

СТАТЬЯ

УДК 530.145

DOI 10.17513/snt.40296

**ПАРАЛЛЕЛЬНОЕ КОМПАКТНОЕ МОДЕЛИРОВАНИЕ
КВАНТОВЫХ СХЕМ****Тырышкин С.Ю.***ФГБОУ ВО «Алтайский государственный технический университет им. И.И. Ползунова»,
Барнаул, e-mail: service.vip-spe@yandex.ru*

Цель статьи заключается в рассмотрении особенностей и перспектив реализации параллельного компактного моделирования квантовых схем. Материалы и методы исследования – модели, методы, алгоритмы и процедуры параллельного синтеза и анализа специализированных логических схем, численные методы решения алгебраических уравнений, группировка, сравнение, абстракция, дедукция. Результаты: в статье описывается подход к параллельному компактному моделированию квантовых схем, который основан на моделировании равномерно-структурированных гамильтонианов. Отдельный акцент в процессе исследования сделан на изучении того, каким образом параллелизм может ускорить квантовое моделирование. В частности, рассмотрен параллельный квантовый алгоритм для моделирования динамики большого класса гамильтонианов. Кроме того, освещены общие вычислительные аспекты параллелизации с выделением особенностей компактного формата вычислений, который предполагает хранение и вычисление только ненулевых элементов для повышения эффективности моделирования. Для достижения максимальной пропускной способности ввода-вывода предложена техника управления памятью на основе наложения. Отдельный акцент сделан на методах оптимизации параллелизма для ускорения компактного моделирования квантовых схем. Выводы: стратегия параллельной декомпозиции задач для межзатворных и внутризатворных операций на основе разделенных блоков данных открывает дополнительные возможности оптимизации для моделирования квантовых схем.

Ключевые слова: квантовые вычисления, параллелизм, данные, ворота, оптимизация, буфер памяти, ввод-вывод, вычисления

PARALLEL COMPACT MODELING OF QUANTUM CIRCUITS**Tyryshkin S.Yu.***Altai State Technical University named after I.I. Polzunov, Barnaul,
e-mail: service.vip-spe@yandex.ru*

The aim of the article is to consider the peculiarities and prospects of realization of parallel compact modeling of quantum circuits. Material and methods of research – models, methods, algorithms and procedures of parallel synthesis and analysis of specialized logic circuits, numerical methods of solving algebraic equations, grouping, comparison, abstraction, deduction. Results: the paper describes an approach to parallel compact modeling of quantum circuits, which is based on modeling of uniformly structured Hamiltonians. A particular focus of the research process is to investigate how parallelism can accelerate quantum simulations. In particular, we consider a parallel quantum algorithm for modeling the dynamics of a large class of Hamiltonians. In addition, general computational aspects of parallelization are highlighted, highlighting the features of a compact computation format that involves storing and computing only non-zero elements to improve simulation efficiency. An overlay-based memory management technique is proposed to maximize I/O throughput. Separate emphasis is placed on parallelism optimization techniques to accelerate compact simulation of quantum circuits. Conclusions: a parallel task decomposition strategy for inter-gate and intra-gate operations based on separated data blocks opens up additional optimization opportunities for quantum circuit simulations.

Keywords: quantum computing, parallelism, data, gates, optimization, memory buffer, I/O, computation

Введение

На протяжении последних нескольких лет поддерживать темпы выполнения закона Мура становится сложной и отнюдь не тривиальной задачей. Отдавая должное различным инновациям, которые уже сегодня позволяют удовлетворить высокие требования к мощности и производительности экзамасштабных систем, в научно-экспертном сообществе прогнозируется, что альтернативный подход даст возможность обеспечить устойчивый рост вычислительной мощности за пределами экзамасштаба. По мнению исследователей, одной из таких

альтернатив являются квантовые вычисления (КВ), которые обладают потенциалом экспоненциального ускорения для некоторых типов вычислений [1]. В этой ситуации, несомненно, доминирует цифровая модель КВ, основанная на вентилях, и, хотя существуют теоретические гарантии того, что другие конкурирующие режимы, например адиабатические КВ, не менее эффективны, они остаются гораздо более нишевой технологией.

