

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Сударевой Ольги Юрьевны

«Встречная оптимизация класса задач трёхмерного моделирования для архитектур многоядерных процессоров»,

представленную к защите 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 оценке влияния наиболее важных характеристик архитектуры гибридных вычислительных систем на производительность таких систем при высокопроизводительных вычислениях, а именно, при решении задач трёхмерного моделирования.

В настоящее время для высокопроизводительных вычислений повсеместно используются гибридные системы с массивно-параллельными сопроцессорами. Однако не для каждой задачи использование таких систем обеспечивает существенный прирост производительности вычислений, по сравнению с классическими системами на базе универсальных процессоров.

Проведение таких оценок целесообразно не только для существующих гибридных систем, но и для систем, находящихся на стадии разработки: это может позволить оптимизировать архитектуру под целевой класс вычислительных процедур и таким образом добиться высокой производительности. Тем более, что гибридные системы, как правило, не являются универсальными.

Таким образом, исследования, направленные на разработку методов оценки производительности процедур на гибридных системах, являются актуальными.

В работе рассматривается подмножество существующих типов гибридных систем, а именно, гибридные узлы на базе универсальных процессоров и графических ускорителей (GPU). Ограничение класса рассматриваемых систем, естественно, упрощает задачу исследования производительности. Оно оправдано тем обстоятельством, что системы на базе GPU используются для масштабных расчётов во многих исследовательских и промышленных организациях, и возможность оценивать преимущества и недостатки таких систем при решении производственных задач имеет большое значение.

Следует отметить, что, как показано в рецензируемой работе, вышеуказанная архитектура во многом схожа с архитектурой отечественных

гибридных микропроцессоров, разрабатываемых в ФГУ ФНЦ НИИСИ РАН, в котором и была выполнена рецензируемая работа. Гибридные микропроцессоры НИИСИ РАН имеют оригинальную отечественную архитектуру, уже сейчас показывают высокую эффективность при решении задач цифровой обработки сигналов и потенциально могут стать альтернативой зарубежным гибридным системам для решения задач трёхмерного моделирования. Возможность влиять на архитектуру вычислительной системы при разработке программного обеспечения позволяет поставить задачу встречной оптимизации архитектуры, т.е. усовершенствования тех подсистем, которые ограничивают производительность вычислительных процедур, применяемых в трёхмерном моделировании. Эта задача тоже несомненно актуальна хотя бы потому, что очень востребована разработчиками систем с перестраиваемой архитектурой.

Следует отметить и то обстоятельство, что задачи оптимизации параллельных программ невозможно решать без моделирования как самих программ, так и вычислительных систем, на которых предполагается их выполнять. Таким образом, можно констатировать, что проблемы, рассматриваемые в диссертационной работе весьма актуальны.

В диссертационной работе О.Ю. Сударевой получены следующие **основные результаты**:

- 1) Построена модель гибридной вычислительной системы с управляющим процессором и сопроцессорами массивно-параллельной архитектуры. Данная модель представляет собой высокоуровневое описание архитектуры, что позволяет рассмотреть в этой модели системы, различающиеся масштабом и деталями архитектуры управляющих процессоров, сопроцессоров, контроллеров памяти и других структурных элементов. В частности, в модели могут быть рассмотрены как гибридные узлы с графическими ускорителями (GPU), так и гибридные микропроцессоры НИИСИ РАН.
- 2) Разработан метод оценки производительности алгоритмов на гибридных вычислительных системах, соответствующих построенной модели. В данном методе для рассматриваемого алгоритма строится последовательность основных операций, наиболее затратных с точки зрения времени работы: вычислений на сопроцессорах и обменов данными между сопроцессорами и системной памятью. Далее на основе этой последовательности выводятся формулы для оценки времени работы процедуры и её производительности, которые зависят от нескольких числовых параметров, характеризующих вычислительную систему и процедуру. Сформулирован критерий сбалансированности вычислительной системы на заданной процедуре: каждая вычислительная подпроцедура (ядро), выполняемая на сопроцессорах, должна занимать столько же

времени, сколько передача её входных и выходных данных на сопроцессоры и обратно. Выведенные формулы позволяют представить этот критерий в виде формального соотношения между параметрами вычислительной системы и процедуры.

