_{k_0 k'_0 k'_1 k_2 i_2}^{(2,1)}$ со вторым вагоном тензорного поезда квантового состояния $\varphi_{k'_1 i'_2 k'_2}^{(2)*}$:

$$c_{k_0 k'_0 k_2 i_2 i'_2 k'_2}^{(2,2)} = \sum_{k'_1=1}^{r'_1} c_{k_0 k'_0 k'_1 k_2 i_2}^{(2,1)} \varphi_{k'_1 i'_2 k'_2}^{(2)*}. \quad (4.51)$$

На третьем шаге второй итерации сворачиваем по индексам i_2 и i'_2 (при этом $i_2 = i'_2$) тензор $c_{k_0 k'_0 k_2 i_2 i'_2 k'_2}^{(2,2)}$ сам с собой:

$$c_{k_0 k'_0 k_2 k'_2}^{\prime(2)} = c_{k_0 k'_0 k_2 k'_2}^{(2,3)} = \sum_{i_2=1}^{d_2} c_{k_0 k'_0 k_2 i_2 i_2 k'_2}^{(2,2)}. \quad (4.52)$$

Затем, чтобы привести полученный тензор 4-го ранга к виду, похожему по структуре на тензор $c_{k_0 k_1 k'_0 k'_1}^{(1)}$, производим транспонирование полученного тензора — меняем местами второй k'_0 и третий индекс k_2 : $c_{k_0 k_2 k'_0 k'_2}^{(2)} = c_{k_0 k'_0 k_2 k'_2}^{(2)}$.

Абсолютно аналогично повторяем итерации с номерами $3, \dots, N$. По результатам выполнения этих итераций получаем тензор 4-го ранга $c_{k_0 k_N k'_0 k'_N}^{(N)}$, но, поскольку для тензорного поезда выполняются граничные условия $r_0 = r_N = 1$ и $r'_0 = r'_N = 1$, то, по-сути, тензор 4-го ранга $c_{k_0 k_N k'_0 k'_N}^{(N)}$ представляет собой просто скаляр, равный скалярному произведению $\langle \varphi | \psi \rangle$. Графическое представление первой и второй итерации алгоритма вычисления скалярного произведения $\langle \varphi | \psi \rangle$ показано на рисунке 4.11.

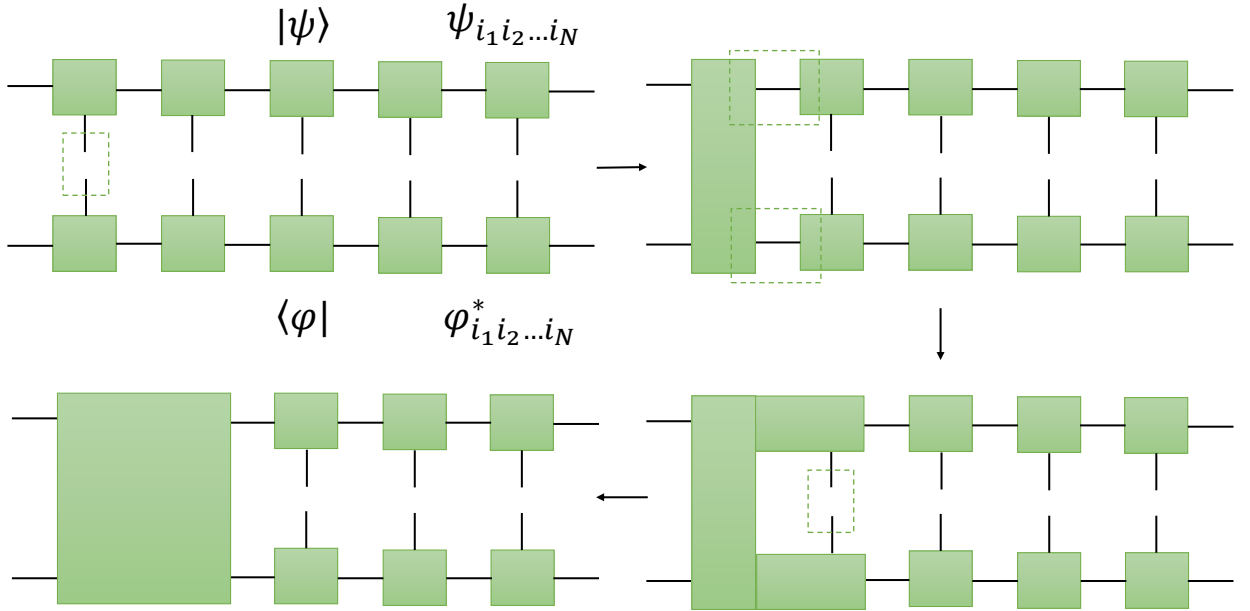


Рисунок 4.11 — Графическое изображение первой и второй итерации алгоритма, позволяющего вычислить скалярное произведение $\langle \varphi | \psi \rangle$ двух квантовых состояний в формате тензорных поездов.

Используя данный алгоритм эффективного вычисления скалярного произведения двух состояний $\langle \varphi | \psi \rangle$, можно вычислять фиделити между двумя состояниями как $|\langle \varphi | \psi \rangle|^2$, а также среднее значение гамильтониана $\hat{H}$ в квантовом состоянии $|\psi\rangle$: $\langle \psi | \hat{H} | \psi \rangle$.

Для вычисления среднего значения удобно использовать разложение гамильтониана по тензорному базису матриц Паули. Любой эрмитов оператор $\hat{H}$, действующий на N кубитах, может быть представлен в виде:

$$\hat{H} = \sum_{\alpha} c_{\alpha} \hat{P}_{\alpha}, \quad (4.53)$$

где $c_{\alpha} \in \mathbb{R}$, а $\hat{P}_{\alpha}$ — строки Паули, то есть операторы вида

$$\hat{P}_{\alpha} = \bigotimes_{k=1}^N \hat{\sigma}_k^{(\alpha)}, \quad \hat{\sigma}_k^{(\alpha)} \in \{\hat{I}, \hat{X}, \hat{Y}, \hat{Z}\}. \quad (4.54)$$

Тогда математическое ожидание гамильтониана принимает вид

$$\langle \psi | \hat{H} | \psi \rangle = \sum_{\alpha} c_{\alpha} \langle \psi | \hat{P}_{\alpha} | \psi \rangle. \quad (4.55)$$

Следовательно, задача сводится к вычислению вкладов $\langle \psi | \hat{P}_{\alpha} | \psi \rangle$ для отдельных строк Паули. Поскольку каждый оператор $\hat{P}_{\alpha}$ является тензорным произведением локальных однокубитных операторов, его действие на состояние $|\psi\rangle$ может быть эффективно вычислено алгоритмами, описанными в разделе 4.4.1, после чего применяется алгоритм вычисления скалярного произведения.

4.4.5 Вычисление матрицы плотности одного кубита

Для некоторых приложений нам нужно будет вычислять матрицу плотности квантового состояния одного кубита с номером n в цепочке кубитов, состояние которых представлено тензорным поездом. Покажем, как вычислять такую матрицу плотности.

Пусть чистое состояние системы из N кубитов есть $|\psi\rangle$. Тогда оператор плотности всей системы из N кубитов вычисляется как $\hat{\rho} = |\psi\rangle\langle\psi|$. Тогда оператор плотности n -го кубита вычисляется с помощью взятия частичного следа по переменным, относящимся к остальным кубитам:

$$\hat{\rho}^{(n)} = \text{Tr}_{i_k, k \neq n} \hat{\rho} = \text{Tr}_{i_k, k \neq n} |\psi\rangle\langle\psi|. \quad (4.56)$$

Пусть состояние N кубитов описывается тензором ранга N $\psi_{i_1 \dots i_N}$. Тогда матрицу плотности n -го кубита в вычислительном базисе можно записать

В виде:

$$\begin{aligned}
\rho_{i_n, j_n}^{(n)} &= \sum_{i_1=1}^{d_1} \cdots \sum_{i_{n-1}=1}^{d_{n-1}} \sum_{i_{n+1}=1}^{d_{n+1}} \cdots \sum_{i_N=1}^{d_N} \psi_{i_1 \dots i_n \dots i_N} \psi_{i_1 \dots j_n \dots i_N}^* = \\
&= \sum_{i_k, k \neq n} \sum_{k_1, \dots, k_N} \sum_{k'_1, \dots, k'_N} \psi_{k_0 i_1 k_1}^{(1)} \cdots \psi_{k_{n-1} i_n k_n}^{(n)} \cdots \psi_{k_{N-1} i_N k_N}^{(N)} \\
&\quad \psi_{k'_0 i_1 k'_1}^{(1)*} \cdots \psi_{k'_{n-1} j_n k'_n}^{(n)*} \cdots \psi_{k'_{N-1} i_N k'_N}^{(N)*}.
\end{aligned} \tag{4.57}$$

Аналогично тому, как мы делаем в разделе 4.4.4, где мы вычисляем скалярное произведение между двумя квантовыми состояниями $|\psi\rangle$ и $|\varphi\rangle$ в ТТ-формате, в выражении (4.57) мы можем группировать множители и менять порядок суммирования. Кратко опишем алгоритм, по которому мы вычисляем однокубитную матрицу плотности. Сначала мы вычисляем тензор 4-го ранга из вагонов с номерами $1, \dots, n-1$ для первого и второго поезда, а затем вычисляем тензор 4-го ранга из вагонов с номерами $n+1, \dots, N$, как показано на рисунке 4.12. Дальше делаем свёртку по двум индексам полученного левого тензора 4-го ранга с двумя вагонами с номером n , а затем свёртку по двум индексам полученного тензора 6-го ранга с правым тензором 4-го ранга. В итоге получаем тензор 6-го ранга $\rho_{k_0 k'_0 i_n j_n k_N k'_N}$, но, в силу граничных условий $r_0 = r_N = 1$ и $r'_0 = r'_N = 1$ этот тензор на самом деле является тензором 2-го ранга - то есть искомой матрицей плотности одного кубита.

4.5 Численные эксперименты

В этом разделе мы рассмотрим результаты, полученные с помощью библиотеки QTensor [90], написание которой является одной из задач оригинальной части работы.

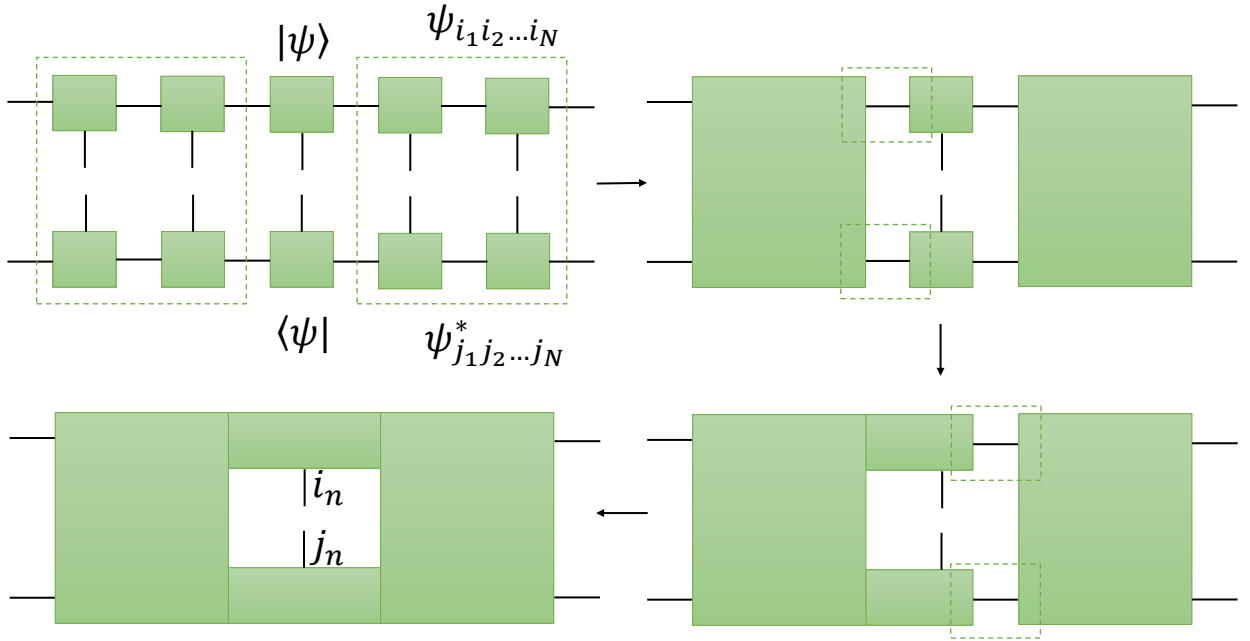


Рисунок 4.12 — Графическое изображение алгоритма, вычисляющего матрицу плотности одного кубита с номером n . На выходе алгоритма получается тензор 6-го ранга, который, в силу граничных условий $r_0 = r_N = 1$ и $r'_0 = r'_N = 1$ является матрицей.

4.5.1 Случайные квантовые схемы

На рисунке 4.13 показана квантовая схема, которую мы используем в наших численных экспериментах. Она состоит из чередующихся слоёв однокубитных и двухкубитных гейтов и относится к классу случайных квантовых схем (random quantum circuits), предложенному в работе [92]. Схемы данного типа предназначены для быстрой генерации сильно запутанных состояний и позднее использовались в задачах random circuit sampling, в том числе в экспериментах по демонстрации квантового превосходства [93]. Кроме того, аналогичная схема была использована в работе [85], что позволяет протестировать нашу библиотеку QTensor путём сравнения с опубликованными результатами.

Случайные квантовые схемы данного типа относятся к числу наиболее трудных для классического моделирования [92; 93], поскольку уже при сравнительно небольшой глубине они приводят к формированию сильно запутанных состояний и сложного выходного распределения вероятностей.