Следует отметить, что КВ – это в большей степени комплексный термин, который охватывает ряд различных архитектур

турных, логических и вычислительных подходов для решения некоторых типов задач, которые классически неразрешимы. Среди широкого спектра приложений КВ классическое моделирование необходимо исследователям в целях разработки оптимизированных квантовых алгоритмов, проверки функциональности квантовых схем и более четкого понимания квантовых операций [2]. В то же время следует отметить, что моделирование квантовых схем – это достаточно сложная и трудновыполнимая задача, поскольку предполагает необходимость проведения больших вычислений, требует наличия значительных объемов памяти для оценки амплитуд состояний кубитов. Например, чтобы реализовать полное моделирование n -кубитной схемы, требуется экспоненциальный размер вектора для хранения 2^n амплитуд [3]. Для решения данной проблемы ученые активно изучают сжатие данных, рассматривают возможности оптимизации схем и перспективы проведения параллельных вычислений.

В данном контексте особого внимания заслуживает симулятор квантовых схем на основе вектора состояния, который эффективно поддерживает инкрементальность с помощью параллелизма задач. Параллелизм данных хорошо работает для высокорегулярных вычислений, которые можно представить как операции над плотными матрицами. Центральным моментом данного предложения является разделение квантовой системы на ряд подэлементов, которые слабо запутаны и могут моделироваться параллельно на разных схемах. Влияние различных областей системы друг на друга можно суммировать эффективным состоянием на гораздо меньшем количестве квантовых битов. Кроме того, параллелизм данных позволяет разделить вычислительные задачи и данные на не взаимодействующие подмножества, что значительно снижает накладные расходы на взаимодействие с внешним окружением. Но в то же время, к примеру, параллелизм с трудом справляется с нерегулярными вычислениями, где объем работы может варьировать между различными «итерациями», подлежащими распараллеливанию.

В связи с этим дальнейшие исследования стратегий параллельного моделирования квантовых схем и перспектив их расширения на квантовые процессоры составляют важную научно-практическую задачу, которая и предопределила выбор темы данной статьи.

Один из первых примеров параллельно-симулятора в научной литературе пред-

ставлен в работах Francisco Pasadas и Pedro C. Feijoo [4]. Он был реализован на языке C и давал возможность провести распараллеливание с помощью интерфейса передачи сообщений. Отличительной характеристикой данного симулятора является использование терминов экспериментальных параметров затвора для реализации лазерных импульсов в ионных ловушках.

Ряд наработок, касающихся создания инструментальной цепочки для симулятора, которая дает возможность выполнять синтезированные квантовые схемы и запускать открытый стандартный квантовый ассемблерный код, а также проводить сверхглубокие и крупномасштабные симуляции, представлен в трудах Sukin Sim, Peter D. Johnson, Alán Aspuru-Guzik [5]. Стоит также упомянуть работы, где описывались специализированные симуляторы, среди которых выделяются те, которые основаны на тензорных сетях, такие как ExaTN, QTensor и ITensor [6, 7].

Вопросы, связанные с возможностями параллельного выполнения во многих потоках КВ на основе графических процессоров (GPU), а также ключевые аспекты новой методологии программирования для нескольких GPU под названием MG-BSP, которая строит виртуальную BSP-машину поверх современных платформ, рассматривают Bakr Ahmed Taha, Ali J. Addie, Adawiya J. Haider [8].

В то же время при высокой оценке имеющихся на сегодняшний день исследований и наработок некоторые проблемные вопросы, связанные с возможностями параллельных вычислений, остаются открытыми. Так, отдельного внимания заслуживают проблемы с объемом памяти, которые препятствуют расширению моделирования для охвата большего числа кубитов. Помимо этого, влияние квантования с фиксированной точкой на квантовые схемы с увеличивающимся числом кубитов требует дополнительного исследования для всесторонней оценки.

Цель исследования – рассмотреть особенности и перспективы реализации параллельного компактного моделирования квантовых схем.

Материал и методы исследования

Материал и методы исследования – модели, методы, алгоритмы и процедуры параллельного синтеза и анализа специализированных логических схем, численные методы решения алгебраических уравнений, группировка, сравнение, абстракция, дедукция.

