

Отзыв официального оппонента на диссертационную работу
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»

Проблема параллелизма является триединой: здесь переплетаются фундаментальные проблемы и вычислительной техники, и алгоритмов, и программирования.

На любом уровне развития вычислительной машины остается понятие большой задачи, с которой за разумное время не может справиться машина типа RAM. А также возникают задачи, которые не могут быть эффективно решены на системах PRAM из-за большого числа операций сравнения или поиска. И тогда возникает необходимость в использовании ассоциативных параллельных систем.

В диссертационной работе Снытниковой Т. В. приводится реализация абстрактной модели ассоциативных вычислений на графических процессорах. А также рассматриваются ассоциативные алгоритмы и их использование с помощью представленной реализации. Таким образом, тема исследования представленной диссертации является **актуальной** для современной науки и ее приложений.

Диссертационная работа состоит из введения, трех глав, списка сокращений, терминов и библиографии, изложена на 114 страницах. Список цитируемой литературы содержит 79 наименований.

Во **Введении** обоснована актуальность выбранной темы исследований, сформулированы цель работы и задачи, которые необходимо решить для достижения цели. Приводятся основные результаты и положения, выносимые на защиту, дается оценка научной и практической значимости.

Первая глава посвящена обзору ассоциативных архитектур от первой до современной. Также приводится описание ассоциативных моделей ASC, MASC и STAR-машины.

Во **второй главе** представлена эффективная реализация STAR-машины на графических ускорителях с помощью технологии CUDA. Реализация состоит из трех частей: реализации типов данных и операторов языка Star, отвечающих за ассоциативную обработку; реализации библиотеки стандартных ассоциативных

алгоритмов; модуля ввода/вывода данных, поскольку ассоциативная обработка производится над бинарными таблицами. Эффективность реализации обосновывается теоретической сложностью и сравнением времени работы библиотеки стандартных алгоритмов с библиотекой CUDA thrust. Также в этой главе предложена методика оптимизации ассоциативных алгоритмов для выполнения на графических ускорителях, учитывающая особенности архитектур.

В **третьей главе** рассматривается модификация известной ассоциативной версии алгоритма Уоршалла транзитивного замыкания и приводятся новые ассоциативные версии динамического алгоритма Дейкстры поиска кратчайших путей и динамического алгоритма Рамалингама для проблемы достижимости в потоковом графе с одним источником. Приведенная адаптация ассоциативного алгоритма Уоршалла к выполнению на графических ускорителях на графе 5000 вершин приводит к ускорению более чем в 2000 раз по сравнению с последовательным вычислением и более чем в 7 раз по сравнению с известными оптимизациями алгоритма на GPU. При использовании более мощных графических ускорителей адаптация ассоциативного алгоритма Уоршалла также показывает большую производительность. Параллельные реализации динамических алгоритмов Дейкстры и Рамалингама являются уникальными.

В **заключении** приведены основные результаты диссертационной работы.

Автореферат в полном объеме отражает содержание диссертации. В нем изложены основные идеи и выводы диссертационной работы, отражен личный вклад соискателя в проведенное исследование, охарактеризованы степень новизны и практическая значимость полученных результатов. Результаты исследования соответствуют научной специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Основные результаты работы:

1. Построена реализация абстрактной модели ассоциативной обработки данных (STAR-машины) на современной параллельной архитектуре (графических ускорителях).
2. Указаны операции языка Star, критичные к синхронизации.
3. Разработана классификация библиотеки стандартных процедур языка Star по способу обработки данных.
4. Выработаны методы оптимизации ассоциативных алгоритмов для выполнения на графических ускорителях, учитывающие различия Star-машины и GPU.

5. Эффективность реализации обоснована в теории и на примере выполнения ассоциативных алгоритмов.

Полученные результаты являются **новыми**. В частности, предложенная реализация рассчитана на выполнение ассоциативных алгоритмов модели, а не прогнозирования ее свойств. Выделение операций, критичных к синхронизации и методы оптимизации ассоциативных алгоритмов для выполнения на графических ускорителях позволяют легко локализовать точки синхронизации, неправильное определение которых может привести к снижению производительности параллельных программ и к появлению ошибок, трудно обнаруживаемых традиционными методами отладки. Также предложена новая модификация алгоритма Уоршалла, дающая большее ускорение, чем известные реализации на GPU. Новыми являются и ассоциативные версии динамических алгоритмов Дейкстры поиска кратчайших путей и Рамалингама для проблемы достижимости в потоковом графе с одним источником. Реализация этих версий на графических ускорителях с помощью реализации ассоциативной модели вычислений не имеет аналогов.

