

Управляемые пакеты оптимизаций

А.Ю. Дроздов, А.М. Степаненков

Аннотация. В работе предлагается метод организации оптимизаций в виде управляемых пакетов. По сравнению с классической организацией, представляющей собой последовательный запуск преобразований, он позволяет более эффективно использовать оптимизации и при этом сократить время компиляции программы.

Введение

При создании современных компиляторов, отличающихся большим количеством оптимизирующих преобразований, на первый план выходят проблемы структурирования оптимизаций и организации их взаимодействия. Ввиду того, что оптимизации могут быть конфликтующими (применение одной исключает применение другой) или связанными (невозможность применить одну распространяется на другую) для эффективного использования возможностей компилятора необходимо подбирать последовательность запуска оптимизаций. Очень важно также, чтобы несмотря на внушительный список оптимизаций, реализованных в компиляторе, процесс компиляции укладывался в допустимое время. Разрабатывая конкурентоспособный оптимизирующий компилятор, необходимо предусмотреть и возможность его развития. В этой связи важно, чтобы уже введенные оптимизации просто адаптировались к нововведениям.

Все эти факторы требуют организовать оптимизации таким образом, чтобы обеспечить контроль сразу над несколькими из них, иметь возможность выполнять новые требования и обеспечивать необходимые свойства.

В работе рассматривается способ организации преобразований в виде управляемых пакетов оптимизаций, позволяющий эффективно решать упомянутые проблемы. Он является альтернативой классической последовательной организации оптимизаций [3]. Тема комбинирования различных методов анализа потока данных и оптимизаций затрагивалась и ранее [4]. В настоящей работе даётся общий подход к решению проблемы организации оптимизаций, не зависящий от их внутренней структуры? и приводятся примеры конкретных реализаций управляемых пакетов оптимизаций в рамках промышленного компилятора, разработанного для архитектуры «Эльбрус» [5,6].

1. Определение оптимизации

Объектом оптимизации является выработанное компилятором представление фрагмента кода программы (например, операции? пары операций, узла управляющего графа, цикла, процедуры), которое достаточно для получения его двоичного кода. Оптимизации на уровне операций принято называть низкоуровневыми, на уровне узлов управляющего графа – макрооперациями.

Оптимизация выполняется в результате следующих действий (Рис.1):

1. Анализ, при котором проверяется возможность, подав на вход оптимизирующего преобразования исходное промежуточное представление, получить на выходе промежуточное представление, функционально эквивалентное исходному.

2. Проверка эффективности оптимизации.

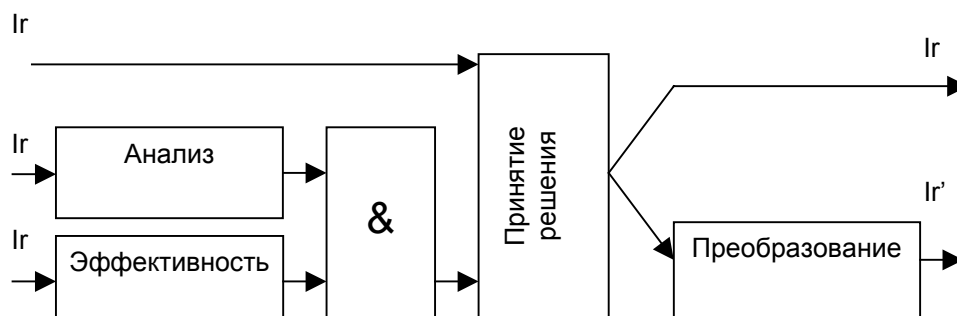


Рис. 1. Схема процесса оптимизации

Пусть зафиксирована Q – функция на промежуточном представлении, принимающая значение на множестве неотрицательных вещественных чисел (*функция качества оптимизации*), Ir – промежуточное представление до оптимизации, Ir' – промежуточное представление после оптимизации. Принимается, что оптимизация эффективна, если $Q(Ir') \geq Q(Ir)$.

3. Преобразование исходного промежуточного представления.

2. Структура управляемого пакета оптимизаций

Управляемый пакет оптимизаций – это пара $\langle P, M \rangle$, где P – пакет оптимизаций, M – менеджер пакета оптимизаций. Пакет оптимизаций представляет собой набор оптимизаций с одним и тем же объектом применения. Менеджер оптимизаций определяет порядок запуска оптимизаций пакета. Он может выполнять дополнительный анализ на применимость и эффективность оптимизаций, входящих в его пакет. Помимо этого, менеджер оптимизаций может сам выполнять ряд преобразований промежуточного представления, имеющих тот же тип, что и элементы пакета. Например, если объект оптимизации – операция, то менеджер пакета может преобразовать узел управляющего графа, если – цикл, то преобразуется дерево циклов [1]. В этом случае у оптимизаций пакета есть возможность не выполнять преобразования, а заказать композицию преобразований своему менеджеру. Иллюстрация этого принципа приведена на Рис. 2.