**Результаты исследования
и их обсуждение**

Одной из первых работ в направлении параллельного квантового моделирования является параллельная квантовая схема $O(\log n)$ -глубины Клива и Уотруса для n -кубитного квантового преобразования Фурье, которая может быть использована для распараллеливания алгоритма факторизации Шора с поливременной классической предварительной и последующей обработкой [9]. Кроме того, учеными был предложен квантовый алгоритм с постоянной квантовой глубиной, который позволяет провести оценку многомерных следов. В то же время следует отметить, что изыскания в направлении параллельных квантовых вычислений не ограничиваются только схемной моделью. Например, в квантовых вычислениях, основанных на измерениях, было замечено, что параллелизм способен обеспечить больше преимуществ, чем в модели схемы. Другой параллельной

моделью, которая представляется более близкой к современному квантовому оборудованию, являются распределенные КВ, которые имеют все возможности эффективно моделировать квантовую схемную модель с низкой глубиной накладных расходов [10]. Параллелизм также изучается на более абстрактных уровнях, таких как квантовое программирование.

Параллелизация

Определим параллельное квантовое прохождение, которое может быть реализовано квантовой схемой постоянной глубины с запросами к равномерно структурированным гамильтонианам. Пусть $j \in H^r$, если в графе H существует путь $j = (j_0, \dots, j_r)^{12}$ длины $r + 1$.

Предположим, что $\mathcal{H} = \mathcal{H}^A \otimes \mathcal{H}^B$ – пространство состояний, где

$$\mathcal{H}^A = \mathcal{H}^B = (\mathbb{C}^{2N})^{\otimes(r+1)}.$$

Для каждого $j_0 \in [N]$ определим:

$$|\psi_{j_0}^{(r)}\rangle = \frac{1}{\sqrt{d^r}} \sum_{\substack{j \in H^r \\ \in \mathcal{H}^A}} |j\rangle \otimes |j_0\rangle \underbrace{\otimes_{s \in [r]} \left(\sqrt{H_{j_s j_{s+1}}^*} |j_{s+1}\rangle + \sqrt{1 - |H_{j_s j_{s+1}}|} |j_{s+1} + N\rangle \right)}_{\in \mathcal{H}^B} \quad (1)$$

где $|j\rangle = |j_0\rangle \dots |j_r\rangle \in \mathcal{H}^A$.

Пусть:

$T^{(r)} : \mathcal{H} \rightarrow \mathcal{H}$ любой унитарный оператор, такой, что:

$$T^{(r)}(|j\rangle|z_1\rangle \dots |z_{2r+1}\rangle) = \begin{cases} |\psi_j^{(r)}\rangle & j \in [N], z_1 = \dots = z_{2r+1} = 0 \\ \text{любое состояние} & \end{cases} \quad (2)$$

для всех $j, z_1, \dots, z_{2r+1} \in [2N]$.

$S^{(r)} : \mathcal{H} \rightarrow \mathcal{H}$ оператор обратного порядка:

$$S^{(r)}|a_0, \dots, a_{2r+1}\rangle = |a_{2r+1}, \dots, a_0\rangle \quad (3)$$

для всех $a_s \in [2N], s \in [2r + 2]$.

В данном случае шаг r -параллельного квантового прохода для H вычисляется следующим образом: $Q^{(r)} := T^{(r)\dagger} S^{(r)} T^{(r)}$.

Ключевая идея параллельного квантового прохождения заключается в том, что происходит расширение состояния $|\Psi_j\rangle$, которое является суперпозицией одного шага движения $(j, k) \in H$, до состояния $|\psi_{j_0}^{(r)}\rangle$, которое представляет собой суперпозицию r шагов движения (т.е. пути) $j \in H^r$. Оператор движения $Q^{(r)}$ становится блок-кодом монома $(H/d)^r$, обобщением H/d , полученного из исходного квантового движения. В этом смысле оператор движения $Q^{(r)}$ явля-

ется расширением $T^r Q T$ вместо Q . Следует подчеркнуть, что r -параллельное квантовое прохождение не эквивалентно r последовательным шагам исходного квантового прохода, который вместо этого кодирует полином Чебышева $T_r(H/d)$.

Набор квантовых ворот, находящихся на одном уровне, создает унитарную матрицу преобразования, которая определяется произведением Кронекера (\otimes) отдельных матриц ворот от первого кубита до n -го кубита [6]. На рисунке 1 показана пятикубитная схема с пятью воротами Хадамарда и четырьмя воротами CNOT.