Достоверность и обоснованность полученных результатов подтверждается доказанными утверждениями, проведенными моделированием и экспериментами на действующих вычислительных системах с графическими процессорами. Результаты работы обсуждались на семинарах «Математическое обеспечение высокопроизводительных вычислительных систем» ИВМиМГ СО РАН, «Дискретные экстремальные задачи» Института математики СО РАН, «Высокопроизводительные вычисления» ССКЦ ИВМиМГ СО РАН кафедры Вычислительных систем НГУ, на семинаре лаборатории программных систем машинной графики ИАиЭ СО РАН, на объединенном семинаре «Интеллектуальные системы и системное программирование» ИСИ СО РАН и кафедры программирования НГУ. Результаты диссертации опубликованы в 8 печатных работах, в том числе входящих в перечень ВАК РФ, индексируемых в базах данных Scopus и Web of Science, докладывались на ряде конференций. Диссертация не содержит заимствованных материалов или отдельных результатов без ссылок на авторов и источники заимствования.

Практическая ценность результатов диссертационной работы заключается в возможности применения ассоциативных версий классических и динамических алгоритмов для решения задач на графах. При этом сохраняются преимущества ассоциативной обработки на не ассоциативной параллельной архитектуре.

Также стоит отметить, что при распараллеливании алгоритмов возникают трудности с определением точек синхронизации. С одной стороны, неоптимальное расположение точек синхронизации потоков вычислений может сильно снизить производительность параллельной программы. С другой стороны, отсутствие необходимой точки синхронизации приводит к появлению ошибок, которые из-за большой степени недетерминизма параллельного исполнения трудно обнаружить традиционными методами отладки. Использование предложенной технологии значительно уменьшает трудозатраты разработчиков при написании и отладке параллельных алгоритмов.

Потенциальными потребителями результатов работы могут явиться организации, в которых востребованным является решение больших задач на графах и информационного поиска, в частности, вычислительная биология, анализ графов (социальные сети, навигация, антифрод). Полученные результаты могут найти применение и в учебном процессе, в частности, в курсах по архитектуре вычислительных систем и параллельным вычислительным технологиям.

Замечания.

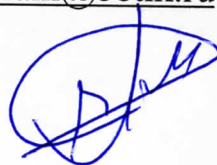
1. Предложенная реализация не проверялась на современной линейке графических ускорителей: Turing, Ampere, Hopper. Было бы интересно понять, какие результаты на них можно ожидать.
2. В диссертации приведены несколько ассоциативных алгоритмов на графах, но не объяснены критерии выбора именно этих алгоритмов.
3. В работе имеются стилистические недочеты и орфографические ошибки.

Приведенные замечания не влияют на общую положительную оценку работы и не снижают научную ценность полученных результатов.

Таким образом, диссертация Снытниковой Татьяны Валентиновны является законченной научно-квалификационной работой, в которой, на основании выполненных исследований и разработок, получено решение актуальной задачи — разработки моделей, алгоритмов и инструментальных средств ассоциативных вычислений на графических процессорах для решения задач на графах. Диссертационная работа удовлетворяет паспорту специальности 2.3.5 — «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей», а также всем критериям, установленным Положением о порядке присуждения ученых степеней ВАК для кандидатских диссертаций, а её автор Снытникова Татьяна

Валентиновна заслуживает присуждения ученой степени кандидата технических наук по искомой специальности.

Официальный оппонент: Абрамов Сергей Михайлович, член-корреспондент РАН, д.ф.-м.н. по специальности 05.13.17 «Теоретические основы информатики», директор Исследовательского центра мультипроцессорных систем Федерального государственного бюджетного учреждения науки Институт программных систем им. А.К. Айламазяна Российской академии наук, 152021, Ярославская область, Переславский район, с. Веськово, ул. Петра Первого, д.4 «а», тел. +7(903) 292-83-08, e-mail abram@botik.ru



Абрамов С. М.
19.12.2022

Подпись Абрамова С.М. заверяю:

Ученый секретарь

ИПС им. А.К. Айламазяна РАН, к.т.н.



С.М. Пономарева

19.12.2022