

Федеральное государственное
бюджетное учреждение науки
**ИНСТИТУТ СИСТЕМ
ИНФОРМАТИКИ
ИМ. А. П. ЕРШОВА**
Сибирского отделения
Российской академии наук
(ИСИ СО РАН)

630090, Новосибирск, пр. Акад. Лаврентьева, 6
тел. (383) 330-86-52, факс (383) 332-34-94
e-mail: iis@iis.nsk.su, http://www.iis.nsk.su/
ОКПО 03945008, ОГРН 1025403664340
ИНН/КПП 5408100201/540801001

От _____ № _____

На № _____ от _____

УТВЕРЖДАЮ

Директор Института систем информатики имени

А. П. Ершова Сибирского отделения РАН



д.ф.м.н. Цалыянов А. Ю.

15 января 2025

Отзыв ведущей организации

Федерального государственного бюджетного учреждения науки

Института систем информатики имени А. П. Ершова Сибирского отделения РАН

на диссертацию Метелицы Елена Анатольевны

"Автоматизация распараллеливания программ со сложными информационными зависимостями",

представленную на соискание ученой степени кандидата физико-математических наук

по специальности 2.3.5 — «Математическое и программное обеспечение вычислительных систем,

комплексов и компьютерных сетей»

Актуальность темы диссертации

Существует много задач, требующих для своего решения больших объемов вычислений. Например, прогноз изменений климата, машинное обучение, исследования космоса, получение новых лекарств.

Для решения таких задач строятся математические модели, которые затем решаются численными методами. Большинство таких задач приводят к математическим моделям, использующим уравнения в частных производных. Для повышения точности берутся большие

размерности задачи (увеличение размеров сетки), это ведёт к повышению размерности соответствующей системы линейных алгебраических уравнений. При решении системы уравнений большой размерности прямыми методами сильно накапливается погрешность, поэтому используются итерационные методы. Возникает задача быстрого выполнения итерационных численных методов. Ускорению таких методов и посвящается диссертация.

Не всегда мощности вычислительных систем используются в полной мере. Одними из главных инструментов, которые улучшают использование ресурсов компьютера, являются оптимизирующие/распараллеливающие компиляторы. Они основаны на теории преобразования программ и могут анализировать исходные тексты программ и автоматически выполнять оптимизирующие преобразования, что позволяет значительно сократить время на разработку и избежать необходимости для прикладного программиста в специализированных знаниях в областях параллельной разработки программ и архитектуры компьютера.

В последнее время очень быстро развивались архитектуры компьютеров, а возможности компиляторов в оптимизации программ не развивались адекватными темпами. В результате, многие прикладные программы используют менее одного процента пиковой производительности процессора. Назрел момент создания оптимизирующих компиляторов нового поколения. Рассматриваемая диссертация представляет собой шаг на пути решения этой проблемы.

Степень обоснованности положений, выводов и рекомендаций

Присутствуют ссылки на другие источники (почти 150 источников), строгий математический язык для формулировок, доказательства основных утверждений, на которых основываются выводы и рекомендации.

Проведено множество численных экспериментов, результаты которых отражены в 20 таблицах. Приводится описание архитектуры компьютера, на котором проводились эксперимент, а также компиляторы.

Оценка новизны и достоверности основных научных положений, выводов и рекомендаций

Достоверность полученных результатов подтверждается корректным использованием математических формулировок и результатов теории преобразования программ, линейной алгебры, теории графов; тестированием корректности и производительности разработанного ПО; успешным применением разработанного ПО в научной деятельности; полученными и опубликованными результатами численных экспериментов по ускорению целевых алгоритмов. В третьей главе диссертации представлено 20 таблиц с результатами численных экспериментов.

Для контроля корректности (достоверности) преобразований программ используется анализ информационных зависимостей. Для этого анализа по тексту программы строится граф

информационных зависимостей, а также, решетчатый граф и его фактор-граф. На основе анализа информационных зависимостей подтверждается, что сходимость преобразованных итерационных программ не меняется по сравнению с исходными.

Приводятся примеры, в которых обозначены границы применимости автоматического распараллеливания: компилятор, не знающий диапазоны значений переменных и требуемую погрешность вычислений, не может автоматически оптимизировать код, но вручную разработчик программы мог бы код распараллелить. Приводимые примеры показывают, что компилятор не может жертвовать точностью вычислений ради ускорения. В 4-й главе диссертации предлагается диалоговый режим компиляции, который в таких случаях может подсказать разработчику программы в каком месте и как этот код можно оптимизировать.

Новизна. Разработана цепочка оптимизирующих преобразований гнёзд циклов итерационного типа, основанная на универсальном высокоуровневом древовидном внутреннем представлении. Известный подход (распараллеливающая система PLUTO), основанный на полиэдральном внутреннем представлении, нацелен только на преобразование гнёзд циклов. Подход, разработанный в диссертации, позволяет расширить цепочку преобразований для оптимизации гнёзд циклов итерационного типа дополнительными преобразованиями, тем самым увеличить класс оптимизируемых алгоритмов.