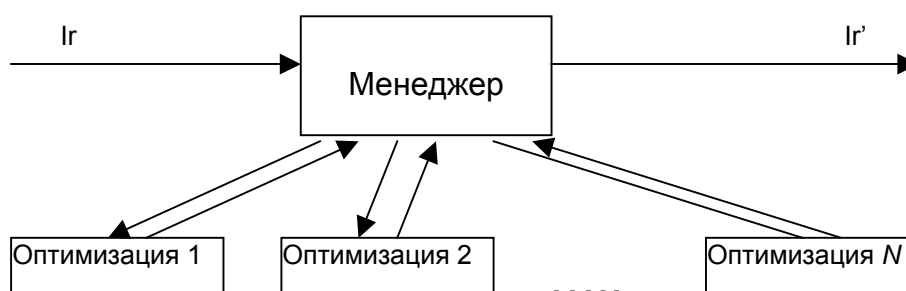


Рис. 2. Структура управляемого пакета оптимизаций

Далее приведём конкретные примеры управляемых пакетов оптимизаций, реализованных в компиляторе.

3. Пакет низкоуровневых оптимизаций на потоке данных

Объектом применения оптимизаций пакета является операция. Дадим краткое описание оптимизаций, входящих в пакет.

1. Удаление мёртвого кода. Удаляются операции, результаты которых далее нигде не используются, а потому их присутствие не влияет на поведение программы при исполнении.

2. Подстановка констант. Если удалось определить, что результатом операции является константа, то операция удаляется, а константа подставляется в места использования результата операции.

3. Сбор общих подвыражений. Если имеются две эквивалентные операции, то одна из них удаляется, а во все места использования ее результата подставляется результат второй. Если операции выполнялись условно, то оставшаяся операция рассматривается как предикат, равный логическому ИЛИ предикатов эквивалентных операций.

4. Эквивалентные преобразования выражений, состоящих из операций представления. Например, $x+0=x$, $x*1=x$, $x*8=x<<3$, $(x+1)+2=x+3$.

5. Удаление избыточных операций чтения из памяти. Если операции чтения можно поставить в соответствие предшествующую запись в память по тому же адресу (с форматом, не меньшим чем у этой операции чтения) и между этими операциями нет других, записывающих по тому же адресу, то операция чтения удаляется, а в местах её использования подставляется записанное ранее значение.

6. Удаление избыточных записей в память. Обрабатываются два случая. В первом записываемое значение является результатом чтения из памяти, ранее выполненного по тому же адресу, и между этим чтением и записью нет операций, модифицирующих ту же ячейку памяти. Здесь операция записи является избыточной и поэтому удаляется. Во втором случае за первой операцией записи следует другая запись (с не меньшим форматом) по тому же адресу, причем между двумя этими операциями записи нет чтений из данной ячейки памяти. Здесь удаляется первая операция записи.

Опишем менеджер пакета, включающего приведенные оптимизации. Минимальным объектом его применения является узел управляющего графа, максимальным – вся процедура. При обработке процедуры узлы управляющего графа обходятся в порядке, который обеспечивает то свойство, что при переходе к следующему узлу, все его предшественники в графе уже пройдены (при этом обратные дуги не учитываются [1]). Внутри узла операции обрабатываются в порядке их следования. Последовательность операций узла можно упорядочить по некоторому правилу и использовать это свойство при оптимизации. Далее для каждой операции запускается пакет оптимизаций в том порядке, в котором они были представлены выше. Запуск пакета для операции осуществляется в цикле, условием выхода из которого является неприменимость ни одной из оптимизаций или удаление операции.

Оптимизации пакета не выполняют преобразования промежуточного представления, а заказывают набор преобразований у менеджера пакета. Менеджер пакета предоставляет следующие типы преобразования промежуточного представления:

- подстановка вместо результата операции константы;
- удаление операции по одному из правил;
- замена результата операции на результат уже существующей операции;
- замена результата операции на результат дерева новых операций, когда новые операции создаются по некоторому их формальному описанию.