- 3) В диссертации продемонстрирована применимость предложенного метода оценки производительности. Для этого выбрано несколько представительных вычислительных процедур и разработаны их оптимизированные реализации, для которых проведены замеры производительности на реальных аппаратных стендах. Во-первых, рассмотрены системы, которые широко применяются для вычислений общего назначения в настоящее время — гибридные узлы и кластеры с GPU, — и распространённый инструмент программирования под такие системы — стандарт OpenCL. Во-вторых, рассмотрены гибридные процессоры НИИСИ РАН оригинальной отечественной архитектуры — VM7 и VM9, обладающие SIMD-сопроцессором CP2; для программирования под эти процессоры применялись специализированные инструментальные средства, которые позволяют осуществить низкоуровневую оптимизацию процедур под архитектуру процессоров.
- 4) Для процессоров НИИСИ РАН автором разработаны оптимизированные процедуры Быстрого Преобразования Фурье длинных векторов и свёртки, которые вошли в промышленную библиотеку цифровой обработки сигналов НИИСИ РАН. При помощи разработанного метода оценки производительности обоснована неулучшаемость указанных процедур на нынешней версии архитектуры процессоров НИИСИ РАН.
- 5) По итогам исследования автором подготовлен проект оптимизации архитектуры гибридных процессоров НИИСИ РАН, который включает улучшение количественных характеристик процессора, усовершенствование программной модели сопроцессора CP2 и контроллера прямого доступа к памяти DMA. Оптимизации направлены на то, чтобы снять архитектурные ограничения, которые не позволяют достичь теоретически возможной максимальной производительности на алгоритмах, рассмотренных в диссертации.

Достоверность перечисленных результатов подтверждается результатами многочисленных вычислительных экспериментов на GPU и на процессорах КОМДИВ, результаты которых приведены в работе. Кроме того, разработанные автором процедуры БПФ и свёртки имеют успешные применения в производственных приложениях.

Все основные результаты диссертационной работы опубликованы в 10 публикациях в открытой печати, из них: 2 статьи в изданиях, включенных в список

ВАК, 1 свидетельство на программу для электронных вычислительных машин, 1 монография, 4 статьи в прочих научных изданиях, 2 публикации в тезисах докладов. Публикации в полной мере отражают основное содержание диссертационной работы. Работа докладывалась на научных конференциях и семинарах, в том числе международных.

Диссертация состоит из введения, четырёх глав, заключения и двух приложений. Её полный объём составляет 175 страниц, иллюстративный материал представлен в виде 19 рисунков и 24 таблиц. Список литературы содержит 158 наименований.

Во введении раскрывается актуальность выбранной темы диссертации, описываются поставленные цели, задачи и методы исследования. Перечисляются основные положения, выносимые на защиту, обосновывается их научная новизна, а также теоретическая и практическая значимость.

Глава 1 посвящена построению модели гибридной вычислительной системы и метода оценки ожидаемой производительности вычислительной процедуры на такой системе. Для этого автор проводит обзор существующих архитектур вычислительных систем, которые используются для высокопроизводительных вычислений, и методов программирования под такие системы. Обосновывается популярность и преимущества гибридных систем с графическими ускорителями (GPU). Далее рассматриваются существующие отечественные разработки, при этом основное внимание уделяется гибридным многоядерным процессорам КОМДИВ разработки ФГУ ФНЦ НИИСИ РАН. Строится модель гибридной вычислительной системы, которая отражает общие черты двух рассмотренных классов гибридных систем, и метод априорной оценки производительности, основные этапы которого находятся в соответствии с этапами разработки оптимизированных процедур под гибридные системы. Также в первой главе выбирается тестовый набор вычислительных процедур, которые рассматриваются в дальнейших главах, а именно, три процедуры из набора тестов NAS Parallel Benchmarks: быстрое преобразование Фурье, многосеточный метод и метод сопряжённых градиентов, который сводится к умножению разреженной матрицы на вектор. В конце главы формулируется критерий сбалансированности вычислительной системы на заданной вычислительной процедуре.

В главе 2 подробно рассматриваются гибридные системы на базе GPU. К этим системам и к выбранным вычислительным процедурам применяется построенный метод и выводятся оценки производительности. Делается вывод о том, что производительность многомерного БПФ больших массивов ограничена пропускной способностью шины PCI Express. Кроме того, подробно описываются разработанные автором оптимизированные реализации процедур многосеточного метода и метода сопряжённых градиентов, который сводится к умножению разреженной матрицы на вектор. Приводятся результаты вычислительных

экспериментов, проведённых с этими реализациями. Результаты подтверждают применимость построенного метода оценки производительности для гибридных систем на базе GPU.

В главе 3 проводятся оценки производительности аналогичных процедур на отечественных микропроцессорах КОМДИВ ВМ7 и ВМ9, снабженных SIMD-сопроцессором CP2, и описываются разработанные автором оптимизированные реализации процедур для этих процессоров. При этом отмечаются отличия выбранных схем вычислений от тех, что использовались во второй главе, а также особенности вычислительных ядер для CP2. Делаются выводы о том, для каких процедур производительность определяется производительностью ядра на CP2, а для каких — скоростью пересылки данных по DMA. Приводятся результаты вычислительных экспериментов на серийных процессорах, которые подтверждают применимость построенного метода для процессоров КОМДИВ. Эти результаты, кроме того, сравниваются с производительностью аналогичных процедур на других процессорах.

