

Отзыв

официального оппонента на диссертационную работу Афанасьева Ильи Викторовича «ИССЛЕДОВАНИЕ И РАЗРАБОТКА МЕТОДОВ ЭФФЕКТИВНОЙ РЕАЛИЗАЦИИ ГРАФОВЫХ АЛГОРИТМОВ ДЛЯ СОВРЕМЕННЫХ ВЕКТОРНЫХ АРХИТЕКТУР», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Актуальность выбранной темы

Работа посвящена исследованию эффективных реализаций графовых алгоритмов для современных векторных архитектур с быстрой памятью. С одной стороны, на сегодняшний день эти проблемы проработаны недостаточно глубоко. Главная причина заключается в необходимости использования регулярной по своей природе векторной обработки данных для нерегулярных структур таких, как графы. С другой стороны, графы широко используются для моделирования многих объектов реального мира из различных прикладных областей.

В настоящее время для векторных систем не предложено подходов к эффективной реализации графовых алгоритмов, а также не существует специализированных библиотек для решения графовых задач. Известные библиотечные реализации для многоядерных центральных процессоров на векторных системах имеют низкую производительность. Реальная производительность существующих реализаций на графовых задачах при использовании графических ускорителей крайне низка.

Данная диссертация посвящена вопросам создания эффективных реализаций графовых задач для современных векторных архитектур с быстрой памятью. Решение этих проблем востребовано практикой. При этом открываются возможности создания программных комплексов для решения графовых задач на современных векторных системах с быстрой памятью.

Эти аспекты исследования методов эффективных реализаций графовых задач определяют актуальность диссертации.

Содержание диссертации

Диссертация состоит из введения, пяти глав, заключения, и списка литературы.

Первая глава посвящена выбору целевых архитектур, позволяющих существенно ускорить решение исследуемого в работе класса графовых задач, а также исследованию существующих подходов к реализации графовых алгоритмов на выбранных архитектурах.

Выделены основные свойства графовых алгоритмов, а именно, большое количество обменов с оперативной памятью при малом числе арифметических операций, нерегулярность и потенциально высокий уровень параллелизма.

Проведен анализ основных современных архитектур с быстрой памятью. На основе анализа выделен такой класс систем, как векторные архитектуры с быстрой памятью, предназначенных для ускорения решения графовых задач.

Выделены фундаментальные задачи для эффективного решения на векторных системах.

Рассмотрены существующие подходы к реализации и оптимизации графовых алгоритмов для современных многоядерных центральных процессоров, графических ускорителей NVIDIA GPU и векторных систем.

Проведен анализ методов, позволяющих оценить производительность, локальность и эффективность разрабатываемых реализаций графовых алгоритмов (roofline-модель).

Приведены основные характеристики синтетических графов и графов для реальных приложений, которые в ходе исследования используются для оценки производительности, эффективности и энергоэффективности разрабатываемых реализаций.

Вторая глава содержит результаты исследования зависимостей между различными представителями выделенного класса векторных систем с быстрой памятью и возможностей эффективной реализации на них графовых алгоритмов.

Для векторных архитектур и графических ускорителей NVIDIA выделено большое число схожих особенностей организации потока управления, вычислений, а также работы с подсистемой памяти, большая часть которых обусловлена использованием векторного подхода к обработке данных в обоих типах систем. Это позволяет для разработанных реализаций иметь высокую степень переносимости между различными целевыми архитектурами.

Рассматриваются различные классы приложений, имеющие как сравнимую, так и принципиально различную эффективность для векторных архитектур и графических ускорителей NVIDIA, в частности, бенчмарки, операции над плотными и разреженными матрицами, алгоритмы сортировки. При этом эффективность определяется как доля использования пиковых аппаратных характеристик каждой из архитектур – производительности или пропускной способности памяти в зависимости от свойств приложения.

Третья глава посвящена разработке и исследованию метода создания эффективных реализаций графовых алгоритмов для векторных систем с быстрой памятью на основе выделения типовых алгоритмических абстракций.