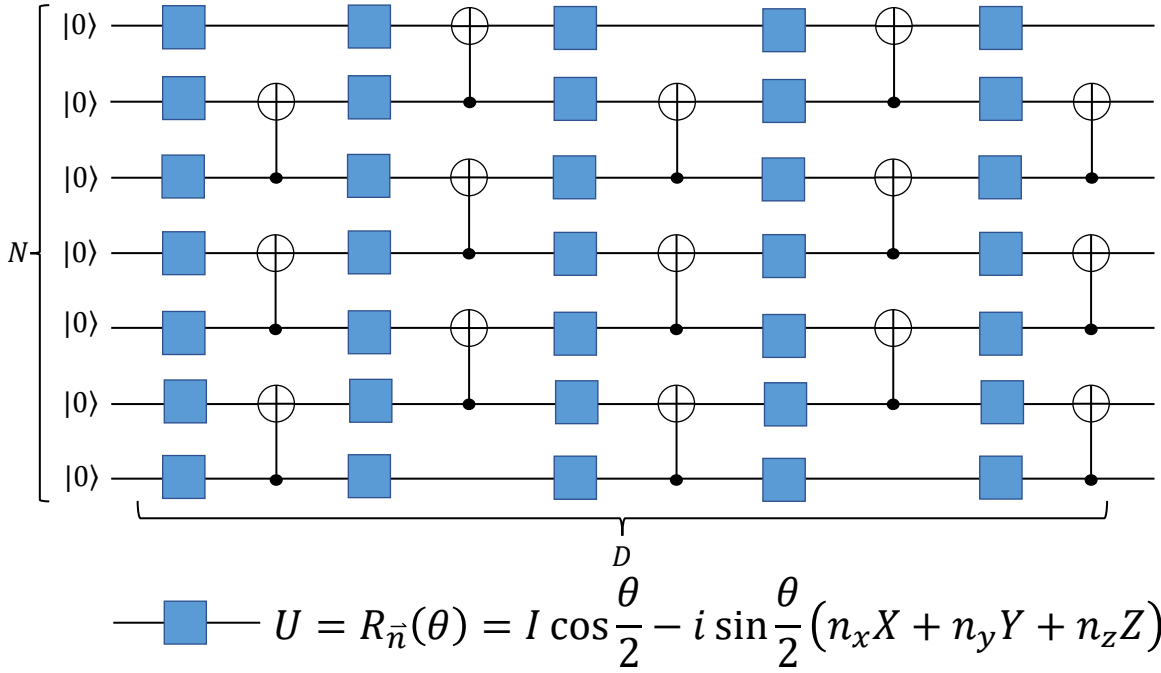


Рисунок 4.13 — N -кубитная случайная квантовая схема. Синие квадраты обозначают случайные однокубитные гейты $\hat{R}_{\vec{n}}(\theta)$, а точки, соединённые с крестиком в кругу, обозначают двухкубитные гейты C_X (в некоторых случаях мы дальше будем использовать гейты C_Z вместо гейтов C_X). Глубина схемы D определяется как количество слоёв двухкубитных гейтов, то есть на данном рисунке изображена схема, имеющая глубину $D = 5$.

При этом итоговое многокубитное состояние оказывается весьма чувствительным к локальным изменениям схемы, в частности к замене отдельных однокубитных гейтов. Однокубитные вентили $\hat{U}$, представленные цветными квадратами на рисунке 4.13, выбираются случайным образом. Однокубитный гейт $\hat{U}$ — это поворот $\hat{R}_{\vec{n}}(\theta)$:

$$\hat{R}_{\vec{n}}(\theta) = e^{-i\frac{\theta}{2}\vec{n}\hat{\vec{\sigma}}} = \hat{I} \cos \frac{\theta}{2} - i \sin \frac{\theta}{2} (n_x \hat{X} + n_y \hat{Y} + n_z \hat{Z}) \quad (4.58)$$

на угол θ вокруг единичного вектора $\vec{n} = (\sin \alpha \cos \varphi, \sin \alpha \sin \varphi, \cos \alpha)$, где $\hat{\vec{\sigma}}$ — вектор, составленный из матриц Паули. Углы θ , α и φ мы берём из равномерного распределения на заданных отрезках.

Следует отметить, что в экспериментах Google по random circuit sampling использовался фиксированный дискретный набор однокубитных гейтов, обусловленный особенностями аппаратной реализации сверхпроводникового процессора. В отличие от этого, в работе [85] и в настоящей

диссертации рассматривается более общий ансамбль случайных однокубитных операций, удобный для численного моделирования. Несмотря на это различие, все указанные схемы принадлежат к одному классу *random quantum circuits* и обладают сходными свойствами с точки зрения генерации сильной запутанности и сложности классического моделирования.

Двухкубитные гейты $\hat{C}_X$ (CNOT) и $\hat{C}_Z$ являются стандартными двухкубитными гейтами. Их действие на базисные состояния задаётся следующими соотношениями:

$$\hat{C}_X|a, b\rangle = |a, b \oplus a\rangle, \quad \hat{C}_Z|a, b\rangle = (-1)^{ab}|a, b\rangle, \quad (4.59)$$

где $a, b \in \{0, 1\}$, а $\oplus$ обозначает сложение по модулю 2.

Глубиной схемы D мы называем количество применённых слоёв двухкубитных гейтов, поскольку они являются единственным источником ошибок в наших расчётах. В реальных квантовых компьютерах двухкубитные гейты также преобладают над ошибками однокубитных гейтов с точки зрения величины фиделити. Однако настоящие квантовые компьютеры также имеют другие источники ошибок (декогеренция, неизвестные связи между кубитами, утечка в невычислительные состояния и т.д.), отсутствующие в нашем алгоритме.

В дальнейшем изложении мы будем придерживаться следующих обозначений. Основной характеристикой, от которой зависит точность вычислений, является величина максимально допустимого ТТ-ранга $\chi = \max r_n$ тензорного разложения, с помощью которого представляется квантовое состояние. Квантовое состояние, которое вычисляется точно, мы будем обозначать как $|\psi\rangle$, а состояние, которое вычислено с учётом ограничения ТТ-ранга до величины χ , мы обозначаем как $|\psi_\chi\rangle$.

4.5.2 Распределение Портера-Томаса

После достижения достаточно большой глубины случайная квантовая схема, показанная на рисунке 4.13, генерирует состояния, статистические свойства которых приближаются к свойствам случайных состояний по мере Хаара в гильбертовом пространстве размерности 2^N . Для случайного

состояния по мере Хаара квадрат модуля скалярного произведения этого состояния с любым фиксированным базисным состоянием, в частности с состоянием $|0\rangle^{\otimes N}$, подчиняется распределению Портера–Томаса [94]. В терминах кумулятивной функции распределения оно имеет вид

$$P_{PT}(\rho) = 1 - (1 - \rho)^{2^N - 1}. \quad (4.60)$$

Это демонстрируется на рисунке 4.14, где кумулятивное распределение величины

$$p_x = |\langle 0|^{\otimes N} |\psi_\chi\rangle|^2 \quad (4.61)$$

сравнивается с распределением Портера–Томаса для различных максимальных ТТ-рангов χ . Поскольку случайные квантовые схемы при достаточной глубине аппроксимируют хааровски-случайные состояния, можно ожидать, что распределение величины p_x будет приближаться к распределению Портера–Томаса. Как видно из рисунка, с ростом χ это согласование действительно улучшается.

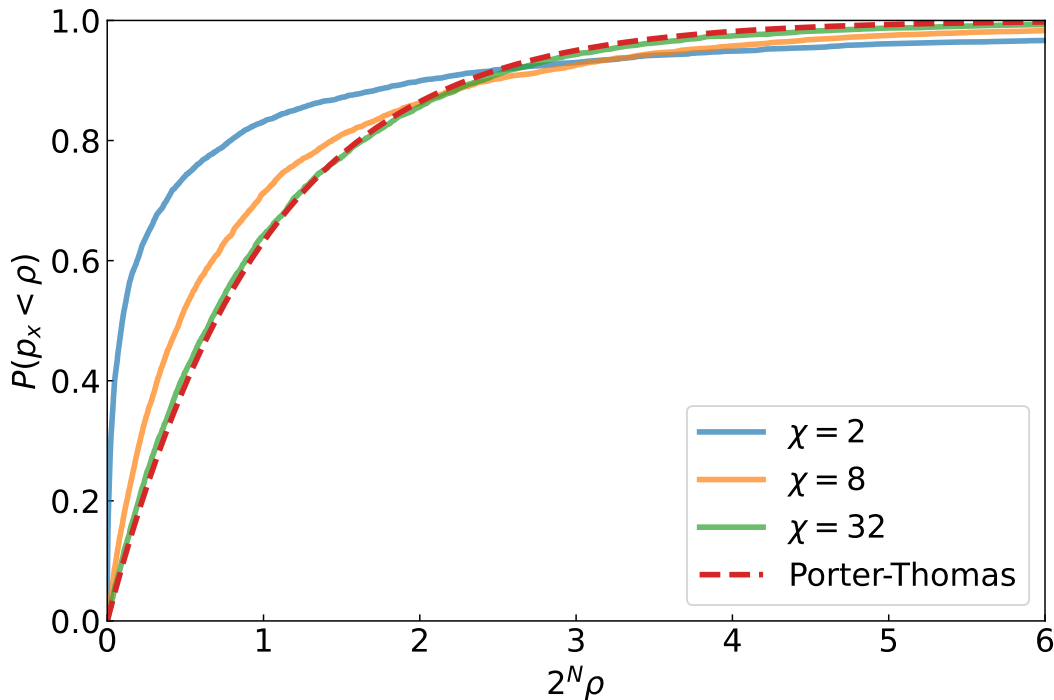


Рисунок 4.14 — Функция распределения вероятности для величины $p_x = |\langle 0|^{\otimes N} |\psi_\chi\rangle|^2$ для квантовой схемы с $N = 15$ и $D = 24$. Пунктирной линией показано точное распределение Портера–Томаса для $N = 15$.

4.5.3 Эффективное двухкубитное фиделити

Введём основную интересующую нас величину — эффективное двухкубитное фиделити f_n . Эффективное двухкубитное фиделити f_n является вычислительным аналогом экспериментального фиделити для двухкубитных гейтов. Для идеального расчёта $f_n = 1$, но ограничение максимального ТТ-ранга χ для MPS приведёт к $0 \leq f_n < 1$.

Состояние MPS после последовательности n отдельных двухкубитных гейтов ($n \approx \frac{ND}{2}$ для схемы на рисунке 4.13) обозначим как $|\psi_\chi(n)\rangle$. Мы определяем эффективное двухкубитное фиделити для схемы с C_Z -гейтами следующим образом:

$$f_n = |\langle \psi_\chi(n) | \hat{C}_Z | \psi_\chi(n-1) \rangle|^2, \quad (4.62)$$

аналогично для схем с C_X -гейтами. Обратим внимание, что в выражении (4.62) операция $\hat{C}_Z$ применяется точно без ограничения ТТ-ранга к состоянию MPS $|\psi_\chi(n-1)\rangle$, и полученное состояние скалярно умножается на уже усечённый результат $|\psi_\chi(n)\rangle$.

Результаты численных симуляций для вычисления эффективного двухкубитного фиделити для C_X -гейтов приведены на рисунке 4.15а, а для C_Z -гейтов на рисунке 4.15б. Видно, что фиделити выше для C_Z -гейтов и улучшается после применения ортонормализации.

Важный вывод, который можно сделать по результатам, представленным на рисунке 4.15, состоит в том, что обработка квантовых состояний в формате MPS с ограничением ТТ-ранга представляет собой практический алгоритм, который позволяет моделировать такие случайные квантовые схемы за умеренное время за счёт конечной точности аппроксимации двухкубитных преобразований, характеризуемой величиной f_n , причём эта точность достигает для схемы с C_Z -гейтами значений порядка 0.99, которая является типичной для современных экспериментальных прототипов таких схем [93].

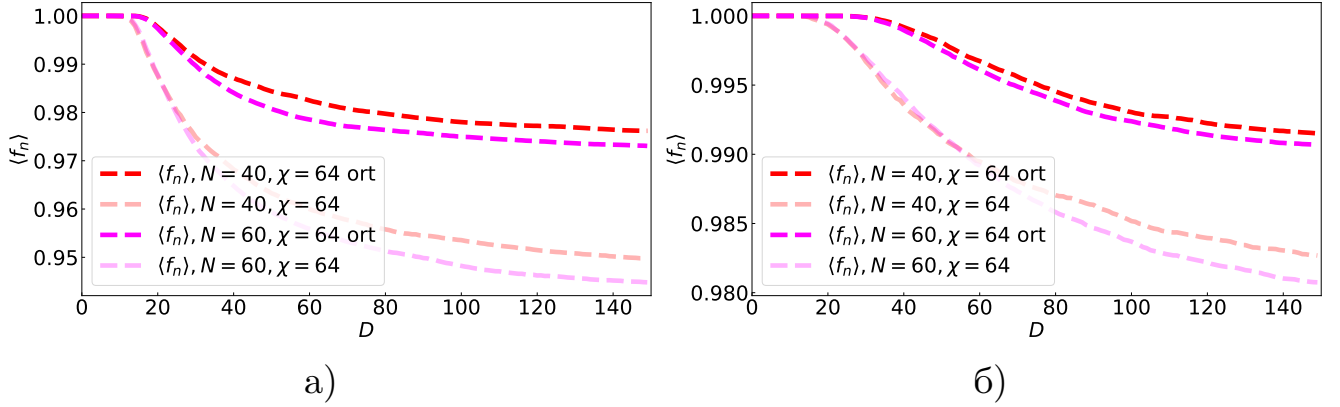


Рисунок 4.15 — Эффективное двухкубитное фиделити f_n в зависимости от глубины D схемы для $\chi = 64$ и для $N = 40$ (красный) и $N = 60$ (пурпурный). Линии соответствуют $\langle f_n \rangle$ — среднему геометрическому f_n по всем двухкубитным вентилям с начала схемы до глубины D . а) — двухкубитные гейты C_X без ортонормализации, б) — двухкубитные гейты C_Z с ортонормализацией и без неё.

4.5.4 Многокубитное фиделити

Определим теперь величину многокубитного фиделити после определённого количества слоёв двухкубитных гейтов D следующим образом:

$$F = |\langle \psi | \psi_\chi \rangle|^2, \quad (4.63)$$

где $|\psi_\chi\rangle$ — состояние после прохождения D слоёв двухкубитных гейтов случайной схемы, вычисленное с использованием процедуры обрезания ТТ-ранга до значения χ , а $|\psi\rangle$ — точно вычисленное без обрезания ранга состояние после прохождения той же глубины схемы D с теми же случайными параметрами.

Результаты численных симуляций для вычисления многокубитного фиделити для схемы с C_X -гейтами приведены на рисунке 4.16. Видно, что ортонормализация существенно улучшает величину фиделити.

Покажем также для большей наглядности диаграмму рассеяния (рисунок 4.17а) и гистограммы распределения многокубитного фиделити (рисунок 4.17б) для схемы с параметрами $N = 10$, $D = 10$ и $\chi = 2$ для большого числа проходов квантовой схемы для разных случайных параметров.

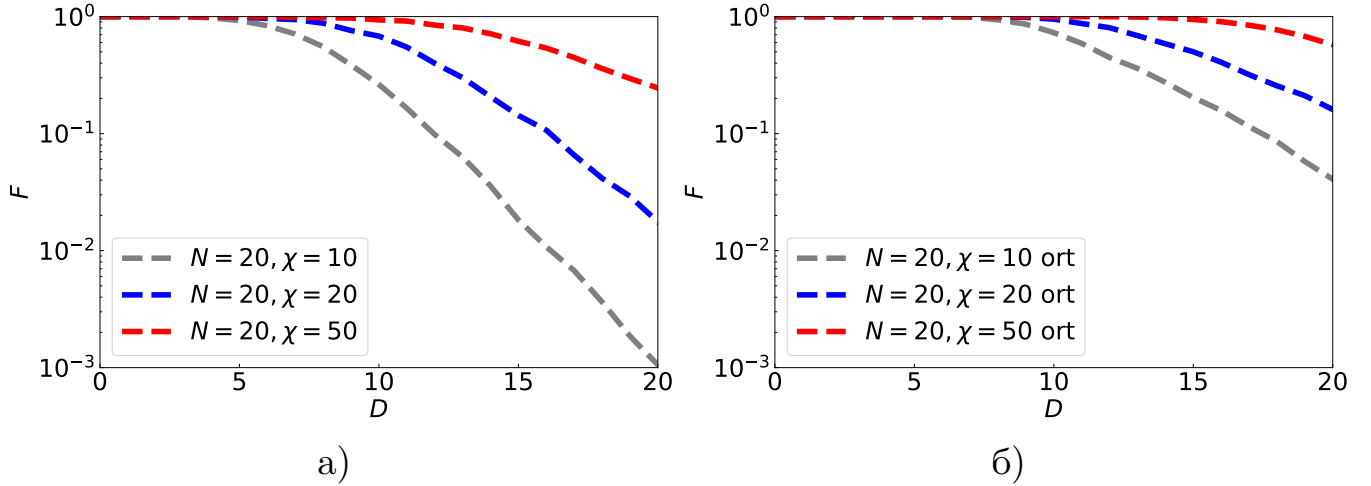


Рисунок 4.16 — Многокубитное фиделити F в зависимости от глубины D для $N = 20$ и различных значений максимального ТТ-ранга $\chi = 10, 20, 50$ для схемы с C_X -гейтами. а) — без ортонормализации, б) — с ортонормализацией.

