

«УТВЕРЖДАЮ»

Директор Федерального государственного
бюджетного учреждения науки Институт
автоматики и электрометрии Сибирского отделения
Российской академии наук, чл.-корр. РАН



Бабин Сергей Алексеевич

«19» августа 2022 г.

ОТЗЫВ

ведущей организации на диссертацию **Снытниковой Татьяны Валентиновны** на тему: **«Эффективная реализация модели ассоциативных вычислений на графических ускорителях для решения задач на графах»**, представленную на соискание ученой степени кандидата технических наук по специальности 2.3.5 - «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Актуальность темы диссертации.

Цифровая информация стремительно растет, поэтому обеспечение быстрого поиска и обработка информации, особенно в режиме реального времени, становятся очень важной задачей. Примером может служить БАК, детектор ATLAS, на который каждую секунду поступает порядка сотни ТВ данных. И никакая машина фон-неймановского типа не может справиться с такой работой, только ассоциативные машины дают такую возможность. Ключевой момент в том, что не только поиск, но и обработку информации можно делать с помощью ассоциативных процессоров. В разработке ассоциативных вычислений существует три направления. Во-первых, разработка аппаратного обеспечения: как устройства ассоциативной памяти, так и чипы ассоциативных процессоров. Второе направление — это, соответственно, разработка ассоциативных моделей и алгоритмов для этих моделей. И, наконец, третье направление — это реализация ассоциативных моделей на существующем оборудовании.

Диссертационная работа Т.В. Снытниковой посвящена созданию программных средств для эффективной реализации ассоциативной модели вычислений на графических ускорителях с использованием технологии CUDA.

В силу вышеизложенного, актуальность темы диссертационной работы Снытниковой Т. В. не вызывает сомнений.

Целью диссертации является построение эффективной реализации STAR-машины на графических ускорителях с использованием технологии CUDA. Это позволяет использовать на практике ассоциативные параллельные алгоритмы, разработанные для этой модели.

Научная новизна полученных в диссертации результатов заключается в реализации модели ассоциативных вычислений на графических ускорителях. Реализация эффективно сохраняет ассоциативные свойства и рассчитана на выполнение ассоциативных алгоритмов модели, а не прогнозирование ее свойств. Предложена реализация ассоциативной модели вычислений на неассоциативной архитектуре является уникальной.

Методы оптимизации ассоциативных алгоритмов для выполнения на графических ускорителях, учитывающие архитектурные различия STAR-машины и GPU. Оптимизация алгоритмов направлена на локализацию точек синхронизации, что значительно сокращает трудозатраты разработчиков.

Также для данной реализации разработаны ассоциативные параллельные версии динамических алгоритмов решения задач теории графов, для которых нет других параллельных версий.

Теоретическая и практическая значимость

Для STAR-машины разработаны как классические, так и динамические алгоритмы для решения задач на графах. Реализация этих алгоритмов на графических ускорителях дает возможность их практического использования, сохраняя преимущества ассоциативной обработки.

Для динамических алгоритмов решения задач теории графов нет неассоциативных параллельных алгоритмов, поскольку последовательные алгоритмы используют структуры данных, сложные для распараллеливания. Но использование данной технологии позволяет разрабатывать параллельные динамические алгоритмы.

Разработанные автором методы оптимизации ассоциативных алгоритмов для выполнения на графических ускорителях позволяют легко локализовать точки синхронизации в ассоциативных алгоритмах при реализации на GPU. Это существенно снижает вероятность появления ошибок синхронизации, которые из-за большой степени недетерминизма параллельного исполнения трудно обнаружить традиционными методами отладки. Использование предложенной технологии значительно уменьшает трудозатраты разработчиков при разрабатывании и отладки параллельных алгоритмов.

Основное содержание работы

Диссертация состоит из введения, 3-х глав, заключения, списка сокращений и терминов, а также библиографии. Общий объем диссертации 114 страниц, из них 105 страницы текста, включая 28 рисунков. Библиография включает 79 наименований на 8 страницах.

Во **Введении** обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследований, показана практическая значимость полученных результатов, представлены выносимые на защиту научные результаты.

В **первой главе** описана история развития ассоциативных архитектур от первого коммерчески успешного ассоциативного параллельного процессора STARAN до крупнейшего современного проекта с использованием ассоциативной архитектуры ATLAS Fast Tracker. Также приводятся описания моделей ассоциативных вычислений STAR-машины и MASC.

Вторая глава посвящена реализации модели STAR-машины на графических ускорителях с помощью технологии CUDA. Приведены ключевые отличия ассоциативных систем от PRAM-архитектур и обоснован выбор предложенной технологии для реализации STAR-машины.