Анализ информационных графов позволяет выделить типовые алгоритмические структуры графовых алгоритмов, которые могут быть эффективно реализованы на векторных системах с быстрой памятью.

В исследуемых алгоритмах выделены так называемые типовые алгоритмические структуры, каждая из которых представляет собой подграф в информационном макрографе алгоритма, изоморфный для всех исследуемых алгоритмов при условии их применения к одним и тем же входным данным.

Введено понятие алгоритмического шаблона – группы из элементарных операций в информационном графе, применяемых в рамках выделенных типовых алгоритмических структур для обработки различных вершин и ребер входных графов.

Показано, что алгоритмы решения всех рассматриваемых в работе графовых задач обладают четырьмя типовыми алгоритмическими структурами.

Для обозначения совокупности выделенных информационных структур и соответствующим им подходов к реализации для векторных архитектур с быстрой памятью вводится понятие алгоритмической абстракции, представляемой соответствующим макрографом.

Первая выделенная алгоритмическая абстракция - применение заданного алгоритмического шаблона к каждому из ребер графа, смежных некоторому подмножеству вершин.

Вторая - применение алгоритмического шаблона к каждой из вершин заданного подмножества вершин графа.

Третья абстракция представляет собой процедуру создания заданного подмножества вершин графа.

Наконец, четвертой алгоритмической абстракцией является процедура применения алгоритмического шаблона к каждой из вершин заданного подмножества вершин графа, с последующей аккумуляцией некоторых значений, вычисленных для каждой из вершин.

Вводится понятие типовой абстракции данных. Выделенные алгоритмические абстракции оперируют с тремя типами абстракций данных.

Предложены подходы к эффективной реализации выделенных абстракций данных для векторных систем с быстрой памятью. Разработан векторно-ориентированный формат хранения графа.

Рассмотрены пути эффективной реализации и оптимизации выделенных алгоритмических абстракций для векторных систем с быстрой памятью.

Приведены оценки эффективности реализованных абстракций для различных рассматриваемых векторных архитектур с быстрой памятью, полученные на основе гоofline-модели. На основе предложенных метрик даны оценки эффективности для всех четырех реализованных абстракций.

Замеренные значения динамических характеристик реализованных абстракций, подтверждают их значительную эффективность для целевых векторных архитектур в смысле выполнения транзакций к памяти, использования векторных инструкций (нитей варпа), загрузки GPU.

Разработан метод создания реализаций графовых алгоритмов для векторных систем с быстрой памятью.

В четвертой главе описан разработанный программный комплекс для создания архитектурно-независимых реализаций графовых алгоритмов.

Он включает в себя набор вычислительных абстракций и структур данных, используя которые пользователи могут реализовывать подклассы графовых алгоритмов, самостоятельно определяя несложные последовательные фрагменты кода, описывающие основную вычислительную логику графовых алгоритмов.

Проведено сравнение производительности оптимизированных «вручную» реализаций алгоритмов поиска кратчайших путей и поиска в ширину с реализациями, разработанными на основе предложенного программного комплекса.

Пятая глава содержит результаты анализа производительности, эффективности и энергоэффективности реализаций на основе разработанного программного комплекса, а также сравнения его с существующими аналогами, в частности, для решения задач ранжирования вершин в графе и поиска связанных компонент.

Проведенное сравнение показывает, что реализации на основе разработанного комплекса для векторных систем с быстрой памятью обеспечивают существенное ускорение по сравнению с существующими библиотечными аналогами для многоядерных процессоров (от 3 до 14 раз) и аналогами для графических ускорителей (от 1.5 до 3 раз).

Даны оценки, демонстрирующие значительно более высокую энергоэффективность реализаций на основе разработанного комплекса по сравнению с аналогами для многоядерных центральных процессоров.

Сама работа написана хорошим языком с четкой логикой в изложении. Автореферат полно отражает содержание диссертационной работы, ее структуру, положения, выводы.