По рисунку 4.17а можно сделать вывод о том, что ортонормализация действительно увеличивает многокубитное фиделити на выходе случайной квантовой схемы, так как все точки лежат выше серой пунктирной линии $F_{\text{ort}} = F$. Также видно, что ошибки двухкубитных C_Z -гейтов меньше ошибок двухкубитных C_X -гейтов при фиксированном значении параметра максимального ранга χ .

Отметим, что меньшее значение ошибки для схем с C_Z -гейтами по сравнению со схемами с C_X -гейтами при одинаковом ограничении ТТ-ранга χ не противоречит их локальной эквивалентности. При точном моделировании оба гейта эквивалентны с точностью до локальных однокубитных преобразований. Однако в рассматриваемом численном алгоритме после применения гейтов выполняется обрезание ТТ-ранга, и ошибка такой аппроксимации зависит от локального базиса, в котором представлено состояние. Гейт C_Z , будучи диагональным в вычислительном базисе, в среднем сохраняет структуру состояния, более удобную для низкорангового ТТ-представления, тогда как C_X приводит к дополнительному перемешиванию амплитуд в этом базисе. В результате при фиксированном χ состояние после действия C_Z -гейтов аппроксимируется точнее, чем после действия C_X -гейтов, что и проявляется в большем значении многокубитного фиделити.

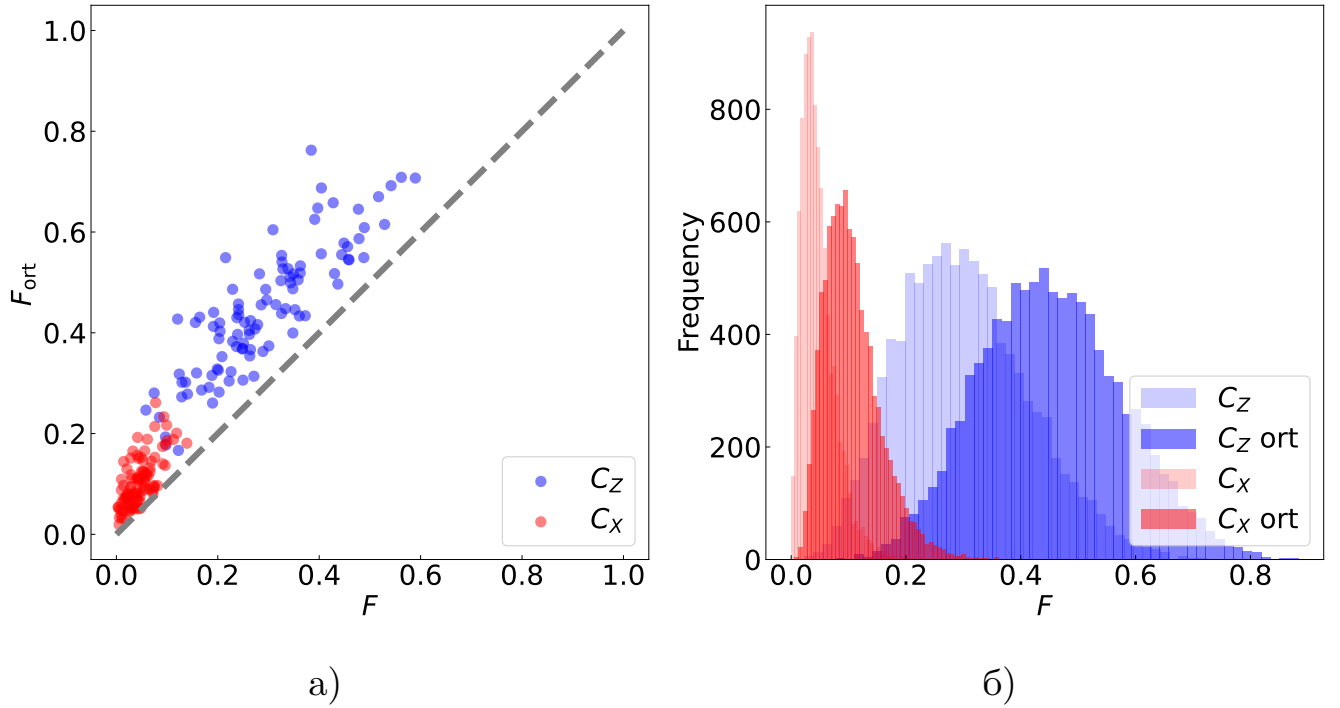


Рисунок 4.17 — а) — Диаграмма рассеяния для многокубитного фиделити F для схемы с параметрами $N = 10$, $D = 10$, $\chi = 2$ для C_X и C_Z гейтов, построенная по 100 случайным запускам (по горизонтальной оси отложено многокубитное фиделити на выходе схемы без ортонормализации, по вертикальной оси - многокубитное фиделити с ортонормализацией для схемы с теми же параметрами, серая пунктирная линия - линия равенства $F_{\text{ort}} = F$), б) — Гистограмма многокубитного фиделити F для такой же схемы, построенная по 10000 случайным запускам.

4.5.5 Распределение точности и чистоты состояний

Рассмотрим вопрос о том, как распределяется точность хранения информации о состоянии отдельных кубитов в ТТ-формате. Поставим вопрос более конкретно. При идеальном хранении квантового состояния состояние каждого отдельного кубита с номером n полностью описывается с помощью однокубитной матрицы плотности $\rho^{(n)}$. При приближённом хранении квантового состояния всей системы в ТТ-формате мы можем вычислить согласно алгоритму из раздела 4.4.5 однокубитовые операторы плотности $\rho_{\chi}^{(n)}$. Тогда точностью хранения состояния одного кубита мы будем называть

величину фиделити для двух однокубитных матриц плотности $\rho^{(n)}$ и $\rho_\chi^{(n)}$:

$$F(\rho_\chi^{(n)}, \rho^{(n)}) = \left(\text{Tr} \sqrt{\sqrt{\rho^{(n)}} \rho_\chi^{(n)} \sqrt{\rho^{(n)}}} \right)^2. \quad (4.64)$$

На рисунке 4.18а мы строим зависимость этого однокубитного фиделити от номера кубита в цепочке. На рисунке 4.18б показано, как распределена чистота $\text{Tr} \rho^2$ состояний кубитов.

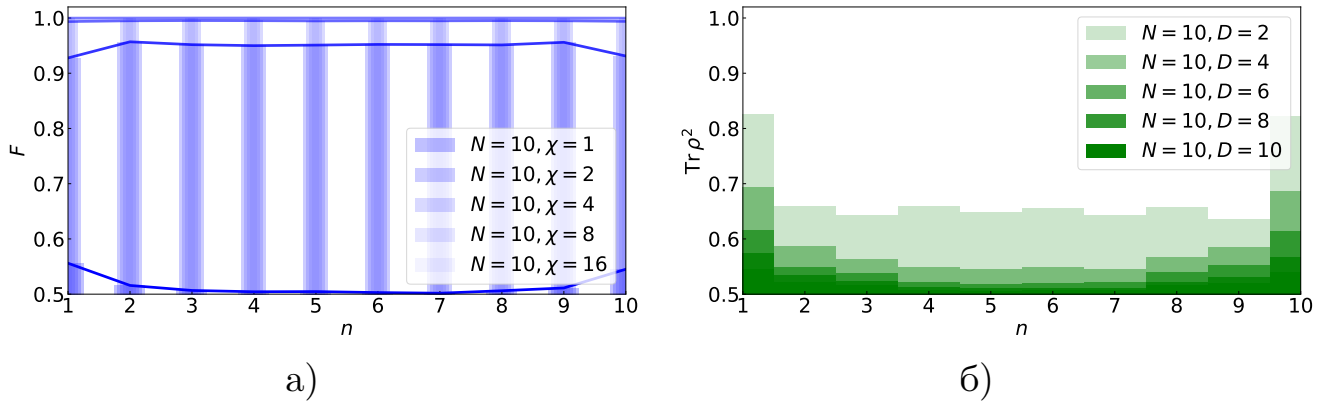


Рисунок 4.18 — Изучение характеристик квантового состояния одного кубита в цепочке в ТТ-формате, параметры схемы $N = 10$ и $D = 10$. а) — распределение однокубитного фиделити между точным состоянием кубита и состоянием в ТТ-формате с ограниченным ТТ-рангом χ (с использованием ортонормализации вагонов MPS), б) — распределение чистоты точных однокубитовых состояний. Графики построены с усреднением по 1000 случайным запускам схемы.

По данным графикам можно сделать следующие выводы. Во-первых, анализ рисунка 4.18б приводит нас к выводу о том, что состояния крайних кубитов в цепочке более чистые, чем в центре, то есть точка изображающая квантовое состояние кубита на сфере Блоха находится ближе к центру сферы для центральных кубитов в цепочке. Такое распределение чистоты связано с распределением ТТ-рангов по различным кубитам в цепочке $r_0, \dots, r_N$ — для нашего конкретного примера случайной квантовой схемы с параметрами $N = 10$ и $D = 10$ мы имеем следующее распределение ТТ-рангов (если не ограничивать максимальный ТТ-ранг χ): $[1, 2, 4, 8, 16, 32, 16, 8, 4, 2, 1]$. То есть в центре ТТ-ранги больше, чем по краям цепочки кубитов, а, следовательно, в центре цепочки состояние кубитов ближе к полностью смешанному состоянию, описываемому оператором

плотности $\hat{\rho}^{(n)} = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$. Таким распределением чистоты состояний кубитов можно объяснить результат распределения однокубитного фиделити, представленного на рисунке 4.18а. По этому графику мы видим, что при $\chi > 1$ мы наблюдаем большее фиделити для центральных кубитов, что объясняется меньшей чистотой их однокубитных состояний. Действительно, состояние центральных кубитов изображается на сфере Блоха точкой, которая лежит близко к центру, а такое состояние легко приблизить с помощью MPS даже с небольшим максимальным ТТ-рангом χ (пример — N -кубитное состояние ГХЦ, представление которого дано выражением (4.26), имеет максимальный ТТ-ранг $\chi = 2$, при этом все кубиты в таком состоянии находятся в полностью смешанном состоянии с операторами плотности $\hat{\rho}^{(n)} = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$).

При $\chi = 1$ мы наблюдаем обратную ситуацию, когда точность представления состояний выше для крайних кубитов, что объясняется тем, что состояние с максимальным ТТ-рангом $\chi = 1$ является факторизованным состоянием, а, значит, все однокубитные матрицы плотности $\rho_{\chi}^{(n)}$ являются матрицами плотности чистых состояний. Фиделити, определённое выражением (4.64) равно $\frac{1}{2}$, если одно из состояний является чистым, а другое — полностью смешанным, поэтому при $\chi = 1$ мы наблюдаем на рисунке 4.18а значение фиделити порядка $\frac{1}{2}$.

Главный вывод, который можно сделать из данного исследования распределения однокубитной точности, состоит в том, что для длинных цепочек кубитов для таких случайных схем точность распределена почти равномерно, за исключением краёв.

Глава 5. Обучение тензорной сетевой модели квантового состояния с использованием однокубитных измерений

В предыдущей главе тензорные сети рассматривались как инструмент эффективного представления и классической симуляции многокубитных квантовых состояний. Формат тензорных поездов позволяет компактно описывать квантовые состояния, выполнять базовые операции, соответствующие элементам квантовой схемы, и вычислять ряд практически значимых характеристик без обращения к полному вектору состояния. Рассмотренный в предыдущей главе подход естественно распространяется не только на задачи моделирования квантовых систем, но и на задачи реконструкции квантового состояния по данным измерений.

Данная глава посвящена обучению тензорной сетевой модели квантового состояния по результатам однокубитных измерений. В отличие от предыдущей главы, где основное внимание уделялось представлению чистых состояний и их эволюции в рамках квантовых схем, здесь тензорно-сетевой подход используется для параметризации оператора плотности и решения задачи реконструкции квантового состояния. Такой переход является естественным развитием изложенного ранее формализма, поскольку компактность тензорно-сетевого представления делает возможным построение вычислительно эффективных алгоритмов, применимых в условиях ограниченного объёма измерительных данных. В данной главе кратко рассматриваются подходы, связанные с теневой томографией и реконструкцией квантовых состояний. Основное внимание уделяется разработке алгоритма обучения тензорно-сетевой модели квантового состояния по результатам однокубитных измерений и исследованию его свойств в численных симуляциях.

5.1 Обзор литературы

Область квантовых вычислений переживает беспрецедентный рост в последнее десятилетие. Основной причиной этого является появление экс-

периментальных прототипов квантовых процессоров с десятками хорошо контролируемых кубитов. Эти устройства пока не достигли порога коррекции ошибок и относятся к так называемому классу процессоров NISQ (Noisy Intermediate-Scale Quantum), однако их сложность достигает предела мощности классического компьютерного моделирования. Появление технологии NISQ увенчалось демонстрацией преимущества квантовых вычислений командой Google AI с использованием сверхпроводящего процессора с 53 кубитами [93]. Другие экспериментальные платформы, такие как захваченные ионы [95], нейтральные атомы [96] и фотоны [97], также становятся всё более конкурентоспособными. Дальнейшее развитие и масштабирование этих устройств требуют новых инструментов для описания их характеристик и оценки их производительности.

Первостепенная задача — определить характеристики многокубитной квантовой системы [98; 99]. Полная реконструкция оператора плотности $\hat{\rho}$ для N кубитов требует экспоненциального числа $O(4^N)$ повторных измерений и классической обработки [100]. Следовательно, полная квантовая томография состояний (QST) не может рассматриваться как практический инструмент для систем среднего и большого масштаба. Однако существует несколько подходов, призванных частично решить эту проблему в некоторых случаях, в частности, с помощью нейросетевого подхода [101—104], а также с помощью тензорных сетей [105; 106] удастся добиться сокращения объема измерений.