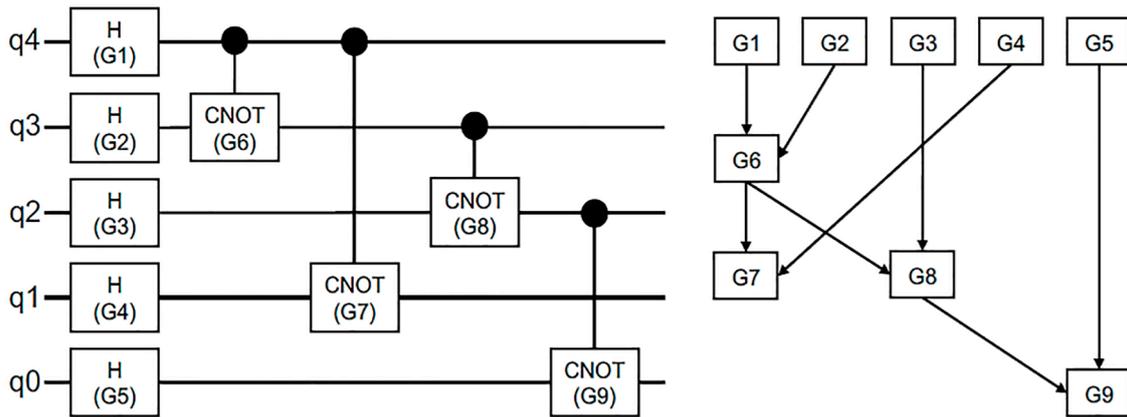


Рис. 1. Пятикубитовая квантовая схема из девяти вент (слева) и ее граф зависимости от вент (справа) [6]

Первые пять хадамардовских вент создают матрицу преобразования 32×32 , $H^{\otimes 5}$, что позволяет получить суперпозицию. Вент CNOT формируют запутанность. Поиск унитарных матриц является обязательным элементом моделирования квантовых схем. Вначале осуществляется упорядочивание вент слева направо и заполняется пустое место матрицей тождества соответствующей размерности. В свою очередь, параллельные вент могут быть упорядочены произвольно, например G7 и G8.

Общие вычислительные аспекты параллелизации

Время процессора и имеющиеся объемы памяти формируют комплекс ограничений на размер квантового компьютера, который может быть смоделирован на обычном цифровом компьютере. Необходимое время процессора определяется числом операций, которые требуется реализовать с кубитами. Состояние L -кубитного кван-

тового компьютера служит комплексно-значимым вектором длины $D=2^L$. Принимая во внимание потенциально большое количество арифметических операций, которые необходимо осуществить, целесообразно использовать 13–15-разрядную арифметику с плавающей точкой (что соответствует 8 байтам для вещественного числа) [7]. Таким образом, для представления состояния квантовой системы из L кубитов в обычном цифровом компьютере потребуется не менее 2^{L+4} байт. Следовательно, объем памяти, который требуется для создания модели квантового компьютера с L кубитами, возрастает экспоненциальными темпами с числом кубитов L . Например, для $L=24$ требуется не менее 256 МБ памяти для хранения одного произвольного состояния $|\Phi\rangle$.

Операции U над вектором состояния $|\Phi\rangle$ приводят к преобразованию амплитуд базисных состояний в $|\Phi\rangle$. Более конкретно это может быть представлено следующим образом [7]:

$$|\Phi\rangle = a(00\dots 0)|00\dots 0\rangle + a(0\dots 01)|0\dots 01\rangle + \dots + a(01\dots 1)|0\dots 01\rangle + a(11\dots 1)|11\dots 1\rangle \quad (4)$$

$$|\Phi'\rangle = U|\Phi\rangle = a'(00\dots 0)|00\dots 0\rangle + a'(0\dots 01)|0\dots 01\rangle + \dots + a'(01\dots 1)|0\dots 01\rangle + a'(11\dots 1)|11\dots 1\rangle \quad (5)$$