Разработан и теоретически обоснован новый метод обхода точек тайла, который повышает временную локальность данных (повышается количество чтений из регистров вместо более медленных модулей памяти) и даёт ускорение в 2 раза. Данный метод влияет на выбор оптимальных размеров тайлов, размеры можно выбирать большими. Для данной перестановки был разработан метод расчёта оптимальных размеров тайлов. От предшествующих методов представленный в диссертации метод отличается динамической локализацией данных, у предшественников рассматривались модели, в которых сначала все необходимые данные размещаются в кэш-памяти, после чего ведутся вычисления. В модели, предложенной в диссертации, использованные для вычисления данные вытесняются из кэш-памяти, уступая место новым. В частности, для последовательного выполнения одна из размерностей тайла – чем больше тем лучше (тайл может не помещаться даже в L3-кэш).

В работе предложен новый режим диалоговой оптимизации программ. У существующих компиляторов есть возможность использования прагм для указания команд компилятору. Для этого пользователь должен иметь достаточные знания и квалификации. Предложенный подход подсказывает пользователю нужное место и нужную прагму для использования. При этом приводятся примеры, в которых компилятор автоматически не имеет права вставить такую прагму, потому что будут нарушены информационные зависимости.

Рекомендации по использованию результатов и выводов диссертации

Предложенные в данной работе методы могут быть реализованы при создании инструментов (компиляторов, конверторов) автоматического (и полуавтоматического) распараллеливания и оптимизации программ. Это позволит сократить время разработки высокопроизводительных программ, тем самым повышая производительность труда программистов.

Рассматриваемые в работе алгоритмы итерационного типа встречаются в задаче имитации распределения температуры в организме человека при лечении рака гипертермией, в численных методах решения дифференциальных уравнений электростатики, в моделировании влияния факельной установки на деградацию вечной мерзлоты и в других прикладных задачах.

Замечания по диссертационной работе

1. Используется не стандартная терминология, что затрудняет понимание работы. Например, вместо «совершенного гнезда циклов»¹ (perfect loop nest) используется термин «тесное гнездо циклов», вместо «графа зависимостей по данным»² (data dependence graph) используется термин «граф информационных зависимостей», а термин «сложная информационная зависимость» используется, но явно не определяется.
2. Какова область применимости предложенного метода? В работе хорошо рассмотрен алгоритм Гаусса-Зейделя для решения разных задач, однако помимо него присутствует только алгоритм решения задачи теплопроводности.
3. Приводится сравнение с небольшим числом аналогов, хотя аналоги указаны в обзоре автора (PolyMage³, SUIF⁴, Polly-llvm⁵, MLIR) и есть стандартные наборы тестов для оценки производительности типа NAS Parallel Benchmarks.
4. Непонятно, почему во время режима диалоговой оптимизации программ диалог системы с пользователем происходит на английском языке.
5. Присутствуют также опечатки. Например, в заголовке таблицы 3.12 (стр. 92) написано «без использования циклов внутри тайла», однако, скорее всего, имелось в виду «с использованием перестановки циклов внутри тайла.»

Приведённые замечания не снижают научной и практической ценности полученных результатов и не влияют на общую положительную оценку диссертационной работы. Диссертация выполнена на высоком научном уровне.

1 Векторизация программ: теория, методы, реализация. — М.: Мир, 1991

2 Векторизация программ: теория, методы, реализация. — М.: Мир, 1991

3 Mullapudi R.T., Vasista V., Bondhugula U. Polymage: Automatic optimization for image processing pipelines // SIGARCH Comput. Archit. News. 2015. Vol. 43, № 1. P. 429–443.

4 SUIF (Stanford University Intermediate Format). [Электронный ресурс]: URL: <https://suif.stanford.edu/>

5 Grosser T., Groslinger A., Lengauer C. Polly – performing polyhedral optimizations on a low-level intermediate representation // Parallel Processing Letters. 2012. Vol. 22, № 4. P. 1–27

Заключение

На основании вышеизложенного считаем, что диссертационная работа Метелицы Елены Анатольевны «Автоматизация распараллеливания программ со сложными информационными зависимостями» по своему научному уровню, практической значимости, степени новизны полученных результатов соответствует всем требованиям п. 9 «Положения о присуждении ученых степеней» ВАК РФ, предъявляемым к кандидатским диссертациям (Постановление Правительства РФ от 24.09.2013 № 842 (ред. от 11.09.2021) "О порядке присуждения ученых степеней" (вместе с "Положением о присуждении ученых степеней")), а ее автор Метелица Е.А. заслуживает присуждения ей ученой степени кандидата физико-математических наук по специальности 2.3.5 — «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Диссертация обсуждена на заседании № 1008 Объединенного семинара «Конструирование и оптимизация программ» Института систем информатики СО РАН и Новосибирского государственного университета 14 января 2025 г. и рекомендована к защите.

Руководитель семинара: _____



Касьянов Виктор Николаевич

Доктор физико-математических наук, профессор,
главный научный сотрудник, и.о. завлаб «Конструирование и оптимизация программ»

Института систем информатики им. А.П. Ершова СО РАН
Россия, 630090, Новосибирск-90, Пр. Академика Лаврентьева, 6

Тел.: +7(383) 330-40-47, e-mail: kvn@iis.nsk.su

Сведения о ведущей организации:

Федеральное государственное бюджетное учреждение науки

Институт систем информатики имени А. П. Ершова Сибирского отделения РАН

Адрес: 630090, г. Новосибирск, проспект Академика Лаврентьева, 6.

Веб-сайт: <https://www.iis.nsk.su/>

Адрес электронной почты: iis@iis.nsk.su