На практике описание квантового состояния всегда сопровождается оценкой качества — фиделити или другой интересующей мерой. Однако информация, содержащаяся в полном квантовом состоянии, обычно избыточна для простых бенчмарков. В основополагающей работе [7] теория обучения PAC [107] была применена для ответа на простой вопрос: можно ли эффективно предсказать вероятность успеха двухисходного измерения POVM? Оказалось, что эта задача решается положительно с использованием всего лишь $O(N)$ измерений. Этот результат вызвал интерес, и серия работ [108—114] изучала этот подход, который сейчас обычно называют теневой томографией. За теоретическими исследованиями последовали экспериментальные применения с использованием оптического процессора [115]. Дальнейшие исследования показали, что некоторые характеристики, такие как точность и средние значения наблюдаемых низкоранговых

величин, можно оценить, используя постоянное число копий состояний, независимое от N [8]. Этот результат, как доказано, оптимален с точки зрения теории информации, однако экспериментально он сложен из-за весьма специфичного набора измерений (рандомизированные многокубитные измерения Клиффорда), не свойственного многим экспериментальным платформам. Тем не менее, он был продемонстрирован в ряде экспериментов [116—118].

Помимо доказательств теоретико-информационной сложности [7; 8; 119; 120], поиски практичного протокола без узких мест в вычислительных процессах продолжаются. Здесь на помощь приходит представление квантовых состояний с помощью тензорных сетей. Тензорные сети чрезвычайно эффективны при моделировании квантовых систем большой размерности [85; 121—125]. Недавно было найдено применение для полного QST [126—128].

В основе этого подхода лежит конструктивное доказательство [129] того, что операторы произведения матриц общего вида полностью определяются своими локальными редукциями. Низкоранговое тензорное представление активно используется для томографии квантовых процессов [106; 130—132]. В работе [133] представлен интересный метод, сочетающий теневую томографию с методами локального скремблирования и тензорных сетей.

Здесь мы демонстрируем алгоритм реконструкции, основанный на парадигме теневой томографии, который требует только факторизованных однокубитных измерений и является численно эффективным. Алгоритм эффективно обучается конкретной форме представления квантового состояния в виде тензорной сети. Мы исследуем условия, при которых алгоритм следует линейному масштабированию набора данных $O(N)$, и проводим подробный анализ точности и сходимости алгоритма.

5.2 Тензорно-сетевая модель квантового состояния

5.2.1 Модель измерений

Мы разрабатываем протокол на основе тензорных поездов, реализующий принцип обучения, описанный в [7; 134]. Конечной целью протокола является модель σ , которая точно предсказывает средние значения $\text{Tr}(\hat{E}_m \hat{\rho})$ наблюдаемых $\hat{E}_m$, создаваемых неизвестным квантовым состоянием $\hat{\rho}$. Обучение начинается с обучающего набора данных $\{\hat{E}_m^t, p_m^t\}$, где $\hat{E}_m^t$ — проективное однокубитное измерение, а p_m^t — вероятность получить в этом измерении 1. Затем протокол использует обучающий набор данных для построения правдоподобной модели σ . Начнём с описания математической модели состояния $\hat{\rho}$ и измерения $\hat{E}$, которые мы используем в нашей симуляции.

Смешанное состояние $\hat{\rho}$ представлено в виде тензорной сети МРО (4.25). Измерение, применяемое к регистру, математически описывается положительной операторно-значной мерой (POVM) $\{\hat{E}^{(j)}\}$, где каждый $\hat{E}^{(j)}$ — эрмитово-положительно полуопределённый оператор, такой что $\sum_j \hat{E}^{(j)} = \hat{I}$. Вероятность результата измерения j равна $p_j = \text{Tr}(\hat{E}^{(j)} \hat{\rho})$. Мы сосредоточимся на POVM измерениях с двумя исходами $\{\hat{E}^{(1)} = \hat{E}, \hat{E}^{(2)} = \hat{I} - \hat{E}\}$ в связи с существованием доказательства обучаемости [119] и их непосредственной реализацией в эксперименте [115]. В дальнейшем мы будем рассматривать проективные однокубитные измерения, то есть измерения, описываемые тензорными произведениями операторов однокубитных проективных измерений. Рассмотрение именно такого класса измерений продиктовано прежде всего экспериментальной практикой, поскольку в большинстве реализаций непосредственно доступны именно локальные проективные измерения отдельных кубитов:

$$\hat{E} = \hat{E}_1 \otimes \cdots \otimes \hat{E}_N, \quad (5.1)$$

где $\hat{E}_i = |\psi_i\rangle\langle\psi_i|$ — проектор на чистое состояние $|\psi_i\rangle$. То есть, в представлении тензорным поездом такой оператор может быть представлен тензорным поездом с максимальным рангом $\chi = 1$. Операторы измерения $\hat{E}$ выберем

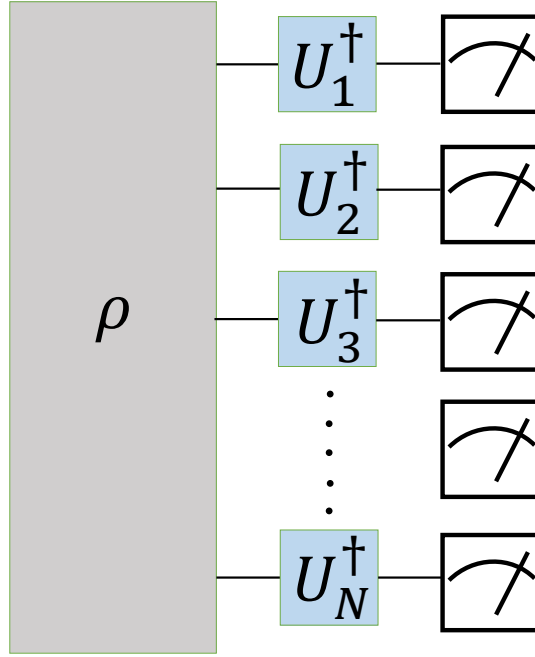


Рисунок 5.1 — Схема генерации обучающего набора данных. Синие квадраты представляют собой случайные однокубитные вентили, содержащиеся в формуле (5.2).

в виде:

$$\hat{E} = \hat{U}_1|0\rangle\langle 0|\hat{U}_1^\dagger \otimes \cdots \otimes \hat{U}_N|0\rangle\langle 0|\hat{U}_N^\dagger, \quad (5.2)$$

где унитарные операторы $\hat{U}_k = \hat{R}_{\vec{n}}(\theta)$ представляют собой повороты сферы Блоха вокруг случайного единичного вектора $\vec{n}$ на угол θ . Единичный вектор $\vec{n}$ равномерно распределён по телесному углу 4π , а угол θ равномерно распределён от 0 до 2π . На рисунке 5.1 представлена схема моделирования, которая выдаёт обучающий набор данных $\{\hat{E}_m^t, p_m^t\}$.

5.2.2 Обучаемая модель

В процессе обучения мы хотим получить модель σ , которая не только воспроизводит обучающие данные $\{\hat{E}_m^t, p_m^t\}$, но и способна предсказать вероятность успеха $\tilde{p}$ нового измерения $\hat{E} \notin \{\hat{E}_m^t, p_m^t\}$. Мы представим мо-

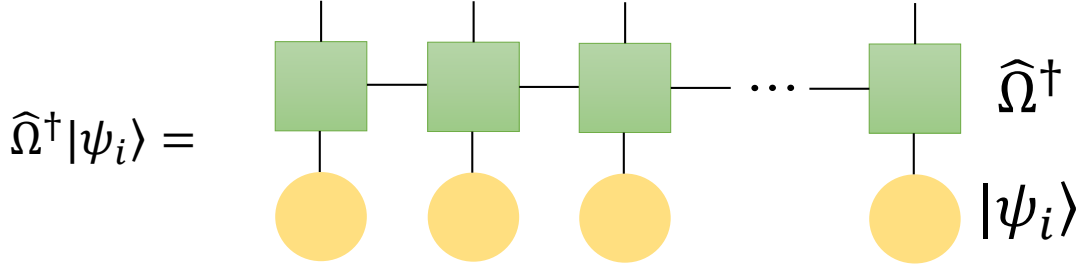


Рисунок 5.2 — Графическое представление процедуры расчета вероятности по формуле (5.4) получения положительного результата в эксперименте с POVM-оператором $\hat{E}_i = |\psi_i\rangle\langle\psi_i|$ в формате тензорного поезда. Жёлтые кружки представляют собой ТТ-ядра $|\psi_i\rangle$ MPS, а зеленые квадраты — ТТ-ядра $\hat{\Omega}^\dagger$ МРО. Линии, соединяющие ТТ-ядра, указывают на то, что суммирование производится по соответствующим индексам. Желтые кружки имеют только одну линию, поскольку все ТТ-ранги $\hat{E}_i$ равны 1.

дель σ как

$$\tilde{p} = \text{Tr}(\hat{E}\hat{\sigma}) = \text{Tr}(\hat{E}\hat{\Omega}\hat{\Omega}^\dagger). \quad (5.3)$$

Поскольку $\hat{E}$ является проектором на чистое состояние $|\psi\rangle$, выражение (5.3) можно сразу переписать в форме, более удобной для численной оценки:

$$\begin{aligned} \tilde{p} &= \text{Tr}(\hat{E}\hat{\Omega}\hat{\Omega}^\dagger) = \text{Tr}(\hat{\Omega}^\dagger\hat{E}\hat{\Omega}) = \text{Tr}(\hat{\Omega}^\dagger|\psi\rangle\langle\psi|\hat{\Omega}) = \\ &= \langle\psi|\hat{\Omega}\hat{\Omega}^\dagger|\psi\rangle = ||\hat{\Omega}^\dagger|\psi\rangle||^2. \end{aligned} \quad (5.4)$$

Алгоритм тензорной сети, который выводит вероятность $\tilde{p}$, проиллюстрирован на рисунке 5.2.

Оператор $\hat{\Omega}$ допускает разложение (4.24), где $\Omega_{i_1\dots i_N j_1\dots j_N}$ — тензор коэффициентов. Мы рассматриваем $\Omega_{i_1\dots i_N j_1\dots j_N}$ как тензорный поезд с максимальным ТТ-рангом $\max r_i = \chi_M$. Более того, мы полагаем, что граничные ТТ-ранги равны 1, а ранги остальных ядер равны максимальному рангу χ_M ($r_0 = r_N = 1$ и $r_k = \chi_M$ для $k \neq 1, N$):

$$\Omega_{i_1\dots i_N j_1\dots j_N} = \sum_{k_1, \dots, k_N=0}^{r_0, \dots, r_N} \omega_{k_0 i_1 j_1 k_1}^{(1)} \cdots \omega_{k_{N-1} i_N j_N k_N}^{(N)}. \quad (5.5)$$

Представление оператора плотности $\hat{\Omega}\hat{\Omega}^\dagger$, где $\hat{\Omega}$ — тензорный поезд, называется локально-очищенным состоянием (locally purified state, LPS) [106].

Действительная и мнимая части ТТ-ядер $\omega_{k_{n-1}i_nj_nk_n}^{(n)}$ являются параметрами обучения в нашем алгоритме. Мы оптимизируем параметры $\hat{\Omega}$ (5.3) для достижения наиболее точного прогнозирования вероятностей успеха $\tilde{p}$ факторизованного измерения. Мы решаем архетипическую задачу обучения с учителем: параметры модели оптимизируются на основе обучающего набора данных $\{\hat{E}_m^t, p_m^t\}$ размера M , а качество модели оценивается на независимом тестовом наборе $\{\hat{E}_n, p_n\} \notin \{\hat{E}_m^t, p_m^t\}$.

Выбор модели предиктора $\sigma = \hat{\Omega}\hat{\Omega}^\dagger$ удовлетворяет условию, что вероятность $\tilde{p} = \text{Tr}(\hat{E}\hat{\Omega}\hat{\Omega}^\dagger)$ должна быть действительным неотрицательным числом. Однако это не гарантирует полноту, а именно, если мы возьмем полный набор операторов измерения $\{\hat{E}_i\}$, которые в сумме дают единицу $\sum_i \hat{E}_i = \hat{I}$, то сумма соответствующих вероятностей не обязательно будет равна единице $\sum_i p_i \neq 1$, то есть след нашей моделируемой матрицы плотности не обязательно будет равен единице. Отметим, что в оригинальной статье [7] предполагалось, что σ не обязательно должна представлять физически значимый объект. Это допущение основано на предположении, что мы больше не ищем описание неизвестного состояния $\hat{\rho}$, а заинтересованы лишь в прогнозировании конкретных характеристик, связанных с этим состоянием.

5.2.3 Функция потерь

Мы оптимизируем параметры $\hat{\Omega}$ путем минимизации функции потерь среднеквадратической ошибки (MSE):

$$\mathcal{L} = \frac{1}{M} \sum_{m=1}^M (\text{Tr}(\hat{E}_m \hat{\Omega} \hat{\Omega}^\dagger) - p_m)^2. \quad (5.6)$$

Функция потерь в формуле (5.6), основанная на метрике *среднеквадратической ошибки* (MSE), широко используется и хорошо подходит [115; 135]

для процесса обучения в контексте наших исследовательских целей — обучения предиктору вероятностей измерений.

Мы используем нормировочный коэффициент $C = (2^N)^2$, то есть $\mathcal{L}' = C\mathcal{L}$, чтобы компенсировать снижение средних значений вероятностей $\sim \frac{1}{2^N}$ в более крупных измерениях. Мы дополняем значение функции потерь $\mathcal{L}'$ коэффициентом детерминации R^2 , что обеспечивает приемлемую оценку качества:

$$R^2 = 1 - \frac{\mathcal{L}}{\sigma_p^2}, \quad (5.7)$$

где σ_p^2 — дисперсия вероятностей на тестовой выборке. Значение R^2 показывает, насколько обученная модель превосходит тривиальную модель, возвращающую постоянное значение вероятности $\bar{p}_n$, равное среднему значению вероятностей для всех операторов измерения $\hat{E}_n$ из тестовой выборки.

5.3 Численные симуляции

5.3.1 Симуляции на случайных смешанных состояниях

Процедура создания смешанного состояния в формате МРО начинается с генерации случайного оператора $\hat{L}$ в формате МРО, у которого все ранги r_i имеют одинаковое значение χ_S , за исключением граничных, которые имеют ранг 1 ($r_0 = r_N = 1$). Каждый элемент ТТ-ядра является комплексным и имеет вид $l_{k_{n-1}i_nj_nk_n}^{(n)} = x_{k_{n-1}i_nj_nk_n}^{(n)} + iy_{k_{n-1}i_nj_nk_n}^{(n)}$, где $x_{k_{n-1}i_nj_nk_n}^{(n)}, y_{k_{n-1}i_nj_nk_n}^{(n)}$ — случайные величины, распределённые в соответствии со стандартным нормальным распределением вероятностей $\mathcal{N}(0, 1)$. Затем мы используем алгоритм сжатия для вычисления $\text{Tr} \hat{L}\hat{L}^\dagger$ и выводим оператор плотности МРО:

$$\hat{\rho} = \frac{\hat{L}\hat{L}^\dagger}{\text{Tr}(\hat{L}\hat{L}^\dagger)}. \quad (5.8)$$

Легко заметить, что такой выбор $\hat{\rho}$ удовлетворяет всем требованиям к оператору плотности: $\hat{\rho}^\dagger = \hat{\rho}$, $\hat{\rho} > 0$, $\text{Tr } \hat{\rho} = 1$. Полученный случайный оператор плотности представляется в ТТ-формате с единичными граничными рангами, а все остальные ранги равны χ_S^2 .