Рассмотрим более подробно одноквантовые операции на кубите j , которые преобразуют $|\Phi\rangle$ в $|\Phi'\rangle = U_j|\Phi\rangle$. Операция Хадамарда на кубите j , H_j преобразует амплитуды в соответствии с:

$$a'(*\dots *0_j*\dots *) = \frac{1}{\sqrt{2}}(a(*\dots *0_j*\dots *) + a(*\dots *1_j*\dots *)) \quad (6)$$

$$a'(*\dots *1_j*\dots *) = \frac{1}{\sqrt{2}}(a(*\dots *0_j*\dots *) - a(*\dots *1_j*\dots *)) \quad (7)$$

В формулах 6, 7 звездочка обозначает, что биты в соответствующих позициях одинаковы. Для X_j , действующего на $|\Phi\rangle$, элементы $|\Phi'\rangle$ получаются следующим образом:

$$a'(* \dots * 0_j * \dots *) = \frac{1}{\sqrt{2}}(a(* \dots * 0_j * \dots *) + ia(* \dots * 1_j * \dots *)) \quad (8)$$

$$a'(* \dots * 1_j * \dots *) = \frac{1}{\sqrt{2}}(a(* \dots * 1_j * \dots *) - ia(* \dots * 0_j * \dots *)) \quad (9)$$

В случае $|\Phi'\rangle = Y_j |\Phi\rangle$ следует, что:

$$a'(* \dots * 0_j * \dots *) = \frac{1}{\sqrt{2}}(a(* \dots * 0_j * \dots *) + a(* \dots * 1_j * \dots *)) \quad (10)$$

$$a'(* \dots * 1_j * \dots *) = \frac{1}{\sqrt{2}}(a(* \dots * 1_j * \dots *) - a(* \dots * 0_j * \dots *)) \quad (11)$$

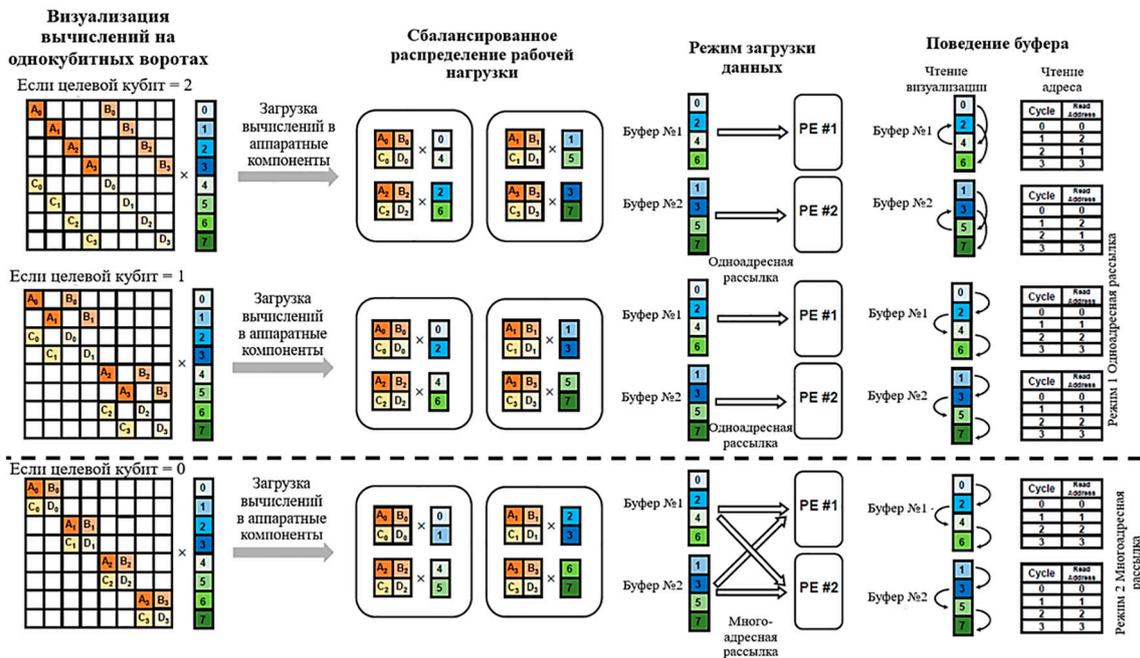


Рис. 2. Два режима (одноадресный и многоадресный режимы загрузки данных) для поддержания баланса рабочей нагрузки

В целом, можно отметить, что выполнение операции над кубитом j требует обновления 2^L элементов $|\Phi\rangle$. Исключение составляет операция сдвига фазы на одном кубите, которая требует обновления только 2^{L-1} единичных амплитуд. Необходимо сделать акцент на том, что все эти операции могут быть выполнены на месте, то есть без использования другого вектора длины 2^L .