Научная новизна, обоснование, достоверность и практическая ценность полученных научных положений, выводов и рекомендаций

Новизна научных положений и выводов диссертации определяется тем, что решена задача разработки методов, позволяющих создавать эффективные реализации графовых алгоритмов для современных векторных архитектур с быстрой памятью.

Автором получен ряд новых научных результатов, к числу которых можно отнести следующие.

1. Для широкого класса графовых алгоритмов выделен компактный набор типовых алгоритмических структур, которые возможно эффективно реализовать на векторных архитектурах с быстрой памятью.

2. Предложен метод, позволяющий создавать эффективные реализации графовых алгоритмов для современных векторных архитектур с быстрой памятью.

3. Разработаны реализации фундаментальных графовых алгоритмов, имеющие существенно более высокую производительность и эффективность по сравнению с существующими библиотечными аналогами для многоядерных центральных процессоров и графических ускорителей NVIDIA.

Практическая значимость результатов работы состоит в разработке и реализации архитектурно-независимого программного комплекса для решения графовых задач. Комплекс позволяет создавать эффективные реализации широкого класса графовых алгоритмов для современных векторных систем с быстрой памятью.

Обоснованность и достоверность положений и выводов диссертации подтверждается соответствием теоретических и экспериментальных результатов, полученных для синтетических графов и графов для реальных приложений, при исследовании алгоритмов для современных векторных архитектур с быстрой памятью – NEC SX-Aurora TSUBASA, NVIDIA GPU, Intel KNL.

Замечания по диссертационной работе

1. В главе 3 (раздел 3.2) вводится понятие типовых алгоритмических структур. На уровне макрографа они изоморфны между собой для любых двух рассматриваемых графовых алгоритмов, при условии их применения к одним и тем же входным данным. Известно, что задача поиска изоморфного подграфа является обобщением задач о максимальной клике и задачи о проверке, не содержит ли граф гамильтонов цикл, и является NP-полной. Из работы не ясно, как устанавливается изоморфизм типовых алгоритмических структуры на уровне информационных макрографов.

2. В работе следовало больше внимания уделить графовым алгоритмам для реальных приложений. Тогда было бы понятнее, о какой нерегулярности структур данных, каковыми являются графы, и о каких приложениях идет речь.

3. В работе автор проводит экспериментальные исследования для вычислительных систем, построенных на основе графических процессоров NVIDIA P100 и V100, но нет данных экспериментов для последних модификаций процессоров A100. Не ясно, будут ли положения верны и для этой новой архитектуры.

Заключение

В диссертационной работе Афанасьева Ильи Викторовича решены задачи создания эффективных реализаций графовых задач для современных векторных архитектур с быстрой памятью.

Отмеченные выше недостатки, в целом, не снижают высокого уровня работы.

Основные результаты диссертации полно отражены в публикациях автора, в том числе в 9 публикациях в рецензируемых научных изданиях, определенных п. 2.3 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова.

Диссертация Афанасьева Ильи Викторовича является законченной научно-квалификационной работой и отвечает требованиям, предъявляемым к кандидатским диссертациям.

Автор заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Официальный оппонент
доктор технических наук, профессор,
заведующий кафедрой Вычислительных технологий
Национального исследовательского университета «МЭИ»

В.В. Топорков

« 8 » декабря 2020 года

Контактные данные:
Телефон: 8(495)362-71-45
Электронная почта: ToporkovVV@mpei.ru

Специальность, по которой официальным оппонентом
защита диссертация:
05.13.15 – «Вычислительные машины, комплексы и компьютерные
сети»

Адрес места работы:
111250, Москва, ул. Красноказарменная, д. 14,
Федеральное государственное бюджетное образовательное учреждение
высшего образования Национальный исследовательский университет
«МЭИ»
Телефон: 8(495)362-71-45
Электронная почта: ToporkovVV@mpei.ru