

## ОТЗЫВ

официального оппонента на диссертационную работу Сударевой Ольги Юрьевны «Встречная оптимизация класса задач трёхмерного моделирования для архитектур многоядерных процессоров», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Рассматриваемая диссертационная работа посвящена встречной оптимизации архитектур вычислительных систем и алгоритмов, используемых в высокопроизводительных вычислениях. Встречная оптимизация подразумевает, что ни реализация алгоритма, ни архитектура не являются фиксированными, а изменяются навстречу друг другу с целью максимально эффективного решения вычислительной задачи.

Анализ репрезентативного набора вычислительных задач, для которых предназначается разрабатываемая архитектура, на основе построения концептуальной схемы вычислений по рассматриваемым алгоритмам, позволяет определить характеристики системы, которые оказывают наибольшее влияние на производительность. Такое исследование может быть проведено не только для существующих вычислительных систем, но и для перспективных систем, которые находятся на стадии проектирования.

В диссертации рассматривается набор алгоритмов, которые имеют отношение к вычислительно-ёмким задачам трёхмерного моделирования. Исследование проводится для систем гибридной архитектуры, в которых сочетаются центральные процессоры и массивно-параллельные ускорители. Целью исследования является выработка методов, которые позволяют оценивать потенциал гибридных систем для выбранных вычислительных алгоритмов.

Мотивацией данной работы было исследование потенциала специализированных процессоров КОМДИВ, имеющих гибридную архитектуру, которые были созданы в ФГУ ФНЦ НИИСИ РАН для приложений обработки сигналов. В настоящее время такие процессоры успешно применяются для данного класса приложений и показывают высокую эффективность.

Исследование производительности процессоров КОМДИВ на широком спектре задач, определение потенциала развития этой архитектуры для применения в высокопроизводительных вычислениях общего назначения представляется весьма актуальным. Особенно учитывая, что аппаратную базу для суперкомпьютеров сейчас составляют исключительно импортные процессоры и ускорители, что в настоящее время несет определенные риски.

Диссертация состоит из введения, четырёх глав, заключения и двух приложений. Работа изложена на 175 страницах. Иллюстративный материал состоит из 19 рисунков и 24 таблиц. Автореферат диссертационной работы представлен на 25 страницах.

Во введении обосновывается актуальность выбранной темы диссертационной работы, формулируются поставленные цели и задачи исследования, описываются методы исследования и основные результаты. Обосновывается научная новизна результатов, их теоретическая и практическая значимость.

В главе 1 приводится обзор современных тенденций в развитии архитектур высокопроизводительных вычислительных систем и обосновывается актуальность разработки эффективных реализаций базовых вычислительных алгоритмов на гибридных системах. Описываются существующие отечественные разработки, главным образом, гибридные микропроцессоры НИИСИ РАН. Приводится обзор методов разработки параллельных реализаций для гибридных систем. Выявляются общие черты в архитектурах и подходах к разработке программных реализаций для систем на базе GPU и процессоров КОМДИВ. Строится модель гибридной системы. Обосновывается выбор трёх вычислительных процедур из набора тестов NAS Parallel Benchmarks. Формулируется метод оценки производительности процедур на гибридной системе.

Глава 2 посвящена вычислениям на GPU. Выбранные алгоритмы исследуются на гибридных системах на базе GPU. Для каждого алгоритма построена схема вычислений, на основе которой оценивается ожидаемая производительность. Описываются разработанные автором реализации, использующие открытый стандарт OpenCL. Приводятся результаты тестирования на различных системах, включая суперкомпьютер K100 ИПМ им. М.В. Келдыша

РАН. Полученные результаты анализируются, сравниваются с предварительно полученными оценками и с данными о достигаемой на аналогичных алгоритмах производительности из открытых публикаций. Также сравнивается полученная производительность на GPU и CPU. Полученные результаты показывают, что в большинстве случаев достигаемая на GPU производительность определяется не пиковой производительностью ускорителя, а пропускной способностью памяти.