Вынесение преобразований с уровня оптимизаций на уровень менеджера пакета имеет много преимуществ. Во-первых, усиливается контроль над оптимизациями и увеличивается гибкость пакета по отношению к изменениям промежуточного представления. Зная тип преобразования, которое необходимо выполнить, менеджер может в зависимости от контекста, в котором он вызывается, отменять ту или иную оптимизацию, некорректную в этом контексте. Поскольку все преобразования выполняет менеджер, при изменении свойства промежуточного представления достаточно отразить это изменение только в нём. Во-вторых, многие потоковые оптимизации существенно различаются только анализом и сильно пересекаются по типам преобразований промежуточного представления. Выделение основных классов преобразований, применяющихся в низкоуровневых оптимизациях, и их поддержка менеджером пакета избавляет от излишнего дублирования текстов и существенно облегчает отладку компилятора.

Управляемый пакет потоковых оптимизаций имеет высокую эффективность при сравнительно небольших затратах по времени. Большинство оптимизаций используют одинаковые структуры данных, требующие инициализации. При объединении оптимизаций в пакет инициализация дополнительных структур данных делается один раз. Известно, что применение одних оптимизаций даёт возможность применять другие. Добиться эффективности, соизмеримой с эффективностью управляемого пакета, можно многократными вызовами отдельных оптимизаций, но этот способ требует больших временных ресурсов. В управляемом пакете многократные вызовы оптимизаций осуществляются только для тех операций, которые этого требуют. Для большинства операций делается всего один проход по всем оптимизациям.

4. Пакет цикловых макрооптимизаций

Объектом применения оптимизаций пакета является цикл. Перечислим оптимизации, входящие в пакет.

1. Слияние внутреннего цикла с охватывающим циклом. Если i – индуктивная переменная внешнего цикла, а j – индуктивная переменная внутреннего цикла, то вводится такая индуктивная переменная k , что любому индексу цикла $f(i,j)$ ставится в соответствие индекс $f'(k)$, $f(i,j) = f'(k)$ на каждой итерации цикла. Внутренний и внешний циклы преобразуются в новый цикл с индуктивной переменной k [1].

2. Перестановка циклов. Если i – индуктивная переменная внешнего цикла, а j – индуктивная переменная внутреннего цикла, то в качестве индуктивной переменной внешнего цикла вводится j , внутреннего – i , причем индексы $f(i,j)$ заменяются на $f'(j,i)$ так, что $f(i,j) = f'(j,i)$ на каждой итерации внутреннего цикла.

3. Полная раскрутка цикла. Если число итераций цикла постоянно и задается константой C , то цикл можно представить в виде повторения его тела C раз.

4. Частичное повторение тела цикла (открутка). В случае, когда число повторений цикла не является константой, можно сделать несколько повторений тела цикла на основе профильной информации с последующей передачей управления на исходный цикл. При этом вероятность перехода на цикл после открутки приближается к нулю [1].

5. Расщепление цикла по инвариантному условию. Для цикла создаётся точная копия. Управление передаётся на цикл или его копию в зависимости от предиката инвариантного условия. В цикле инвариантное условие считается тождественной истиной, в копии – тождественной ложью.

6. Расщепление цикла по индексу. Если в цикле имеется условие, ограничивающее изменение его индуктивной переменной, то он разбивается на два цикла. В первом ограничение, заданное условием, берётся в качестве верхней границы, во втором цикле – в качестве нижней границы. В обоих циклах убираются вычисления, соответствующие условию выхода значения индуктивной переменной за ее границы.

7. Расщепление цикла по произвольному условию. В случае более общего условия для некоторого класса циклов возможно расщепление одного цикла на три. Первый цикл не содержит альтернатив условия. Второй цикл содержит первую альтернативу условия. Третий цикл содержит вторую альтернативу условия. Число итераций первого цикла равно числу итераций исходного цикла. Среднее число итераций второго цикла равно среднему числу повторений первой альтернативы. Среднее число итераций третьего цикла равно среднему числу повторений второй альтернативы. Значения переменных исходного цикла, используемых в альтернативах, запоминаются в массивы в первом цикле и считываются во втором и третьем циклах.

8. Разрыв статически неразрешённых зависимостей по памяти. Во многих случаях для пары операций чтения или записи, заданных в цикле, невозможно статически определить, могут ли они обращаться к одной и той же ячейке памяти во время его исполнения. Иногда вне цикла можно сформировать цепочку условий, выполнимость которых обеспечивает независимость по памяти

некоторых пар. В этом случае строится точная копия цикла, в которой указанные пары операций считаются независимыми; на нее передаётся управление в случае истинности цепочки условий.

Объектом применимости управляемого пакета является дерево циклов. Перед описанием менеджера пакета сформулируем основные задачи, которые он должен решать.