Далее мы выбираем случайный оператор плотности $\hat{\rho}$, используя формулу (5.8), генерируем M различных операторов измерения $\hat{E}_m$ (5.2) и вычисляем вероятности $p_m = \text{Tr}(\hat{E}_m \hat{\rho})$. В результате мы получаем обучающий набор размером M , который затем подаётся на вход протокола обучения в виде кортежа $\{\hat{E}_m, p_m\}$.

Затем мы инициализируем модель σ , используя тот же метод (5.8) с максимальным ТТ-рангом χ_M^2 . Мы рассматриваем случаи, когда $\chi_M < \chi_S$, $\chi_M = \chi_S$ и $\chi_M > \chi_S$. Для минимизации функции потерь мы используем алгоритм стохастического градиентного спуска (5.6). На каждом шаге значение функции потерь усредняется по подмножеству из $m = 5$ элементов, случайно выбранных из обучающего набора данных. Этот метод существенно снижает вычислительные затраты и дает столь же достоверные результаты по сравнению с усреднением по всему обучающему набору данных на каждом шаге.

На рисунке 5.3 показана метрика R^2 обученной модели после 1000 шагов оптимизации. R^2 рассчитывается с использованием независимого тестового набора данных. Размеры обучающего и тестового наборов данных равны M . В качестве порогового значения, указывающего на достижение моделью приемлемого качества, мы используем $R^2 = 0.5$. Точный выбор этого порога в некоторой степени условен, однако он не влияет на свойства масштабирования размера обучающего набора данных, которые представляют наибольший интерес.

На рисунке 5.3(а, б, в) наблюдается насыщение R^2 , достигающее предельного уровня. При увеличении числа кубитов это предельное значение уменьшается. Одной из причин такого поведения может быть трудность невыпуклой многомерной оптимизации функции потерь со сложным ландшафтом. Сравнение рисунков 5.3а и 5.3б показывает, что в режиме $\chi_S \leq \chi_M$ предельное значение R^2 не демонстрирует выраженной зависимости от соотношения χ_M/χ_S , что согласуется с гипотезой о существенной роли оптимизационных эффектов. В то же время сравнение рисунков 5.3б и 5.3в указывает на то, что в режиме $\chi_S > \chi_M$, когда максимальный ранг моде-

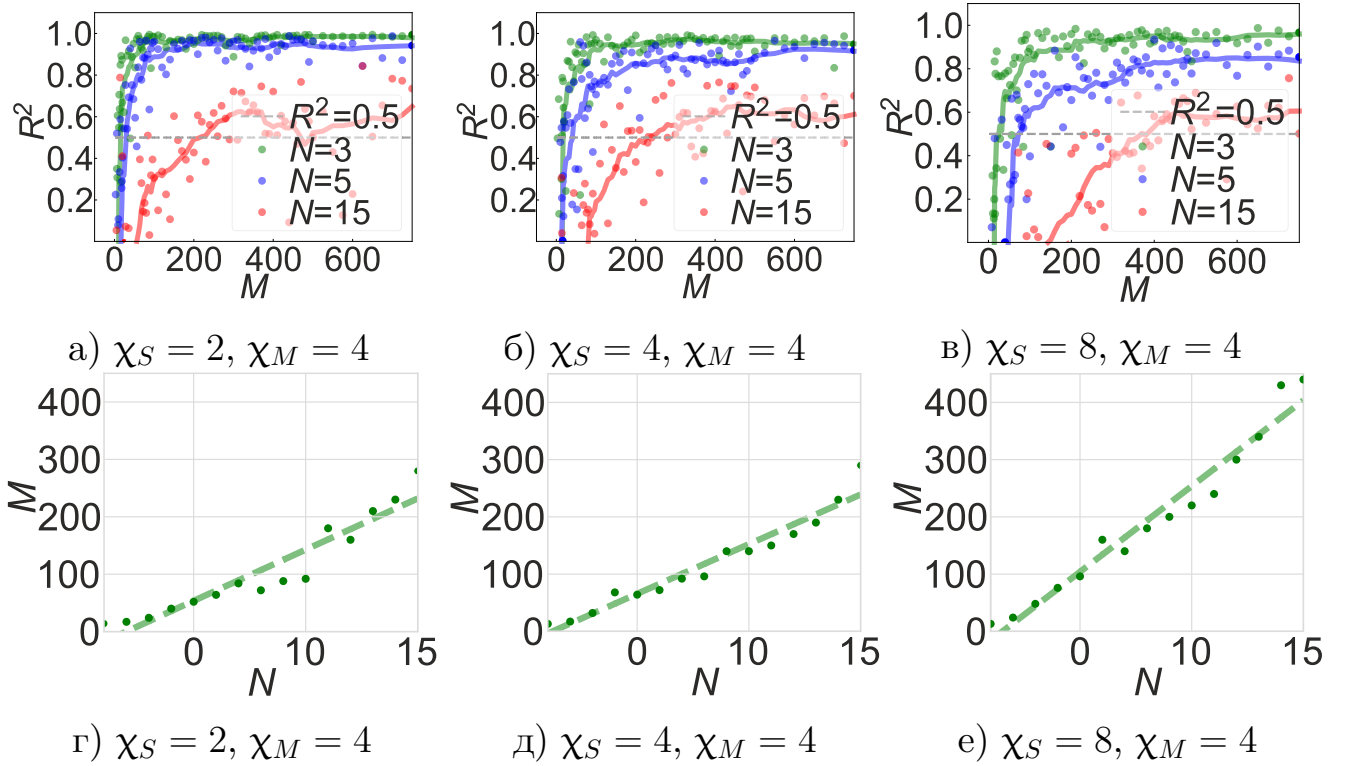


Рисунок 5.3 — Обобщённые результаты численного моделирования процедуры обучения тензорного поезда. На верхних графиках (а, б, в) показана зависимость ключевого показателя качества обучения R^2 от размера обучающей выборки (точки), а также построены скользящие средние с симметричным окном из 11 точек (сплошная линия). На нижних графиках (г, д, е) показано минимальное значение размера обучающей выборки для достижения успеха в обучении по критерию $R^2 = 0.5$ на тестовой выборке (для скользящего среднего).

ли χ_M меньше ранга целевого состояния χ_S , предельное значение метрики R^2 снижается. Это свидетельствует о том, что одним из факторов, ограничивающих качество восстановления в этом случае, является недостаточная выразительная способность модели.

В рамках РАС-подхода допустимо, что небольшая доля измерений предсказывается с заметно большей ошибкой, тогда как в среднем модель сохраняет высокое качество обобщения. Подробности представлены на рисунке 5.4(а, б, в), где мы представляем расчет выборочного стандартного отклонения σ_{ε_p} для оценки доли сильных выбросов и вычисления доли случаев, когда $|\varepsilon_p| > 3\sigma_{\varepsilon_p}$: (а) 1.7%, (б) 1.6%, (в) 0.9%.

Зависимости минимального размера обучающего набора данных, необходимого для достижения значения $R^2 \geq 0.5$, показаны на рисунке 5.3(г, д,

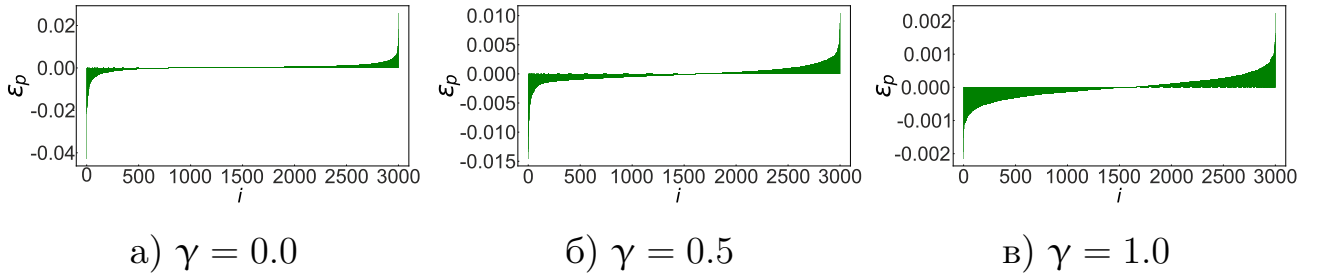


Рисунок 5.4 — Функция распределения вероятностных ошибок $\varepsilon_p = p_{\text{predict}} - p_{\text{exact}}$, полученная путем рассмотрения 30 различных состояний, сгенерированных в соответствии со схемой на рисунке 5.5, и измеренных по 100 различным операторам POVM. Для оценки доли сильных выбросов мы рассчитали для каждого случая (а, б, в) выборочное стандартное отклонение σ_{ε_p} и вычислили долю случаев, когда $|\varepsilon_p| > 3\sigma_{\varepsilon_p}$: а) 1.7%, б) 1.6%, в) 0.9%.

е). Мы провели тесты в трех различных случаях — $\chi_S < \chi_M$, $\chi_S = \chi_M$ и $\chi_S > \chi_M$. Результаты моделирования подтвердили, что модель σ может быть менее информативной, чем истинная модель квантового состояния, и при этом давать точные предсказания $\tilde{p}$. В каждом из трех случаев мы наблюдаем линейную зависимость размера обучающего набора данных M , необходимого для успешного обучения σ для N -кубитного состояния. Однако мы достигаем такой линейной зависимости, ограничивая максимальный ТТ-ранг исследуемого квантового состояния (сравнение рисунка 5.3(д) и рисунка 5.3(е) показывает, что эти линейные зависимости имеют разные наклоны).

5.3.2 Симуляции на реалистичном физическом устройстве

Рассмотрим теперь квантовую схему, показанную на рисунке 5.5, которая генерирует стандартный аппаратно-эффективный анзац [136], который является хорошим примером NISQ-устройства. Схемы с малой глубиной хорошо описываются состояниями на основе тензорного произведения с низким рангом [85; 121] и должны эффективно обучаться с помощью нашего протокола.

Начальное квантовое состояние $|0\rangle^{\otimes N}$ проходит через случайную схему глубины D , состоящую из двух чередующихся слоев случайных одноку-

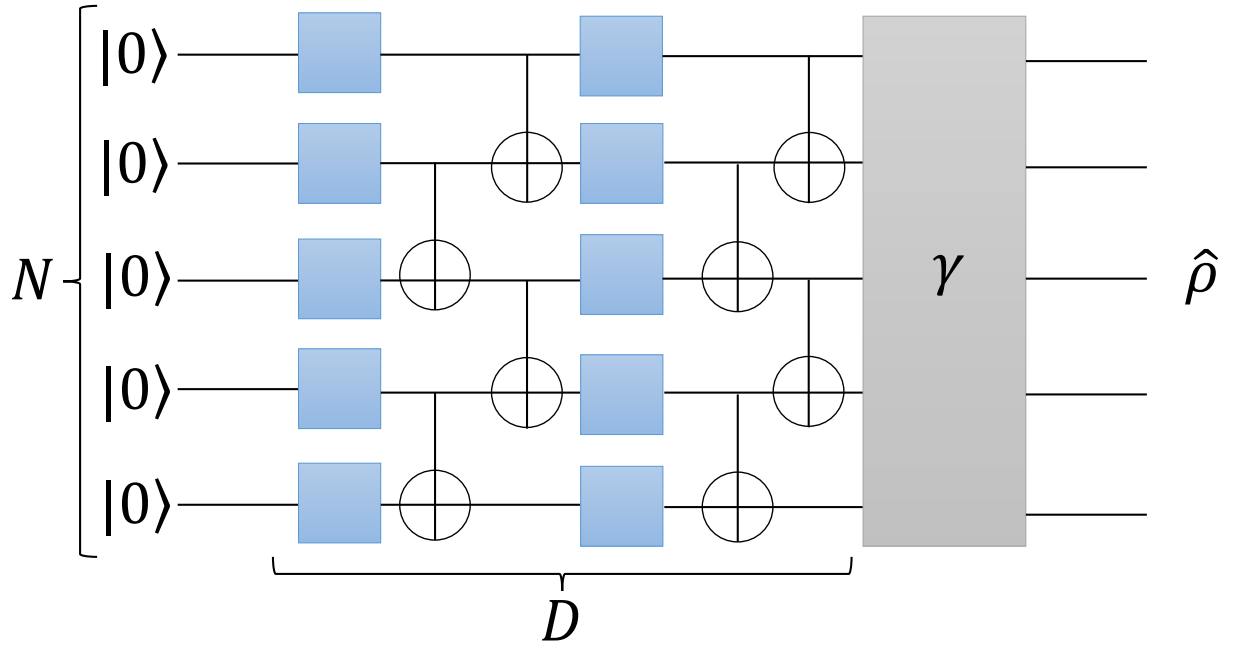


Рисунок 5.5 — Стандартный аппаратно-эффективный анзац, используемый для аппаратного обеспечения NISQ. Синие квадраты представляют собой вращения одного кубита $\hat{U} = \hat{R}_{\vec{n}}(\theta)$ вокруг случайного единичного вектора $\vec{n}$ на угол θ . Единичный вектор $\vec{n}$ равномерно распределен по телесному углу 4π , а угол θ равномерно распределен от 0 до 2π . Серый прямоугольник в конце обозначает канал деполяризации (5.9) с коэффициентом деполяризации γ .

битных вентилях и запутывающих вентилях CNOT. В общем случае, при полном описании квантового состояния на выходе такой схемы, максимальный ТТ-ранг МРО будет равен $\min(4^{\lceil N/2 \rceil}, 4^D)$. Физические квантовые системы всегда подвержены источнику шума, который неизбежно приводит к перемешиванию состояний системы. Поэтому впоследствии контролируемый уровень шума вводится путём применения деполяризующего канала с коэффициентом деполяризации γ :

$$\Delta_\gamma(\hat{\rho}) = (1 - \gamma)\hat{\rho} + \frac{\gamma}{2^N}\hat{I}. \quad (5.9)$$

В нашем численном примере мы использовали схему с $N = 8$ кубитами и глубиной $D = 2$, а ранг обученной модели был выбран равным $\chi_M^2 = 16$.

На рисунке 5.6 сравнивается распределение вероятностей результатов измерений, предсказанных обученной моделью σ , и точное распределение

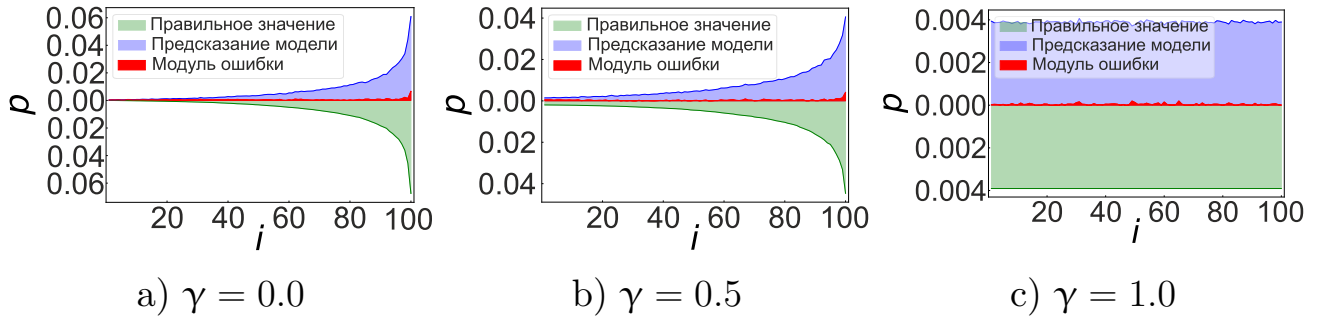


Рисунок 5.6 — Сравнение точных распределений вероятностей с распределениями, сгенерированными обученной моделью. Синим цветом показано распределение вероятностей, предсказанное обученной моделью, зеленым — истинное распределение вероятностей, а красным — модуль их разности. Модель была обучена на наборе данных размером $M = 500$ и 5000 шагами оптимизации. Квантовое состояние — это 8-кубитное состояние $|0\rangle^{\otimes 8}$ на выходе случайной схемы глубины $D = 2$, дополненной деполяризующим каналом с коэффициентом деполяризации γ . Здесь $\chi_S^2 = 16$ и $\chi_M^2 = 16$. Результаты показаны путем усреднения по 30 различным наборам операторов POVM $\hat{E}_i$.

