

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»

Актуальность темы исследования. Ассоциативные вычислительные системы и модели ассоциативных вычислений позволяют эффективно решать задачи поиска информации в больших массивах данных. Помимо развития специализированного аппаратного обеспечения ассоциативных вычислений развиваются и соответствующие модели и алгоритмы на их основе. Один из путей практической реализации алгоритмов на базе моделей ассоциативных вычислений — это реализация языков ассоциативных вычислений и сред их выполнения на базе архитектур общего назначения (x86-64, MIPS, ARM, RISC-V). В диссертации Снытниковой Т.В. предлагается подобный актуальный подход — эффективная реализация Star-машины ассоциативных вычислений на базе графических процессоров. Созданная система позволила реализовать на практике семейство эффективных параллельных ассоциативных алгоритмов решения задач на графах.

Содержание диссертации и автореферата. Диссертация состоит из введения, трёх глав, заключения, списка сокращений, терминов и библиографии. Общий объем диссертации 114 страниц, 28 рисунков, библиографический список включает 79 источников. Автореферат в полном объеме отражает содержание диссертации. В нем изложены основные идеи и выводы диссертации, отражен личный вклад автора 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».

В **первой главе** дан обзор развития проблемно-ориентированных вычислительных систем с ассоциативной памятью. Приведено описание ассоциативных моделей вычислений ASC, MASC, описан язык Star и его базовые операции.

Вторая глава посвящена вопросам эффективной реализации операций языка Star на CUDA для графических процессоров. Выделены основные свойства ассоциативных моделей вычислений и подходы к их сохранению при реализации операций языка Star на CUDA. Для работы со столбцом матричной памяти (slice) предложено использовать битовый массив (вектор) с элементарными блоками по 64 бита, что обусловлено ограничениями архитектуры CUDA. Все операции с матричной памятью (таблицей) разбиты на 4 группы, предложены способы их оптимизации с учетом архитектуры GPU.

В **третьей главе** рассмотрены известные и предложены модифицированные ассоциативные параллельные алгоритмы для решения задач на графах. Предложенная адаптация Star-алгоритма Уоршалла к архитектуре CUDA приводит к большему ускорению, чем известные оптимизации алгоритма на GPU. Реализованы статическая и динамическая версии ассоциативного алгоритма Дейкстры нахождения кратчайших путей от одной из вершин ориентированного взвешенного графа до всех остальных вершин в случае положительных весов. Приведена ассоциативная версия алгоритма Рамалингама для решения динамической проблемы достижимости в потоковых графах с одним источником при добавлении новой дуги.

Новизна и значимость результатов определяется эффективностью разработанных ассоциативных параллельных алгоритмов решения задач на графах по сравнению с известными методами их реализации на графических процессорах.

1. Разработана среда выполнения ассоциативных параллельных Star-программ на графических процессорах. Эффективная реализация среды на CUDA позволила реализовать эффективные ассоциативные параллельные алгоритмы и экспериментально исследовать их производительность.

2. Разработан адаптированный ассоциативный алгоритм Уоршалла транзитивного замыкания, который характеризуется большим ускорением чем известные реализации алгоритма для графических процессоров.

3. Созданы статическая и динамическая версии ассоциативного алгоритма Дейкстры.

4. Предложена ассоциативная параллельная версия динамического алгоритма Рамалингама для решения задачи достижимости в потоковых графах с одним источником при добавлении новой дуги.

Достоверность и обоснованность научных положений, выводов и рекомендаций. Достоверность проведенных исследований 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» ИВМиМГ СО РАН, 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. Основные результаты диссертации докладывались и обсуждались на Международной конференции «Мърчуковские научные чтения» (2017 г., 2020 г.).

Основные результаты диссертации опубликованы в научных изданиях: 4 работы в журналах из перечня ВАК РФ; 2 – в материалах конференций; получено свидетельств о государственной регистрации программ для ЭВМ. Диссертация не содержит

заимствованных материалов или отдельных результатов без ссылок на авторов и источники заимствования.

Рекомендации по использованию результатов. Разработанные в диссертации алгоритмы и программный инструментарий могут найти применение в организациях, 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 вычислительных технологиям.

Замечания.

1. При реализации Star-машины на GPU один из ключевых аспектов является производительность операций над битовыми векторами, так соискатель использует функцию CUDA `_ffsll` в операции FND, но в том же время использует итеративный алгоритм подсчета числа установленных разрядов (операция NUMB), хотя в CUDA имеется функция `_popc1l` и др.

2. В тексте диссертации стоило уделить больше внимания сути ассоциативных алгоритмов, их ключевые отличия от алгоритмов для RAM-машин.

3. При описании результатов экспериментов не отражены детали методики измерений: версии программного обеспечения, количество запусков и оценка погрешности измерений.

4. По тексту диссертации имеются стилистические замечания, некоторые операции описаны без требуемой полноты:

- название языка Star (название из первоисточника) в разных частях диссертации используется в разных стилях: «STAR», «Star»;
- на С. 26 для операции FND(Y) указано ограничение « $i \geq 0$ », хотя в операции Y(i) нумерация строк начинается с 1;
- на С. 27 в описании операций MIN, MAX, MATCH не указано назначение аргумента Z;
- в описании операций TRIM, REP (С. 27) используется операция взятия модуля от слова («|w|»), не понятен смысл этого ограничения, все слова/строки имеют фиксированную длину k бит/столбцов, следовательно ограничение имеет вид $i < j \leq k$.

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

Заключение. Диссертация Снытниковой Татьяны Валентиновны «Эффективная реализация модели ассоциативных вычислений на графических ускорителях для решения задач на графах» является самостоятельной, 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. Диссертация отвечает требованиям Положения ВАК РФ о порядке присуждения ученых степеней, а её автор, Снытникова Татьяна Валентиновна, заслуживает присуждения ей ученой степени кандидата технических наук по специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Официальный оппонент:

Курносков Михаил Георгиевич, доктор технических наук по специальности 2.3.2 «Вычислительные системы и их элементы», профессор Кафедры вычислительных систем федерального государственного бюджетного образовательного учреждения высшего образования «Сибирский государственный университет телекоммуникаций и информатики» (СибГУТИ), директор Центра параллельных вычислительных технологий СибГУТИ, 630102, г. Новосибирск, ул. Кирова, 86, тел. +7 (383) 269-82-75, e-mail: mkurnosov@sibsutis.ru.

 / Михаил Георгиевич Курносков /

«14» декабря 2022 г.

Подпись официального оппонента заверяю

Зав. кафедрой





14.12.2022