Для реализации модели разработаны форматы внутреннего представления данных, реализованы операции ассоциативной обработки языка STAR и библиотека стандартных ассоциативных процедур. Эффективность реализации доказана как теоретической оценкой сложности, так и сравнением времени работы процедур с временем работы аналогов библиотек CUDA thrust и STL.

Выработаны методы оптимизации ассоциативных алгоритмов под исполнение на графических ускорителях, учитывающие особенности архитектур.

В **третьей главе** приводятся ассоциативные алгоритмы для решения задач графах: алгоритм Уоршалла транзитивного замыкания, динамический алгоритм Дейкстры нахождения кратчайших путей и динамический алгоритм Рамалингама для проблемы достижимости в потоковом графе с одним источником. Демонстрируется оптимизация этих алгоритмов и результаты вычислительных экспериментов.

Автореферат соответствует содержанию диссертации.

Степень обоснованности и достоверности научных результатов.

При получении основных результатов диссертационной работы использовались методы параллельного программирования, методы анализа

информационной структуры параллельных алгоритмов, элементы теории графов, а также формальные модели оценки эффективности параллельных программ и степени локальности данных. При разработке реализаций графовых алгоритмов и создании программного комплекса использовались методы объектно-ориентированного анализа и проектирования, а также программно-аппаратная архитектура параллельных вычислений CUDA, а также средства анализа эффективности и производительности – nvprof.

По теме диссертации в рецензируемых научных изданиях автором опубликовано 8 работ, получено свидетельство о регистрации программ для ЭВМ. Результаты исследования обсуждались на конференциях и семинарах профильных институтов.

Замечания по диссертационной работе.

По тесту работы следует сделать несколько замечаний. Следствием того, что автор глубоко погружен в тематику исследований является недостаточное внимание к пояснению рисунков. Так, например, рис. 2.1 на стр. 32 следовало бы пояснить. Почему приведенная модель графических ускорителей обусловила их выбор для реализации STAR-машины? В представленном виде этот рисунок представляется лишним. Аналогично, рис. 2.2 на стр. 35 ничего не добавляет к тексту работы и не служит информативной иллюстрацией.

Далее, фраза на стр. 59, что «реализации на GPU существенно превосходят по производительности последовательные реализации» является просто констатацией того, что параллельное исполнение соответствующих алгоритмов быстрее последовательного – краеугольного камня использования GPU.

Следовало бы дать пояснения к рис. 3.3 на стр. 73, который свидетельствует о достижении большего ускорения исполнения программы на последних архитектурных решениях от Nvidia.

Указанные замечания носят скорее технический характер и ни в коей мере не снижают высокий уровень проведенного исследования и не влияют на общую положительную оценку работы Т. В. Снытниковой.

Заключение по диссертации.


Диссертационная работа Т.В. Снытниковой представляет собой законченную научно-исследовательскую работу. В диссертационной работе получены новые результаты, имеющие теоретическую и прикладную ценность.

Тема, содержание и результаты диссертации Т. В. Снытниковой отвечают Паспорту специальности 2.3.5 - «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» (технические науки, ранее специальность 05.13.11). Автореферат полностью отражает содержание диссертации. Работа полностью соответствует требованиям п.п. 9, 10, 11, 13, 14 Положения о присуждении ученых степеней № 842 от 24.09.2013 г. с изменениями и дополнениями от 21.04.2016 г., а соискатель Т.В. Снытникова заслуживает присуждения ученой степени кандидата технических наук.

Отзыв на диссертационную работу обсужден и одобрен на заседании научного семинара «Инжиниринг современных информационных систем» Федерального государственного бюджетного учреждения науки Институт автоматизации и электрометрии Сибирского отделения Российской академии наук «9» декабря 2022 г., протокол № 2022/12/09.

Отзыв составлен профессором, д.ф.-м.н., заместителем директора по научной работе Федерального государственного бюджетного учреждения науки Институт автоматизации и электрометрии Сибирского отделения Российской академии наук Лаврентьевым Михаилом Михайловичем.

Подпись:

 Лаврентьев М. М.

Руководитель семинара:

д.т.н., заведующий лабораторией 19 ИАиЭ СО РАН  Зюбин В. Е.

Федеральное государственное бюджетное учреждение науки Институт автоматизации и электрометрии Сибирского отделения Российской академии наук

Адрес: 630090, Новосибирская область, г. Новосибирск, просп. Академика Коптюга, д.1

Веб-сайт: <https://www.iae.nsk.su>

Телефон: +7(383) 330-79-69

E-mail: iae@iae.nsk.su