вероятностей. Использовались три различных значения коэффициента деполяризации γ : $\gamma = 0$ соответствует чистому состоянию (см. рисунок 5.6(a)), $\gamma = 0,5$ даёт состояние с промежуточной чистотой (см. рисунок 5.6(б)), а $\gamma = 1,0$ соответствует полностью смешанному состоянию (см. рисунок 5.6(в)).

Индексы i элементов POVM $\hat{E}_i$ нанесены на горизонтальную ось и отсортированы по абсолютному значению точной вероятности. Мы наблюдаем качественно хорошее соответствие между результатами модели и точными значениями. Подходящим количественным показателем качества является точность соответствия между двумя вероятностными распределениями:

$$F = \frac{\sum_i \sqrt{p_i q_i}}{\sqrt{\sum_i p_i \sum_i q_i}}. \quad (5.10)$$

Мы провели серию тестов для глубин D от 0 до 4 (полный ранг для 8 кубитов) и для трех значений коэффициента деполяризации и обнаружили, что значения точности варьировались от 0.96 до 0.999 (рисунок 5.7).

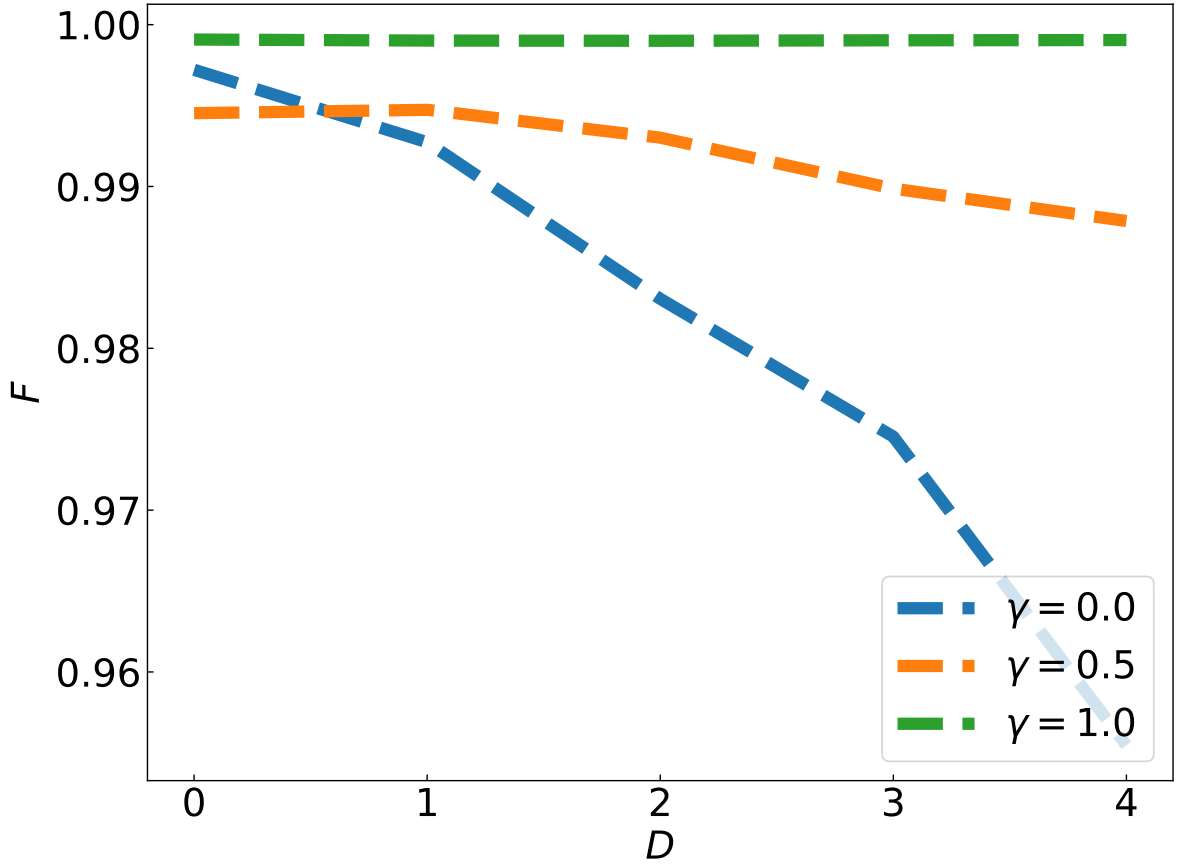


Рисунок 5.7 — Фиделити F между распределениями вероятностей, сгенерированными точным эталонным состоянием, и обученной моделью. Размер обучающего набора данных составляет $M = 500$, количество шагов оптимизации равно 5000 в зависимости от глубины цепи D и коэффициента деполяризации γ .

5.4 Выводы по главе

В данной главе предложен алгоритм восстановления характеристик неизвестного квантового состояния на основе низкоранговой тензорной сетевой модели. В отличие от классической теневой томографии и ряда близких подходов, рассматриваемый метод использует только локальные измерения отдельных кубитов, что делает его применимым в широком классе существующих экспериментальных платформ без введения дополнительных требований к измерительной схеме.

По результатам численного моделирования показано, что минимальный размер обучающей выборки, необходимый для успешного предсказания результатов последующих измерений, растёт линейно с числом кубитов системы. Это указывает на хорошую масштабируемость метода в рассматриваемом классе задач.

Алгоритм исследован для условий, характерных для современных квантовых устройств класса NISQ, а именно для состояний, формируемых квантовыми схемами малой глубины с аппаратно-эффективной структурой, состоящей из чередующихся слоёв параметризуемых однокубитных и фиксированных двухкубитных вентилях. Поскольку такие состояния хорошо описываются низкоранговыми тензорными разложениями [85], предложенный подход может эффективно использоваться для извлечения соответствующего тензорного представления непосредственно из экспериментальных данных.

Все численные результаты в главе получены с использованием разработанной библиотеки QTensor для работы с квантовыми состояниями в формате тензорных последовательностей на базе фреймворка PyTorch. Программный пакет опубликован в открытом доступе на GitLab [90].

Результаты пятой главы опубликованы в работе [137].

Заключение

Основные результаты работы заключаются в следующем.

1. Рассмотрена и формализована задача программирования реконфигурируемых линейно-оптических интерферометров в общем (архитектурно-независимом) виде как задача идентификации параметров модели устройства по экспериментальным данным и последующего подбора фазовых настроек для реализации заданного преобразования.
2. Исследован алгоритм обучения цифровой модели интерферометра по обучающей выборке вида «настройки фаз — реконструированная матрица преобразования». Для целевой функции, основанной на норме Фробениуса, использованы аналитические выражения градиентов и проекция промежуточных оценок на унитарное множество, что обеспечивает практическую реализуемость обучения.
3. В численных экспериментах изучены условия сходимости архитектурно-независимого обучения модели при росте числа мод. Установлено, что в режиме «чёрного ящика», без априорного знания базисных матриц, обучение устойчиво достигается для размерностей до $N = 6$ включительно, тогда как при больших N для успешной сходимости требуется использование априорной информации (в виде начального приближения, близкого к реальным или проектным базисным преобразованиям).
4. Экспериментально продемонстрирована применимость подхода к программированию фотонного процессора на основе обученной модели для случая $N = 4$ в условиях аппаратных неидеальностей, включая наличие кросс-токов между фазовыми элементами, ограничение числа фазовых слоёв (один фазовый слой) и использование отличающихся измерительных процедур; выполнен анализ влияния этих факторов на качество программирования.
5. Предложен и исследован линеаризованный (признаковый) подход к обучению модели линейно-оптического интерферометра, в котором зависимость элементов матрицы преобразования от фазовых пара-

метров представляется линейной по весам в специально выбранном признаковом пространстве.

6. Показано, что в режиме обучения «по всем слоям» (all-layers) линеаризованный подход требует объёма обучающих данных, растущего как $O(N^L)$ по числу мод N и глубине L интерферометрической структуры, что ограничивает применимость этого режима для глубоких схем.
7. Показано, что режим «послойного обучения» (layer-wise) в линеаризованном подходе приводит к линейной по числу мод масштабируемости требуемого объёма данных ($O(N)$ на слой/шаг), что делает метод применимым для более глубоких интерферометрических структур при сохранении практической вычислительной сложности.
8. Разработан и исследован протокол обучения тензорно-сетевой модели квантового состояния по данным факторизованных однокубитных измерений, ориентированный на предсказание вероятностей исходов новых измерений без восстановления полной матрицы плотности.
9. В численных экспериментах для тензорно-сетевого протокола показано, что минимально необходимый объём обучающих данных для достижения заданного качества предсказания демонстрирует линейное масштабирование по числу кубитов ($O(N)$) в рассмотренных сценариях.

Полученные результаты могут быть использованы при разработке процедур калибровки и программирования фотонных процессоров, а также при построении вычислительно эффективных методов характеристики квантовых систем на основе обучаемых компактных моделей.

В заключение автор выражает благодарность и большую признательность научному руководителю Страупе С. С. за поддержку, помощь, обсуждение результатов и научное руководство. Отдельная благодарность выражается Кондратьеву И. В. за неоценимую помощь в проведении экспериментов с линейно-оптическим интерферометром. Также автор благодарит соавторов научных публикаций Михайлову В. И., Дьяконова И. В. и Кулика С. П. за содействие в проведённых исследованиях. Дополнительно автор благодарит создателей Russian-Phd-LaTeX-Dissertation-Template

за предоставление шаблона оформления диссертации. Автор также благодарит всех, кто оказал поддержку и содействие в выполнении настоящей работы.

Список сокращений и условных обозначений

ALS	alternating least squares (метод попеременных наименьших квадратов), стр. 64
DMRG	density matrix renormalization group (алгоритм ренормгруппы матрицы плотности), стр. 85
GD	gradient descent (градиентный спуск), стр. 63
GHZ	квантовое состояние Гринбергера-Хорна-Цайлингера (ГХЦ), стр. 92
LPS	locally purified state (локально-очищенное состояние), стр. 122
MERA	multi-scale entanglement renormalization ansatz (многомасштабный анзац ренормализации запутанности), стр. 86
MPO	matrix product operator (оператор в виде произведения матриц), стр. 86
MPS	matrix product state (чистое квантовое состояние в виде произведения матриц), стр. 78
NISQ	noisy intermediate-scale quantum (квантовые устройства промежуточного масштаба с шумом)
PAC	probably approximately correct (вероятностно приближённо корректное обучение)
PEPS	projected entangled pair states (проецируемые состояния запутанной пары), стр. 86
POVM	positive operator-valued measure (разложение единицы, положительная операторно-значная мера), стр. 120
QML	quantum machine learning (квантовое машинное обучение), стр. 86
SGD	stochastic gradient descent (стохастический градиентный спуск), стр. 63
SLM	spatial light modulator (пространственный модулятор света), стр. 75
SVD	singular value decomposition (сингулярное разложение), стр. 79

TEBD	time-evolving block decimation (алгоритм разрежения блоков во времени), стр. 85
TR	tensor ring (тензорное кольцо), стр. 86
TT	tensor train (тензорный поезд), стр. 78
VQC	variational quantum circuits (вариационные квантовые схемы), стр. 86

Список литературы

1. Experimental realization of any discrete unitary operator / M. Reck [и др.] // Physical review letters. — 1994. — Т. 73, № 1. — С. 58.
2. Optimal design for universal multiport interferometers / W. R. Clements [и др.] // Optica. — 2016. — Т. 3, № 12. — С. 1460—1465.
3. *Laing, A.* Super-stable tomography of any linear optical device / A. Laing, J. L. O'Brien // arXiv preprint arXiv:1208.2868. — 2012.
4. Direct characterization of linear-optical networks / S. Rahimi-Keshari [и др.] // Optics express. — 2013. — Т. 21, № 11. — С. 13450—13458.
5. Rapid characterisation of linear-optical networks via PhaseLift / D. Suess [и др.] // arXiv e-prints. — 2020. — Окт. — arXiv:2010.00517. — arXiv:2010.00517 [physics.optics].
6. Linear programmable nanophotonic processors / N. C. Harris [и др.] // Optica. — 2018. — Т. 5, № 12. — С. 1623—1631.
7. *Aaronson, S.* The learnability of quantum states / S. Aaronson // Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. — 2007. — Т. 463, № 2088. — С. 3089—3114.
8. *Huang, H.-Y.* Predicting many properties of a quantum system from very few measurements / H.-Y. Huang, R. Kueng, J. Preskill // Nature Physics. — 2020. — Т. 16, № 10. — С. 1050—1057.
9. Universal linear optics / J. Carolan [и др.] // Science. — 2015. — Т. 349, № 6249. — С. 711—716.
10. *Nielsen, M. A.* Quantum computation and quantum information / M. A. Nielsen, I. L. Chuang. — Cambridge university press, 2010.
11. *Scully, M. O.* Квантовая оптика / М. О. Scully, М. S. Zubairy ; под ред. В. В. Самарцев ; пер. А. А. К. и др. — Москва : Физматлит, 2003. — С. 512. — Перевод с английского.
12. *Scheel, S.* Permanents in linear optical networks / S. Scheel // arXiv preprint quant-ph/0406127. — 2004.

13. *Aaronson, S.* The computational complexity of linear optics / S. Aaronson, A. Arkhipov // Proceedings of the forty-third annual ACM symposium on Theory of computing. — 2011. — С. 333—342.
14. Boson sampling with 20 input photons and a 60-mode interferometer in a 10^{14} -dimensional hilbert space / H. Wang [и др.] // Physical review letters. — 2019. — Т. 123, № 25. — С. 250503.
15. Multidimensional quantum entanglement with large-scale integrated optics / J. Wang [и др.] // Science. — 2018. — Т. 360, № 6386. — С. 285—291.
16. Integrated photonic quantum technologies / J. Wang [и др.] // Nature photonics. — 2020. — Т. 14, № 5. — С. 273—284.
17. An optical neural chip for implementing complex-valued neural network / H. Zhang [и др.] // Nature communications. — 2021. — Т. 12, № 1. — С. 457.
18. Large-scale optical neural networks based on photoelectric multiplication / R. Hamerly [и др.] // Physical Review X. — 2019. — Т. 9, № 2. — С. 021032.
19. Inference in artificial intelligence with deep optics and photonics / G. Wetzstein [и др.] // Nature. — 2020. — Т. 588, № 7836. — С. 39—47.
20. Multipurpose self-configuration of programmable photonic circuits / D. Pérez-López [и др.] // Nature communications. — 2020. — Т. 11, № 1. — С. 6359.
21. Using an imperfect photonic network to implement random unitaries / R. Burgwal [и др.] // Optics Express. — 2017. — Т. 25, № 23. — С. 28236—28245.
22. Direct dialling of Haar random unitary matrices / N. J. Russell [и др.] // New journal of physics. — 2017. — Т. 19, № 3. — С. 033007.
23. Robust architecture for programmable universal unitaries / M. Y. Saygin [и др.] // Physical review letters. — 2020. — Т. 124, № 1. — С. 010501.

