

Отзыв научного руководителя

Кандидатская диссертация Елены Анатольевны Метелицы относится к теории преобразования программ и посвящена автоматизации ускорения (оптимизации и распараллеливанию) программ со сложными информационными зависимостями. В большей части диссертации исследуется ускорение гнезд циклов, которые требуют наибольшего времени выполнения программы и их ускорение наиболее существенно для повышения производительности программы.

До работы Елены Метелицы была статья

S. G. Ammaev, L. R. Gervich, B. Y. Steinberg: Combining Parallelization with Overlaps and Optimization of Cache Memory Usage// Parallel Computing Technologies: 14th International Conference, PaCT 2017, Nizhny Novgorod, Russia, September 4-8, 2017, Proceedings, p. 257-264. DOI: 10.1007/978-3-319-62932-2-24 https://doi.org/10.1007/978-3-319-62932-2_24

В этой статье «вручную» выполнялось ускорение алгоритма Гаусса-Зейделя для решения задачи Дирихле. Перед диссертанткой была поставлена задача создать инструмент для автоматизированной реализации этого преобразования. Автоматизация оправдывалась тем, что подобный метод может быть применим для ускорения многих других алгоритмов, а получаемый быстрый код оказывается большим, логически сложным и требующим много времени для ручной разработки.

После реализации автоматизации ускоряющей цепочки преобразований данного алгоритма были найдены близкие работы (основанные на распараллеливающей системе PLUTO), в которых были описаны реализации некоторых основных преобразований. Эти преобразования были реализованы на полиэдральном внутреннем представлении программ, которое ориентировано только на преобразования пространства итераций гнезда циклов.

В чем достижения Е. Метелицы, если преобразования скашивание, тайлинг, метод гиперплоскостей, вынос инвариантов, перестановка циклов и пр. существуют в других распараллеливающих системах?

Достижение Елены Анатольевны состоит в следующем:

- Собрала с обоснованием эквивалентности в одну цепочку в одной компилирующей системе (ОПС) преобразования 1) скашивание, тайлинг, метод гиперплоскостей, которые существовали в системе PLUTO; 2) вынос инвариантов, перестановка циклов (LLVM); 3) линеаризация выражений, вывод из внутреннего представления в язык C для использования многими компиляторами языка C, в том числе, с закрытым кодом (ОПС).
- Обосновала теоретически с экспериментальным подтверждением целесообразность перестановки циклов при обходе точек тайла и применила это преобразование.
- Определила теоретически с экспериментальным подтверждением оптимальные размеры тайла.
- Продемонстрировала экспериментально целесообразность применения дополнительных преобразований «вынос инвариантов» и «линеаризация выражений».
- Продемонстрировала возможность диалогового режима компиляции, способного существенно расширить множество автоматизировано ускоряемых программ.

Следует отметить, что сборка нескольких компиляторных преобразований в цепочку – это не простой вызов последовательности выполняющих преобразования функций. Сложные преобразования применяются к фрагментам программы некоторого стандартного вида. Например, для практически всех преобразований цикл должен быть с левой границей 1 и инкрементом 1, а, если это не выполняется, то другим преобразованием (do loop normalization) цикл должен быть приведен к такому виду. Преобразования могут портить стандартный вид цикла, а цикл в нестандартной форме может работать быстрее, чем в стандартной. Еще пример: для построения графа информационных связей в индексных выражениях не должно быть переменных, кроме счетчика цикла, но приведение кода к требуемому виду может усложнить индексные выражения и вызвать замедление.

Задача Елене Анатольевне досталась сложная. Есть в Интернете статья Евгения Зуева «Редкая профессия» – это о разработчиках компиляторов. Но разработка оптимизирующих компиляторов – еще более редкая (и сложная) профессия, поскольку к разработке компилятора добавляется область математики «оптимизация». Е.А. Метелица овладела значительной частью этой сложной профессии (полностью этой профессией не владеет никто хотя бы потому, что эта область деятельности развивается).

Отметим также (неожиданный факт), что при новом обходе точек тайла одна из оптимальных размерностей тайла при последовательном вычислении может быть предельно большой и, при этом, используемые тайлом данные не помещаются даже в L3-кэш.

Елене Анатольевне удалось получить в 1.5 раза большего ускорения, чем у предшественников. Такой результат достигнут за счет реализации упомянутых преобразований на основе ОРС (Оптимизирующей распараллеливающей системы) и добавления к ним дополнительных преобразований. Преимущества ОРС в универсальности внутреннего представления, что и позволяет делать некоторые дополнительные преобразования, которые нельзя делать в полиэдральном внутреннем представлении.

Поиск оптимальной цепочки преобразований – основная проблема современных оптимизирующих компиляторов. Об этом есть исследования 2018 года

1. Z. Gong, Z. Chen, Z. Szaday, D. Wong, Z. Sura, N. Watkinson, S. Maleki, D. Padua, A. Veidenbaum, A. Nicolau. An empirical study of the effect of source-level loop transformations on compiler stability. Proceedings of the ACM on Programming Languages. 2. 1-29. 10.1145/3276496. 2018.