В главе 3 аналогичное исследование проводится для процессоров VM7 и VM9 линейки КОМДИВ. Представлены оценки для рассматриваемых алгоритмов на основе производительности ядра сопроцессора CP2 и пропускной способности канала DMA. Также получены оценки максимальной достижимой на рассматриваемых алгоритмах производительности ядер, которые учитывают специфику архитектуры и набора инструкций CP2. Приводятся результаты тестирования, которые демонстрируют применимость метода оценки производительности. При сравнении полученных результатов с оценками указывается, какие именно накладные расходы в алгоритмах и какие именно особенности архитектуры процессоров не позволяют добиться теоретически достижимого максимума производительности. Для демонстрации потенциала процессоров НИИСИ РАН производительность созданных реализаций рассматриваемых алгоритмов сравнивается с производительностью, полученной на других процессорах, в том числе CPU Intel.

В главе 4 подводится итог исследования, проведённого для гибридных процессоров НИИСИ РАН. Описываются преимущества и недостатки архитектурных решений, реализованных в этих процессорах. На основании сформулированного в 1-й главе критерия сбалансированности для рассматриваемых алгоритмов выводятся оптимальные соотношения между пропускной способностью DMA, тактовой частотой и количеством вычислительных секций CP2. Анализируется, какие именно особенности программной модели CP2 и DMA являются наиболее существенным ограничением для производительности. Предлагается проект развития архитектуры, направленный на существенное повышение производительности.

В заклучении приводятся основные результаты диссертации, описаны различные пути их практического применения, намечены направления дальнейших исследований.

Таким образом, диссертация содержит обоснование актуальности выбранной темы и формулировку задач, предлагает решение поставленных задач. Строится метод оценки производительности процедур, показана его применимость для импортных и отечественных вычислительных систем на примере нескольких алгоритмов. Применимость метода подтверждена данными о производительности разработанных автором процедур. По итогам работы сделаны выводы о перспективах использования гибридных многоядерных процессоров НИИСИ РАН в высокопроизводительных вычислениях общего назначения и предложен проект дальнейшего развития этой процессорной архитектуры.

В диссертационной работе О. Ю. Сударевой получены следующие основные результаты, обеспечивающие научную новизну работы:

- 1) Построена модель обобщённой гибридной вычислительной системы.
- 2) Разработан метод оценки производительности на обобщённой гибридной системе на основе представления исследуемого алгоритма в виде набора операций и зависимостей по данным между ними. Набор операций состоит из операций передачи данных между ЦПУ и сопроцессором и вычислительных операций на сопроцессорах. При этом учитывается возможность одновременного выполнения вычислений и обмена данными.
- 3) Репрезентативный набор алгоритмов, применяемых в трёхмерном моделировании, исследован при помощи построенного метода. Созданы переносимые OpenCL реализации этих алгоритмов для гибридных систем с GPU. Впервые созданы реализации для гибридных процессоров VM7 и VM9 НИИСИ РАН оригинальной отечественной архитектуры, обладающих SIMD-сoproцессором CP2.
- 4) Полученные оценки дали строгое обоснование неулучшаемости разработанных автором библиотечных процедур БПФ и свёртки.
- 5) Разработан проект развития архитектуры гибридных процессоров НИИСИ РАН, направленный на повышение производительности вычислений в несколько раз на задачах рассмотренных классов. Предложены улучшения программной модели сопроцессора CP2, контроллера DMA, а также ряда количественных характеристик.

Результаты диссертационной работы были представлены на научных конференциях и семинарах, в том числе международных, и опубликованы в 10 печатных работах, из них две публикации в журналах из перечня ВАК и один РИД в виде свидетельства о регистрации программы для ЭВМ.

В автореферате диссертации представлена необходимая информация об основных положениях и результатах работы. Автореферат в полной мере отражает основное содержание диссертации.

Научные положения, сформулированные в диссертации, выглядят вполне обоснованными. Достоверность научных результатов хорошо подтверждена в тестах на различных существующих гибридных архитектурах.