24. Waveguide-lattice-based architecture for multichannel optical transformations / N. N. Skryabin [и др.] // Opt. Express. — 2021. — АБГ. — Т. 29, № 16. — С. 26058—26067. — URL: <https://opg.optica.org/oe/abstract.cfm?URI=oe-29-16-26058>.
25. *Fldzhyan, S. A.* Optimal design of error-tolerant reprogrammable multiport interferometers / S. A. Fldzhyan, M. Y. Saygin, S. P. Kulik // Optics Letters. — 2020. — Т. 45, № 9. — С. 2632—2635.
26. *Fldzhyan, S. A.* Low-depth, compact, and error-tolerant photonic matrix-vector multiplication beyond the unitary group / S. A. Fldzhyan, M. Y. Saygin, S. S. Straupe // Optics Express. — 2024. — Т. 32, № 26. — С. 46239—46250.
27. *Fldzhyan, S.* Low-depth two-unitary design of programmable photonic circuits / S. Fldzhyan, M. Y. Saygin, S. Straupe // Physical Review Research. — 2026. — Т. 8, № 1. — С. 013021.
28. 60 dB high-extinction auto-configured Mach–Zehnder interferometer / C. M. Wilkes [и др.] // Optics letters. — 2016. — Т. 41, № 22. — С. 5318—5321.
29. Silica-on-silicon waveguide quantum circuits / A. Politi [и др.] // Science. — 2008. — Т. 320, № 5876. — С. 646—649.
30. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer / A. Y. Piggott [и др.] // Nature photonics. — 2015. — Т. 9, № 6. — С. 374—377.
31. Efficient, compact and low loss thermo-optic phase shifter in silicon / N. C. Harris [и др.] // Optics express. — 2014. — Т. 22, № 9. — С. 10487—10493.
32. Thermally reconfigurable quantum photonic circuits at telecom wavelength by femtosecond laser micromachining / F. Flamini [и др.] // Light: Science & Applications. — 2015. — Т. 4, № 11. — e354—e354.
33. Low-loss single-mode integrated waveguides in soda-lime glass / I. Dyakonov [и др.] // Applied Physics B. — 2016. — Т. 122, № 9. — С. 245.

34. *Hong, C.-K.* Measurement of subpicosecond time intervals between two photons by interference / C.-K. Hong, Z.-Y. Ou, L. Mandel // Physical review letters. — 1987. — Т. 59, № 18. — С. 2044.
35. Reconfigurable photonics on a glass chip / I. Dyakonov [и др.] // Physical Review Applied. — 2018. — Т. 10, № 4. — С. 044048.
36. *Tillmann, M.* On unitary reconstruction of linear optical networks / M. Tillmann, C. Schmidt, P. Walther // Journal of Optics. — 2016. — СЕНТ. — Т. 18, № 11. — С. 114002. — URL: <https://doi.org/10.1088%2F2040-8978%2F18%2F11%2F114002>.
37. *Nielsen, M. A.* Neural networks and deep learning. Т. 25 / M. A. Nielsen. — Determination press San Francisco, CA, USA, 2015.
38. *Ohn, I.* Smooth function approximation by deep neural networks with general activation functions / I. Ohn, Y. Kim // Entropy. — 2019. — Т. 21, № 7. — С. 627.
39. *Ferrari, S.* Smooth function approximation using neural networks / S. Ferrari, R. F. Stengel // IEEE Transactions on Neural Networks. — 2005. — Т. 16, № 1. — С. 24—38.
40. *Mezzadri, F.* How to generate random matrices from the classical compact groups / F. Mezzadri. — 2006. — arXiv: math-ph/0609050 [math-ph].
41. Some metric inequalities in the space of matrices / K. Fan, A. J. Hoffman [и др.] // Proc. Amer. Math. Soc. Т. 6. — World Scientific. 1955. — С. 111—116.
42. *Kuzmin, S.* NNOptic / S. Kuzmin. — 2020. — URL: <https://gitlab.com/SergeiKuzmin/nnoptic>.
43. On the saddle point problem for non-convex optimization / R. Pascanu [и др.] // arXiv preprint arXiv:1405.4604. — 2014.
44. Escaping saddle points with adaptive gradient methods / M. Staib [и др.] // International Conference on Machine Learning. — PMLR. 2019. — С. 5956—5965.
45. *Kingma, D. P.* Adam: A method for stochastic optimization / D. P. Kingma // arXiv preprint arXiv:1412.6980. — 2014.

46. *Cai, C.* Femtosecond Laser-Fabricated Photonic Chips for Optical Communications: A Review / C. Cai, J. Wang // *Micromachines*. — 2022. — Т. 13, № 4. — URL: <https://www.mdpi.com/2072-666X/13/4/630>.
47. Two-qubit quantum photonic processor manufactured by femtosecond laser writing / N. Skryabin [и др.] // *Applied Physics Letters*. — 2023. — Т. 122, № 12. — С. 121102.
48. 20-mode universal quantum photonic processor / C. Taballione [и др.] // *Quantum*. — 2023. — Т. 7. — С. 1071.
49. Power-efficient programmable integrated multiport photonic interferometer in CMOS-compatible silicon nitride / S. Lin [и др.] // *Photonics Research*. — 2024. — Т. 12, № 3. — A11—A20.
50. Multiport universal unitary interferometer design / I. Kondratyev [и др.] // *Aip conference proceedings*. Т. 2241. — AIP Publishing. 2020.
51. Fast reconstruction of programmable integrated interferometers / B. Bantysh [и др.] // *Optics Express*. — 2023. — Т. 31, № 10. — С. 16729—16742.
52. *Kuzmin, S.* Architecture agnostic algorithm for reconfigurable optical interferometer programming / S. Kuzmin, I. Dyakonov, S. Kulik // *Optics Express*. — 2021. — Vol. 29, no. 23. — P. 38429—38440.
53. Effective programming of a photonic processor with complex interferometric structure / I. V. Kondratyev [и др.] // *Physical Review Applied*. — 2026. — Т. 25, № 3. — С. 034072.
54. *Comon, P.* Tensor decompositions, alternating least squares and other tales / P. Comon, X. Luciani, A. L. De Almeida // *Journal of Chemometrics: A Journal of the Chemometrics Society*. — 2009. — Т. 23, № 7/8. — С. 393—405.
55. *Liu, D. C.* On the limited memory BFGS method for large scale optimization / D. C. Liu, J. Nocedal // *Mathematical Programming*. — 1989. — АБГ. — Т. 45, № 1. — С. 503—528. — URL: <https://doi.org/10.1007/BF01589116>.

56. *Kay, S. M.* Fundamentals of statistical signal processing: estimation theory / S. M. Kay. — Prentice-Hall, Inc., 1993.
57. Reference-less measurement of the transmission matrix of a highly scattering material using a DMD and phase retrieval techniques / A. Drémeau [и др.] // Optics express. — 2015. — Т. 23, № 9. — С. 11898—11911.
58. 8×8 reconfigurable quantum photonic processor based on silicon nitride waveguides / C. Taballione [и др.] // Optics express. — 2019. — Т. 27, № 19. — С. 26842—26857.
59. Large-scale error-tolerant programmable interferometer fabricated by femtosecond laser writing / I. Kondratyev [и др.] // Photonics Research. — 2024. — Т. 12, № 3. — A28—A40.
60. Electro-optic modulation in integrated photonics / G. Sinatkas [и др.] // Journal of Applied Physics. — 2021. — Т. 130, № 1.
61. *Kuzmin, S.* Leveraging a feature-based model for linear optical interferometer control / S. Kuzmin, I. Dyakonov, S. Straupe // Physical Review A. — 2025. — Vol. 112, no. 5. — P. 053515.
62. *Johnson, R. M.* On a theorem stated by eckart and young / R. M. Johnson // Psychometrika. — 1963. — Сент. — Т. 28, № 3. — С. 259—263. — URL: <https://doi.org/10.1007/BF02289573>.
63. *De Lathauwer, L.* A Multilinear Singular Value Decomposition / L. De Lathauwer, B. De Moor, J. Vandewalle // SIAM Journal on Matrix Analysis and Applications. — 2000. — Т. 21, № 4. — С. 1253—1278. — eprint: <https://doi.org/10.1137/S0895479896305696>. — URL: <https://doi.org/10.1137/S0895479896305696>.
64. *Kolda, T. G.* Tensor Decompositions and Applications / T. G. Kolda, B. W. Bader // SIAM Review. — 2009. — Т. 51, № 3. — С. 455—500. — eprint: <https://doi.org/10.1137/07070111X>. — URL: <https://doi.org/10.1137/07070111X>.

65. Full-dimensional (15-dimensional) quantum-dynamical simulation of the protonated water dimer. I. Hamiltonian setup and analysis of the ground vibrational state / O. Vendrell [и др.] // The Journal of Chemical Physics. — 2007. — Т. 127, № 18. — С. 184302. — eprint: <https://doi.org/10.1063/1.2787588>. — URL: <https://doi.org/10.1063/1.2787588>.
66. *Babuška, I.* A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data / I. Babuška, F. Nobile, R. Tempone // SIAM Journal on Numerical Analysis. — 2007. — Т. 45, № 3. — С. 1005—1034. — eprint: <https://doi.org/10.1137/050645142>. — URL: <https://doi.org/10.1137/050645142>.
67. *Sloan, I. H.* When Are Quasi-Monte Carlo Algorithms Efficient for High Dimensional Integrals? / I. H. Sloan, H. Woźniakowski // Journal of Complexity. — 1998. — Т. 14, № 1. — С. 1—33. — URL: <https://www.sciencedirect.com/science/article/pii/S0885064X97904635>.
68. *Carroll, J. D.* Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition / J. D. Carroll, J.-J. Chang // Psychometrika. — 1970. — Сент. — Т. 35, № 3. — С. 283—319. — URL: <https://doi.org/10.1007/BF02310791>.
69. *Håstad, J.* Tensor rank is NP-complete / J. Håstad // Journal of Algorithms. — 1990. — Т. 11, № 4. — С. 644—654. — URL: <https://www.sciencedirect.com/science/article/pii/0196677490900146>.
70. *Silva, V. de.* Tensor Rank and the Ill-Posedness of the Best Low-Rank Approximation Problem / V. de Silva, L.-H. Lim // SIAM Journal on Matrix Analysis and Applications. — 2008. — Т. 30, № 3. — С. 1084—1127. — eprint: <https://doi.org/10.1137/06066518X>. — URL: <https://doi.org/10.1137/06066518X>.
71. *Oseledets, I. V.* Tensor-Train Decomposition / I. V. Oseledets // SIAM Journal on Scientific Computing. — 2011. — Т. 33, № 5. — С. 2295—2317. — eprint: <https://doi.org/10.1137/090752286>. — URL: <https://doi.org/10.1137/090752286>.
72. *Schollwöck, U.* The density-matrix renormalization group in the age of matrix product states / U. Schollwöck // Annals of Physics. — 2011. —

- T. 326, № 1. — C. 96—192. — URL: <https://www.sciencedirect.com/science/article/pii/S0003491610001752> ; January 2011 Special Issue.
73. *Verstraete, F.* Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems / F. Verstraete, V. Murg, J. Cirac // *Advances in Physics*. — 2008. — Т. 57, № 2. — C. 143—224. — eprint: <https://doi.org/10.1080/14789940801912366>. — URL: <https://doi.org/10.1080/14789940801912366>.
 74. *White, S. R.* Density matrix formulation for quantum renormalization groups / S. R. White // *Phys. Rev. Lett.* — 1992. — Ноябрь. — Т. 69, вып. 19. — C. 2863—2866. — URL: <https://link.aps.org/doi/10.1103/PhysRevLett.69.2863>.
 75. *Vidal, G.* Efficient Classical Simulation of Slightly Entangled Quantum Computations / G. Vidal // *Phys. Rev. Lett.* — 2003. — Октябрь. — Т. 91, вып. 14. — C. 147902. — URL: <https://link.aps.org/doi/10.1103/PhysRevLett.91.147902>.
 76. Tensor ring parametrized variational quantum circuits for large scale quantum machine learning / D. Peddireddy [и др.] // *arXiv preprint arXiv:2201.08878*. — 2022.
 77. *Schuch, N.* PEPS as ground states: Degeneracy and topology / N. Schuch, I. Cirac, D. Pérez-García // *Annals of Physics*. — 2010. — Т. 325, № 10. — C. 2153—2192. — URL: <https://www.sciencedirect.com/science/article/pii/S0003491610000990>.
 78. *Cincio, L.* Multiscale Entanglement Renormalization Ansatz in Two Dimensions: Quantum Ising Model / L. Cincio, J. Dziarmaga, M. M. Rams // *Phys. Rev. Lett.* — 2008. — Июнь. — Т. 100, вып. 24. — C. 240603. — URL: <https://link.aps.org/doi/10.1103/PhysRevLett.100.240603>.
 79. *Luchnikov, I. A.* Simulating quantum circuits using the multi-scale entanglement renormalization ansatz / I. A. Luchnikov, A. V. Berezutskii, A. K. Fedorov. — 2021. — URL: <https://arxiv.org/abs/2112.14046>.

80. *Holtz, S.* The Alternating Linear Scheme for Tensor Optimization in the Tensor Train Format / S. Holtz, T. Rohwedder, R. Schneider // SIAM Journal on Scientific Computing. — 2012. — Т. 34, № 2. — A683—A713. — eprint: <https://doi.org/10.1137/100818893>. — URL: <https://doi.org/10.1137/100818893>.
81. *Dolgov, S. V.* A tensor decomposition algorithm for large ODEs with conservation laws / S. V. Dolgov. — 2014. — URL: <https://arxiv.org/abs/1403.8085>.
82. *Quiñones-Valles, D.* Tensor Product Approach to Quantum Control / D. Quiñones-Valles, S. Dolgov, D. Savostyanov // Integral Methods in Science and Engineering: Analytic Treatment and Numerical Approximations / под ред. C. Constanda, P. Harris. — Cham : Springer International Publishing, 2019. — С. 367—379. — URL: https://doi.org/10.1007/978-3-030-16077-7_29.
83. *Oseledets, I.* ttpy / I. Oseledets. — 2015. — URL: <https://github.com/oseledets/ttpy>.
84. *Fishman, M.* The ITensor Software Library for Tensor Network Calculations / M. Fishman, S. R. White, E. M. Stoudenmire. — 2020. — URL: <https://arxiv.org/abs/2007.14822>.
85. *Zhou, Y.* What Limits the Simulation of Quantum Computers? / Y. Zhou, E. M. Stoudenmire, X. Waintal // Phys. Rev. X. — 2020. — Ноябрь. — Т. 10, вып. 4. — С. 041038. — URL: <https://link.aps.org/doi/10.1103/PhysRevX.10.041038>.
86. *Efthymiou, S.* TensorNetwork for Machine Learning / S. Efthymiou, J. Hidary, S. Leichenauer. — 2019. — URL: <https://arxiv.org/abs/1906.06329>.
87. TensorNetwork on TensorFlow: A Spin Chain Application Using Tree Tensor Networks / A. Milsted [и др.]. — 2019. — URL: <https://arxiv.org/abs/1905.01331>.
88. TensorNetwork: A Library for Physics and Machine Learning / C. Roberts [и др.]. — 2019. — URL: <https://arxiv.org/abs/1905.01330>.
89. *Oseledets, I.* TT-Toolbox / I. Oseledets. — 2014. — URL: <https://github.com/oseledets/TT-Toolbox>.