Компактный формат вычислений

При использовании описания вектора состояний применение ворот U к состоянию $|\psi\rangle$ позволяет без излишних затрат и с допустимой скоростью выполнить матрично-векторное

умножение $U|\psi\rangle$. Например, применение однокубитных ворот $\begin{bmatrix} u_{00} & u_{01} \\ u_{10} & u_{11} \end{bmatrix}$ к состоянию $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ дает состояние:

$$\begin{bmatrix} u_{00} & u_{01} \\ u_{10} & u_{11} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} u_{00}\alpha & u_{01}\beta \\ u_{10}\alpha & u_{11}\beta \end{bmatrix} \quad (12)$$

Далее можно распространить (12) на многокубитные случаи, в результате получаем матрично-векторное представление, которое отображено в левой части на рисунке 2 (применение однокубитных ворот к трехкубитовому вектору состояний с различными целевыми кубитами).

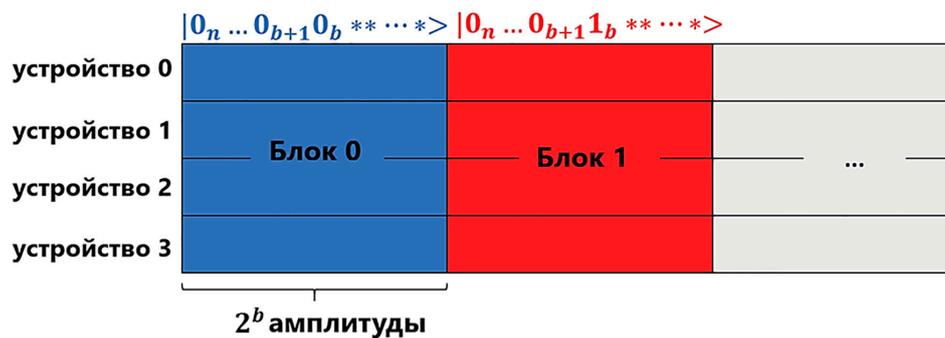


Рис. 3. Схема параллельной компоновки хранилища

Целевой кубит в квантовых воротах отражает запутанность между элементами внутри вектора состояния. Учитывая, что большинство квантовых ворот в обычных схемах являются однокубитными, то есть между большинством элементов нет запутывания, соответствующее матричное представление имеет большое количество нулевых элементов. Для повышения эффективности представляется целесообразным хранить и вычислять только ненулевые элементы, что является, по сути, компактным форматом вычислений.

Направления оптимизации параллелизма для ускорения компактного моделирования квантовых схем

Параллелизм ввода-вывода. Для повышения производительности ввода-вывода эффективным методом может быть организация устройств хранения данных аналогично конфигурации RAID-0, но с большим размером блока (stripe) (рис. 3). При наличии d устройств хранения с параметром размера блока b каждый блок из $2b$ амплитуд может быть распределен между d устройствами. Каждое устройство содержит $2b/d$ амплитуд на блок. После этого алгоритм, используя эту схему и библиотеку асинхронного ввода-вывода, распараллеливает операции ввода-вывода.

Размер любой операции ввода-вывода больше или равен 2^{b+4} байтам. Параметр b является настраиваемым, однако есть ограничение относительно производительности. Принимая во внимание тот факт, что каждая отдельная операция ввода-вывода осуществляется с гранулярностью в размер страницы, каждое устройство должно содержать более 2^8 амплитуд на блок, и, следовательно, $b \geq d \cdot 2^8$.

Балансировка нагрузки между подсхемами. Чтобы уменьшить количество операций ввода-вывода и связи, каждая подсхема должна иметь достаточно вычислений. С целью сбалансирования вычислительной

нагрузки между подсхемами необходимо рассчитать среднюю интенсивность работы всех подсхем. Интенсивность работы – это отношение количества операций с плавающей точкой к объему памяти. Учитывая изложенное, следует стремиться к тому, чтобы каждая из подсхем имела интенсивность работы, близкую к средней, путем перемещения ворот между подсхемами.