Результаты диссертационного исследования обладают несомненной теоретической и практической значимостью. Предложенный метод оценки производительности позволяет получать априорные оценки производительности алгоритмов на существующих и разрабатываемых гибридных системах, оценивать влияние на производительность рассматриваемого алгоритма изменений основных количественных характеристик системы. Наиболее значимым практическим результатом работы является подготовленный проект усовершенствования архитектуры гибридных процессоров НИИСИ РАН.

В то же время к данной диссертационной работе имеются следующие замечания.

- 1) В 1-й главе (с. 28) вводится параметр  $bw_{CP}$ , скорость доступа к памяти на сопроцессоре, актуальный в расчетах на GPU, который далее при анализе алгоритмов в главе 2 практически не используется. Разве пропускная способность памяти ускорителя не существенна для данных процедур?
- 2) В главе 2 в таблицах 2.2, 2.4 теоретическая оценка производительности на четырех устройствах получена умножением на 4 оценки для одного устройства. Такая оценка, естественно, отличается в разы от наблюдаемых на практике значений. Это странно. А как же метод, как же учет скорости передачи данных?

- 3) В разделе 2.3.5 (стр. 77) рассматриваются матрицы 14 типов с какими-то названиями, многие из которых не относятся к трехмерным задачам. Из результатов на рис. 2.9 видно, что производительность сильно зависит от структуры матрицы. Затем показываются результаты для матриц 5 классов, отмеченных буквами, характеризующими размер. Непонятно, какая именно структура у этих матриц. Судя по результатам из таблицы 2.4, где параллельная эффективность на 4-х устройствах невысока, эта структура излишне неудобная для обменов между устройствами.
- 4) В главе 3 в результатах тестирования (3.1.4, 3.2.4, и т.д.) колонка с теоретической оценкой называется Теор. Максимум, в главе 2 просто Теория. В чем разница? Не понятен смысл этих данных. Это априорная оценка, или теоретически достижимый максимум? А где тогда априорная оценка? Если теор. максимум и априорная оценка – это одно и то же, то почему такая большая разница с практикой? Таблица 3.6, 3.7 – наблюдается разница порядка 25 раз. В чем смысл таких оценок?
- 5) В главе 4 предполагается, что предложенные изменения архитектуры выполнимы с технической точки зрения и не потребуют радикального увеличения площади кристалла, энергопотребления и т.п. Однако никаких обоснований этому не приводится.
- 6) В разделе 4.3 (стр. 133) говорится, что для SP2 отсутствуют нормальные общепринятые средства разработки, используется язык ассемблера. При этом нигде в выводах не говорится о первостепенной важности создания таких средств разработки – поддержка OpenCL, компилятора C++ (возможно, с какими-то директивными расширениями). Говорится, что неплохо было бы, например, сделать MPI над Rapid IO. А как же самое главное? В чем причина, что вопрос средств разработки даже не рассматривается?

Следует отметить, что данные замечания носят частный характер и не снижают общей оценки диссертации, полученных в ней результатов, а также качества проделанной работы.

Считаю, что данная диссертация является законченной научно-квалификационной работой, выполненной автором на высоком научном уровне. Диссертация удовлетворяет всем требованиям ВАК, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей», а ее автор, Сударева Ольга Юрьевна, заслуживает присуждения учёной степени кандидата физико-математических наук по указанной специальности.

Официальный оппонент,

доктор физико-математических наук, ведущий научный сотрудник федерального государственного учреждения «Федеральный исследовательский центр Институт прикладной математики им. М. В. Келдыша Российской академии наук»

\_\_\_\_\_

Горобец Андрей Владимирович

23 апреля 2018 г.

125047, Москва, Миусская пл., д.4.

Телефон: 8 499 2207218

E-mail: andrey.gorobets@gmail.com

Подпись А. В. Горобца

Ученый секретарь ИГ.

к. ф.-м. н.

А. И. Маслов