Наиболее эффективная реализация цепочки преобразований «скашивание»-«тайлинг»-«метод гиперплоскостей», кроме реализации Елены, основаны на распараллеливающей системе PLUTO. Ускорение у Е. Метелицы в 1.5 раза больше (в 16 раз у Елены против 11 раз у U.Bondhugula). Для сравнения с PLUTO Елена Анатольевна потратила много времени. Скачала открытый код, изучала условия запуска и тестировала на тех же данных, что и свою программу. Возникает естественный вопрос «за счет чего?». Анализ исходного кода и публикаций показал, что основанная на PLUTO цепочка преобразований не включает в себя такие оптимизирующие преобразования, как «поиск и вынос инвариантов из циклов», «линеаризация арифметических выражений» (упрощение выражений типа школьного формально некорректного понятия «приведение подобных»), перестановку циклов внутри тайла. А другие преобразования не могут быть реализованы на основе внутреннего представления PLUTO. Можно добавить, что преобразование

«линеаризация выражений» есть только в ОРС, а в других компилирующих системах есть только некоторые частные случаи этого преобразования. Таким образом, цепочка преобразований Е. Метелицы отличается от цепочки преобразований PLUTO дополнительными ускоряющими преобразованиями. Более того, наличие в ОРС библиотеки преобразований программ позволяет в перспективе расширить множество оптимизируемых программ, приводя их к нужному шаблону.

Помимо отмеченных преимуществ по сравнению с PLUTO, реализация ускоряющего гнезда циклов алгоритма на основе высокоуровневого внутреннего представления ОРС имеет перспективные преимущества по сравнению с LLVM и другими системами, основанными на низкоуровневом внутреннем представлении.

Реализуемые Еленой Анатольевной преобразования основаны на анализе сложных характеристик информационных зависимостей в программе. В работе используется вектор расстояния зависимости, который в общем случае сложно определяется, но для гнезд циклов итерационного типа его определить можно так, что по нему удастся вычислять параметры скашивания и определять корректность применения прямоугольного тайлинга и перестановки циклов.

Контроль корректности выполняемых преобразований проходит через всю диссертационную работу. Корректность выполняемых преобразований является чем-то само-собой разумеющимся. В отличие от производительности, корректность трудно измерить, чтобы заявить, как техническое достижение. Но соблюдение корректности требует больших усилий. Условия корректности отслеживаются на каждом шаге цепочки преобразований и формулируются в виде теоремы.

Для многих программ бывает сложно определить условия применимости оптимизирующих преобразований, и компиляторы в таких случаях позволяют разработчику вставлять в текст программы специальные прагмы. Но надо знать, в какие места программы вставлять прагмы, и знать, какие прагмы. Это требует высокой квалификации разработчиков программ, близкой к квалификации разработчика компиляторов. Для решения этой проблемы в четвертой главе диссертации предлагается диалоговый режим работы компилятора. Этот режим отличается от известных интерактивных методов преобразования программ тем, что специальный анализатор компилятора должен выдавать разработчику связанные с текстом программы вопросы и, на основании ответов, формировать оптимизированный (например, параллельно выполняемый) код. Следует отметить, что высокоуровневое внутреннее представление ОРС в данном случае имеет преимущество по сравнению с низкоуровневыми внутренними представлениями компиляторов GCC, ICC, LLVM.

Елена Анатольевна отличается большой самостоятельностью: сначала я предлагал Елене статьи для изучения, а последнее время, наоборот, Елена Анатольевна присылает статьи мне для ознакомления. Представленной кандидатской диссертации предшествовали 8 лет аккуратных исследований в области оптимизирующей компиляции: 2 года в бакалавриате, 2 в магистратуре и 4 в аспирантуре. За это время Елена Анатольевна активно участвовала в научных конференциях и семинарах, прошла по конкурсу в Летнюю школу Интел, руководила командой студентов в инженерной олимпиаде, где эта команда заняла 1 место на региональном туре, была старостой группы очень сильных аспирантов, опередив их в написании диссертации и подав им этим самым пример.

Из десятка подготовленных мною кандидатов наук у Е.А. Метелицы диссертационное исследование одно из наиболее трудных.

Результаты исследований Елены Анатольевны предполагают развитие (которое уже ведется новыми участниками группы разработчиков ОРС). Рассматривается возможность сводимости алгоритмов решения широкого класса задач линейной алгебры к вычислению гнезда программных циклов итерационного типа, к которому применимы преобразования Е.А. Метелицы. Рассматриваются дополнительные ускорения самого преобразования Е.А. Метелицы, одно из которых состоит в быстром поиске начального приближения за счет методов типа «квантизации нейронных сетей». А другое – за счет перехода к ромбическим тайлам. Методы ускорения программ в представленной диссертации демонстрируются, в основном, на нескольких примерах численного решения двумерных эллиптических задач математической физики, но уже видно, как их можно распространять и на уравнения параболического типа и другие.

Считаю, что в диссертационном исследовании Елены Анатольевны решается сложная и полезная задача автоматизации ускорения практически значимых программ, Елена Анатольевна владеет сложной (очень) и полезной квалификацией разработчика оптимизирующих компиляторов и заслуживает ученой степени кандидата наук.

27.05.2024г



Штейнберг Борис Яковлевич, доктор технических наук, старший научный сотрудник, заведующий кафедрой Федерального государственного автономного образовательного учреждения высшего образования «Южный федеральный университет», научный руководитель Ростов-на-Дону, Мильчакова 8а
borsteinb@mail.ru, +79094110777

Аргументы прос. подтверждаю



Владислав Березин Б.Я.
(Мильчакова 8А)