Перекрытие вычислений и ввода-вывода. Чтобы уменьшить большие накладные расходы на ввод-вывод хранилища, целесообразным является реализовывать масштабирование квантовой симуляции цепей с использованием трех основных буферов состояния памяти. Благодаря этому во время выполнения операций над одним буфером предварительно считывается другой буфер из пространства хранения и оставшийся буфер сбрасывается в пространство хранения.

Заключение

В статье рассматриваются вопросы, касающиеся возможностей параллелизма ускорить квантовое моделирование. Их актуальность обусловлена тем, что классическое моделирование больших квантовых систем является сложной задачей из-за экспоненциального увеличения пространства и требований к вычислениям.

В статье описывается подход к параллельному компактному моделированию квантовых схем, который основан на моделировании равномерно-структурированных гамильтонианов. Для достижения максимальной пропускной способности ввода-вывода предложена техника управления памятью на основе наложения. Отдельный акцент сделан на методах оптимизации параллелизма для ускорения компактного моделирования квантовых схем. Метод разделения данных адаптирован для разделения вычислительных задач и данных на взаимодополняющие подмножества, это позволяет на практике существенно обра-

зом сократить накладные расходы на обмен с внешними данными.

Список литературы

1. Потапов В.С., Гушанский С.М. Разработка методики построения квантовой системы с использованием аппаратных средств оптимизации // Информационные технологии и вычислительные системы. 2022. № 1. С. 26-32. DOI: 10.14357/20718632220103.
2. Hanxu Zhang, Tian Chen, Naiqiao Pan, Xiangdong Zhang Electric-Circuit Simulation of Quantum Fast Hitting with Exponential Speedup // Advanced Quantum Technologies. 2022. Vol. 5, Is. 4. P. 77-83. DOI: 10.1002/qute.202100143.
3. Amine Zeguendry, Zahi Jarir, Mohamed Quafafou. Quantum-Enhanced K-Nearest Neighbors for Text Classification: A Hybrid Approach with Unified Circuit and Reduced Quantum Gates // Advanced Quantum Technologies. 2024. Vol. 7, Is. 11. P. 107-113. DOI: 10.1002/qute.202400122.
4. Pasadas F., Feijoo P.C. Compact Modeling Technology for the Simulation of Integrated Circuits Based on Graphene Field-Effect Transistors // Advanced Materials. 2022. Vol. 34, Is. 48. P. 87-98. DOI: 10.1002/adma.202201691.
5. Sukin Sim, Peter D. Johnson Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms // Advanced Quantum Technologies. 2019. Vol. 2, Is. 12. P. 66-71. DOI: 10.1002/qute.201900070.
6. Tsung-Wei Huang qTask: Task-parallel Quantum Circuit Simulation with Incrementality // IEEE International Parallel and Distributed Processing Symposium (IPDPS). 2023. DOI: 10.48550/arXiv.2210.01076.
7. Raedt K. De, Michielsen K., Raedt H. De., Trieu B, Arnold G., Richter M., Lippert Th., Watanabe H., Ito N., Massively parallel quantum computer simulator // Computer Physics Communications. 2007. Vol. 176, Is. 2. P. 121-136. DOI: 10.1016/j.cpc.2006.08.007.
8. Taha B.A., Ali J. Addie Exploring Trends and Opportunities in Quantum-Enhanced Advanced Photonic Illumination Technologies // Advanced Quantum Technologies. 2024. Vol. 7, Is. 3. P. 120-129. DOI: 10.1002/qute.202300414.
9. Гушанский С.М., Мушаев А.Я. Анализ эффективности способов моделирования квантовых вычислений на основе экспериментальных данных // Информатизация и связь. 2022. № 2. С. 35-38. DOI: 10.34219/2078-8320-2022-13-2-35-38.
10. Jiang G.-L., Liu W.-Q., Wei H.-R. Optimal Quantum Circuits for General Multi-Qutrit Quantum Computation // Advanced Quantum Technologies. 2024. Vol. 7, Is. 7. P. 65-72. DOI: 10.1002/qute.202400033.