В главе 4 на основе выведенных в предыдущей главе формул для оценки производительности процедур на КОМДИВ формулируется критерий сбалансированности этих процессоров, который наглядно демонстрирует требования, предъявляемые каждой процедурой к пропускной способности канала доступа к памяти. Далее систематизируются архитектурные ограничения процессоров КОМДИВ, с которыми автор столкнулся при разработке оптимизированных процедур и которые ограничивают производительность. По итогам предлагается проект оптимизации программной модели отдельных подсистем процессоров, а именно, сопроцессора CP2 и контроллера DMA. Также автор оценивает потенциал гибридных процессоров НИИСИ РАН для высокопроизводительных вычислений, в сравнении с другими процессорами.

В заключении формулируются основные результаты диссертации и направления дальнейших исследований. Отмечается, что первым этапом реализации предложенного проекта развития архитектуры процессоров НИИСИ РАН должна стать реализация в программных эмуляторах.

Метод оценки производительности, построенный автором, имеет **научную и практическую значимость**. Так, он может использоваться для принятия решения о том, целесообразно ли для текущих производственных задач осуществлять переход с классических систем на гибридные, который требует переработки большого количества программного кода. К достоинствам разработанного автором метода можно отнести его простоту: он оперирует малым количеством параметров и позволяет абстрагироваться от технических деталей реализации процедуры и архитектуры вычислительной системы — в совокупности это позволяет оценить по порядку производительность процедуры путем проводимого «вручную»

расчета, что требует значительно меньше затрат времени и труда, чем реализация процедуры и последующее тестирование.

Кроме того, при разработке оптимизированной процедуры на различных этапах требуется оценить, может ли процедура быть ускорена, и что в первую очередь требует доработки: вычисления на сопроцессорах или же пересылки данных. Здесь также приходит на помощь метод оценки производительности, при этом в формулы вместо пиковых значений параметров подставляются реальные значения, полученные в результате замеров.

Такие оценки, проведённые автором в ходе реализации алгоритма многосеточного метода на GPU, позволили добиться более высокой производительности, чем у аналогичных процедур, описанных в открытых источниках.

Критерий сбалансированности полезно учитывать при проектировании новой гибридной системы: главным образом, он позволяет определить, какая требуется пропускная способность канала обменов данными, чтобы свести к минимуму простои сопроцессоров.

С использованием метода впервые исследована производительность ряда процедур, применяемых в трёхмерном моделировании, на отечественных гибридных процессорах КОМДИВ. При этом разработаны высокооптимизированные процедуры, и результаты их тестирования демонстрируют конкурентоспособность процессоров. В частности, разработанные процедуры обработки сигналов (БПФ и свёртка) применяются при решении реальных задач в гидроакустических комплексах. В ходе работы над реализацией процедур выявлены архитектурные ограничения, присущие процессорам КОМДИВ в настоящее время, и разработан проект развития архитектуры процессоров.

К числу **недостатков** работы можно отнести следующие:

- 1) Недостаточно подробно описаны сходства и различия алгоритмов работы DMA архитектур GPU и КОМДИВ.
- 2) Не приведены явно количественные или качественные ограничения на степень разреженности матрицы, при выполнении которых справедливы прогнозы повышения производительности процедуры SpMV в результате доработки архитектуры отечественных микропроцессоров КОМДИВ.

Несмотря на указанные недостатки, диссертация обладает высоким качеством, научной и практической ценностью.

Автореферат диссертации в полной мере раскрывает содержание представленной работы. Тексты диссертации и автореферата написаны в научном стиле и надлежащим образом оформлены.

Таким образом, диссертация Сударевой Ольги Юрьевны выполнена на актуальную тему, носит законченный характер, содержит новые научные результаты, обладающие практической полезностью, т.е. удовлетворяет требованиям ВАК, предъявляемым к кандидатским диссертациям. Диссертант соответствует требованиям, предъявляемым к научным работникам, и 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,
заведующий отделом компиляторных технологий ИСП РАН,
кандидат физико-математических наук, доцент

С.С. Гайсарян

25» апреля 2018 г.

Подпись Гайсаряна С.С. заверяю,
Учёный секретарь ИСП РАН,
кандидат физико-математических наук

Н.В. Пакулин

апреля 2018 г.

Сведения об оппоненте: Гайсарян Сергей Суменович

1109004, Москва, ул. А. Солженицына, 25. Телефон: +7(495)912-5317 доб. 430

e-mail: ssg@ispras.ru

Федеральное государственное бюджетное учреждение науки «Институт системного программирования им. В.П. Иванникова Российской академии наук»,
заведующий отделом компиляторных технологий