90. *Kuzmin, S.* QTensor / S. Kuzmin. — 2022. — URL: <https://gitlab.com/SergeiKuzmin/qtensor>.
91. Matrix product operators, matrix product states, and ab initio density matrix renormalization group algorithms / G. K. Chan [и др.] // The Journal of chemical physics. — 2016. — Т. 145, № 1.
92. Characterizing quantum supremacy in near-term devices / S. Boixo [и др.] // Nature Physics. — 2018. — ИЮНЬ. — Т. 14, № 6. — С. 595—600. — URL: <https://doi.org/10.1038/s41567-018-0124-x>.
93. Quantum supremacy using a programmable superconducting processor / F. Arute [и др.] // Nature. — 2019. — ОКТ. — Т. 574, № 7779. — С. 505—510. — URL: <https://doi.org/10.1038/s41586-019-1666-5>.
94. *Porter, C. E.* Fluctuations of Nuclear Reaction Widths / C. E. Porter, R. G. Thomas // Phys. Rev. — 1956. — ОКТ. — Т. 104, ВЫП. 2. — С. 483—491. — URL: <https://link.aps.org/doi/10.1103/PhysRev.104.483>.
95. Benchmarking an 11-qubit quantum computer / K. Wright [и др.] // Nature communications. — 2019. — Т. 10, № 1. — С. 5464.
96. Quantum phases of matter on a 256-atom programmable quantum simulator / S. Ebadi [и др.] // Nature. — 2021. — Т. 595, № 7866. — С. 227—232.
97. Interleaving: Modular architectures for fault-tolerant photonic quantum computing / H. Bombin [и др.] // arXiv preprint arXiv:2103.08612. — 2021.
98. *Caves, C. M.* Unknown quantum states: the quantum de Finetti representation / C. M. Caves, C. A. Fuchs, R. Schack // Journal of Mathematical Physics. — 2002. — Т. 43, № 9. — С. 4537—4559.
99. *D'Ariano, G. M.* Optimal data processing for quantum measurements / G. M. D'Ariano, P. Perinotti // Physical review letters. — 2007. — Т. 98, № 2. — С. 020403.
100. Sample-optimal tomography of quantum states / J. Haah [и др.] // Proceedings of the forty-eighth annual ACM symposium on Theory of Computing. — 2016. — С. 913—925.

101. Neural-network quantum state tomography / G. Torlai [и др.] // Nature physics. — 2018. — Т. 14, № 5. — С. 447—450.
102. Experimental neural network enhanced quantum tomography / A. M. Palmieri [и др.] // npj Quantum Information. — 2020. — Т. 6, № 1. — С. 20.
103. *Quek, Y.* Adaptive quantum state tomography with neural networks / Y. Quek, S. Fort, H. K. Ng // npj Quantum Information. — 2021. — Т. 7, № 1. — С. 105.
104. Quantum neural network states: A brief review of methods and applications / Z.-A. Jia [и др.] // Advanced Quantum Technologies. — 2019. — Т. 2, № 7/8. — С. 1800077.
105. Tensor network compressed sensing with unsupervised machine learning / S.-J. Ran [и др.] // Physical Review Research. — 2020. — Т. 2, № 3. — С. 033293.
106. Quantum process tomography with unsupervised learning and tensor networks / G. Torlai [и др.] // Nature Communications. — 2023. — Т. 14, № 1. — С. 2858.
107. *Valiant, L. G.* A theory of the learnable / L. G. Valiant // Communications of the ACM. — 1984. — Т. 27, № 11. — С. 1134—1142.
108. Robust shadow estimation / S. Chen [и др.] // PRX Quantum. — 2021. — Т. 2, № 3. — С. 030348.
109. Measurements of quantum Hamiltonians with locally-biased classical shadows / C. Hadfield [и др.] // Communications in Mathematical Physics. — 2022. — Т. 391, № 3. — С. 951—967.
110. *Zhao, A.* Fermionic partial tomography via classical shadows / A. Zhao, N. C. Rubin, A. Miyake // Physical Review Letters. — 2021. — Т. 127, № 11. — С. 110504.
111. *Levy, R.* Classical shadows for quantum process tomography on near-term quantum computers / R. Levy, D. Luo, B. K. Clark // Physical Review Research. — 2024. — Т. 6, № 1. — С. 013029.

112. Many-body chern number from statistical correlations of randomized measurements / Z.-P. Ciani [и др.] // Physical Review Letters. — 2021. — Т. 126, № 5. — C. 050501.
113. Shadow distillation: Quantum error mitigation with classical shadows for near-term quantum processors / A. Seif [и др.] // PRX Quantum. — 2023. — Т. 4, № 1. — C. 010303.
114. The randomized measurement toolbox / A. Elben [и др.] // Nature Reviews Physics. — 2023. — Т. 5, № 1. — C. 9—24.
115. Experimental learning of quantum states / A. Rocchetto [и др.] // Science advances. — 2019. — Т. 5, № 3. — eaau1946.
116. Experimental estimation of quantum state properties from classical shadows / G. Struchalin [и др.] // PRX quantum. — 2021. — Т. 2, № 1. — C. 010307.
117. Mixed-state entanglement from local randomized measurements / A. Elben [и др.] // Physical Review Letters. — 2020. — Т. 125, № 20. — C. 200501.
118. Experimental quantum state measurement with classical shadows / T. Zhang [и др.] // Physical Review Letters. — 2021. — Т. 127, № 20. — C. 200501.
119. Aaronson, S. Shadow tomography of quantum states / S. Aaronson // Proceedings of the 50th annual ACM SIGACT symposium on theory of computing. — 2018. — C. 325—338.
120. Rocchetto, A. Stabiliser states are efficiently PAC-learnable / A. Rocchetto // arXiv preprint arXiv:1705.00345. — 2017.
121. Density-matrix renormalization group algorithm for simulating quantum circuits with a finite fidelity / T. Ayrat [и др.] // PRX Quantum. — 2023. — Т. 4, № 2. — C. 020304.
122. Orús, R. Tensor networks for complex quantum systems / R. Orús // Nature Reviews Physics. — 2019. — Т. 1, № 9. — C. 538—550.
123. Simulation complexity of open quantum dynamics: Connection with tensor networks / I. Luchnikov [и др.] // Physical review letters. — 2019. — Т. 122, № 16. — C. 160401.

124. *Quiñones-Valles, D.* Tensor product approach to quantum control / D. Quiñones-Valles, S. Dolgov, D. Savostyanov // Integral Methods in Science and Engineering: Analytic Treatment and Numerical Approximations. — Springer, 2019. — C. 367—379.
125. Reconstructing complex states of a 20-qubit quantum simulator / M. K. Kurmapu [и др.] // PRX Quantum. — 2023. — Т. 4, № 4. — C. 040345.
126. Efficient quantum state tomography / M. Cramer [и др.] // Nature communications. — 2010. — Т. 1, № 1. — C. 149.
127. Efficient tomography of a quantum many-body system / B. P. Lanyon [и др.] // Nature Physics. — 2017. — Т. 13, № 12. — C. 1158—1162.
128. *Guo, Y.* Quantum state tomography with locally purified density operators and local measurements / Y. Guo, S. Yang // Communications Physics. — 2024. — Т. 7, № 1. — C. 322.
129. Scalable reconstruction of density matrices / T. Baumgratz [и др.] // Physical review letters. — 2013. — Т. 111, № 2. — C. 020401.
130. *Ahmed, S.* Gradient-descent quantum process tomography by learning Kraus operators / S. Ahmed, F. Quijandría, A. F. Kockum // Physical Review Letters. — 2023. — Т. 130, № 15. — C. 150402.
131. *Yu, N.* Learning marginals suffices! / N. Yu, T.-C. Wei // arXiv preprint arXiv:2303.08938. — 2023.
132. Process tomography on a 7-qubit quantum processor via tensor network contraction path finding / A. Dang [и др.] // arXiv preprint arXiv:2112.06364. — 2021.
133. *Akhtar, A. A.* Scalable and flexible classical shadow tomography with tensor networks / A. A. Akhtar, H.-Y. Hu, Y.-Z. You // Quantum. — 2023. — Т. 7. — C. 1026.
134. Online learning of quantum states / S. Aaronson [и др.] // Advances in neural information processing systems. — 2018. — Т. 31.
135. Efficient quantum mixed-state tomography with unsupervised tensor network machine learning / W.-j. Li [и др.] // arXiv preprint arXiv:2308.06900. — 2023.

136. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets / A. Kandala [и др.] // nature. — 2017. — Т. 549, № 7671. — С. 242—246.
137. Learning the tensor network model of a quantum state using a few single-qubit measurements / S. Kuzmin [et al.] // Physical Review A. — 2024. — Vol. 109, no. 5. — P. 052616.

Список рисунков

1.1	Строительный блок линейно-оптической схемы в стандартном дизайне — интерферометр Маха-Цандера.	21
1.2	Треугольный дизайн линейно-оптических схем, предложенный в работе [1].	22
1.3	Прямоугольный дизайн линейно-оптических схем, предложенный в работе [2].	23
1.4	Схема структуры блочного дизайна многомодового интерферометра.	25
1.5	Пример четырёхканального (четырёхмодового) интерферометра.	26
1.6	Представление линейно-оптической схемы Клементса в виде схемы с блочным дизайном.	27
1.7	Томография когерентными состояниями.	30
1.8	Томография фоковскими состояниями для определения модулей.	31
1.9	Томография фоковскими состояниями для определения фаз. . .	32
2.1	Задача обучения модели линейно-оптического интерферометра. .	36
2.2	Задача настройки модели линейно-оптического интерферометра.	37
2.3	Примеры обучения для различного количества оптических мод N .	43
2.4	Примеры обучения модели интерферометра и зависимость необходимого размера обучающей выборки от числа мод N . . .	44
2.5	Влияние несовершенства обучающего набора на обучение модели интерферометра.	46
2.6	Исследование процесса обучения с использованием априорных знаний о базисных матрицах U_ℓ	47
2.7	Схема структуры фотонного чипа.	49
2.8	Гистограммы распределения фиделити для тестирования модели на основе машинного обучения.	56
3.1	Схема структуры многомодового интерферометра.	60
3.2	Иллюстрация расчёта матричного элемента u_{ij} по уравнению (3.3).	62
3.3	Иллюстрация подхода послойного обучения линейной модели линейно-оптического интерферометра.	65

3.4	Зависимость метрики Фробениуса на тестовой выборке в зависимости от размера обучающей выборки M	68
3.5	Зависимость метрики Фробениуса на тестовой выборке в зависимости от размера обучающей выборки M	69
3.6	Зависимость метрики Фробениуса на тестовой выборке в зависимости от уровня шума ε в обучающей выборке.	70
3.7	Иллюстрация процесса настройки для полнослойного $(L = N + 1)$ линейно-оптического интерферометра.	71
3.8	Иллюстрация процесса настройки полнослойного линейного оптического интерферометра $(L = N + 1)$. На рисунке представлены значения функции потерь на 5000-й итерации для каждого запуска алгоритма настройки.	72
3.9	Иллюстрация процесса настройки для полнослойного $(L = N + 1)$ линейно-оптического интерферометра с послойной томографией, выполненного в условиях шума ε	73
4.1	Графическое представление разложения тензора $A_{i_1 \dots i_N}$ в тензорный поезд (4.7).	82
4.2	Графическое описание первой итерации алгоритма TT-SVD для разложения комплексного тензора $A_{i_1 \dots i_N}$ в тензорный поезд.	83
4.3	Основные архитектуры тензорных сетей.	85
4.4	Представление тензора $\psi_{i_1 \dots i_N}$, описывающего амплитуды разложения квантового состояния (4.21) по базисным состояниям вычислительного базиса N -кубитной системы в формате тензорного поезда (4.22).	90
4.5	Представление квантового состояния в формате MPS для пяти кубитов. При каждом фиксированном наборе значений индексов $i_1, \dots, i_N$ мы имеем совокупность матриц, произведение которых даёт число, зависящее от индексов $i_1, \dots, i_N$	91
4.6	Иллюстрация процедуры применения однокубитного гейта.	96
4.7	Иллюстрация процедуры применения двухкубитного гейта.	97
4.8	Иллюстрация процедуры применения алгоритма Truncated SVD.	99
4.9	Процедура левой ортонормализации n -го вагона тензорного поезда.	101

4.10	Графическое изображение 5-кубитного тензорного поезда, который ортонормализован с центром в точке между 3-им и 4-ым вагоном.	102
4.11	Графическое изображение первой и второй итерации алгоритма, позволяющего вычислить скалярное произведение $\langle \phi \psi \rangle$ двух квантовых состояний в формате тензорных поездов.	104
4.12	Графическое изображение алгоритма, вычисляющего матрицу плотности одного кубита с номером n	107
4.13	N -кубитная случайная квантовая схема.	108
4.14	Функция распределения вероятности для величины $p_x = \langle 0 ^{\otimes N} \psi_\chi \rangle ^2$ для квантовой схемы с $N = 15$ и $D = 24$	110
4.15	Эффективное двухкубитное фиделити f_n в зависимости от глубины D схемы.	112
4.16	Многокубитное фиделити F в зависимости от глубины D для $N = 20$ и различных значений максимального ТТ-ранга $\chi = 10, 20, 50$ для схемы с C_X -гейтами.	113
4.17	Сравнение многокубитного фиделити для C_X и C_Z гейтов для случаев без ортонормализации и с ней.	114
4.18	Изучение характеристик квантового состояния одного кубита в цепочке в ТТ-формате, параметры схемы $N = 10$ и $D = 10$	115
5.1	Схема генерации обучающего набора данных.	121
5.2	Графическое представление процедуры расчета вероятности по формуле (5.4) получения положительного результата в эксперименте с POVM-оператором $\hat{E}_i = \psi_i\rangle\langle\psi_i $ в формате тензорного поезда.	122
5.3	Обобщённые результаты численного моделирования процедуры обучения тензорного поезда.	126
5.4	Функция распределения вероятностных ошибок $\varepsilon_p = p_{\text{predict}} - p_{\text{exact}}$	127
5.5	Стандартный аппаратно-эффективный анзац, используемый для аппаратного обеспечения NISQ.	128
5.6	Сравнение точных распределений вероятностей с распределениями, сгенерированными обученной моделью.	129

5.7	Фиделити F между распределениями вероятностей, сгенерированными точным эталонным состоянием, и обученной моделью	130
-----	--	-----