Первая задача – минимизировать время оптимизации цикловых участков. Обычно как проверка применимости и эффективности цикловых макрооптимизаций, так и преобразования промежуточного представления не являются критичными по времени. Значительную часть времени исполнения таких оптимизаций занимает построение глобальных (для всей процедуры) аналитических структур данных, которые являются общими для всех оптимизаций. Причём в силу нетривиальности оптимизирующих преобразований промежуточного представления, коррекция этих структур может быть нелокальной (в отличие от пакета низкоуровневых оптимизаций, где коррекция является несложной и передана менеджеру пакета). Поэтому часто самым хорошим решением при коррекции аналитических структур данных является их перестроение. Требуется так организовать работу с оптимизациями, чтобы сократить число перестроений глобальных структур данных.

Многие из описанных оптимизаций, входящих в пакет, приводят к значительному увеличению кода процедуры. Бесконтрольное увеличение кода процедуры приводит к резкому возрастанию времени обработки компилятором такой процедуры. От многих глобальных оптимизаций, имеющих нелинейную сложность, нужно будет отказаться, в результате чего полученный код процедуры будет некачественным. Поэтому задача менеджера пакета цикловых макрооптимизаций состоит и в том, чтобы не допускать значительного роста кода, при этом эффективно используя оптимизации, активно дублирующие циклы.

Третья задача – разрешение конфликтов между оптимизациями. Области применимости некоторых оптимизаций существенно пересекаются. Можно применить несколько оптимизаций к одному и тому же циклу и выполнить по сути одно и то же преобразование разными способами (например, оптимизации расщепления цикла). В некоторых случаях применимость одной оптимизации зависит от применимости другой. Это также необходимо учитывать при управлении пакетом.

Опишем алгоритм обработки дерева циклов.

Для каждого цикла введём четыре различных состояния: UNDEF, PROCESSED, WAIT, CLOSED. В начале все циклы находятся в состоянии UNDEF. На Рис. 3 приведён алгоритм, по которому работает менеджер пакета цикловых макрооптимизаций.

```

LoopMacroOptimizationManager()
{
    MAX_INCREASE = CalcMaxIncrease();
    loops_list = FormInnermostLoopsList();
    SortLoopsListByProfile( loops_list);
    ConvertLoopsListState( loops_list, UNDEF, PROCESSED);
    loop = GetFirstLoopFromList(loops_list);
    while ( 1 )
    {
        if ( ApplyOptimizationsForLoop( loop, loops_list ) )
        {
            if ( IsLoopNotDead( loop ) )
            {
                SetLoopState( loop, WAIT);
            }
            loop = GetNextLoopFromList( loops_list, loop);
        } else
        {

```

```

    SetLoopState( loop, CLOSED);
    loop = UpdateLoopsListAfterCloseLoop( loops_list, loop);
  }
  if ( loop != NULL )
  {
    continue;
  }
  loop = GetFirstLoopFromList(loops_list);
  if ( loop == NULL )
  {
    break;
  }
  ReconstrAllStructures();
  SortLoopsListByProfile(loops_list);
  ConvertLoopsListState( loops_list, WAIT, PROCESSED);
  loop = GetFirstLoopFromList(loops_list);
}
}

UpdateLoopsListAfterCloseLoop( loops_list, loop)
{
  pred = GetLoopPredInTree( loop);
  if ( IsAllLoopSucCsClosed( pred ) )
  {
    SetLoopState( pred, PROCESSED);
    InsertLoopInLoopListAfterLoopByProfile( loops_list, pred, loop);
  }
  return RemoveLoopFromList( loops_list, loop);
}

```

Рис. 3. Алгоритм обработки дерева циклов менеджером пакета цикловых макрооптимизаций

CalcMaxIncrease() вычисляет максимальный допустимый размер процедуры после всех преобразований.

FormInnermostLoopsList() формирует список самых вложенных циклов (не имеющих преемников в дереве циклов).

SortLoopsListByProfile(loops_list) сортирует список циклов в соответствии с их важностью (на основе профильной информации).

ConvertLoopsListState(loops_list, state1, state2) заменяет состояние цикла *state1* на *state2* в списке циклов.

GetFirstLoopFromList(loops_list) - получение первого цикла из списка циклов.

ApplyOptimizationsForLoop(loop) пытается реализовать цикловые оптимизации в том порядке, в котором они были описаны выше. При применении некоторой оптимизации происходит выход из процедуры с возможной коррекцией списка циклов (добавлением новых списков, возникших при преобразовании). При этом недопустимо увеличение размера процедуры. Для каждого цикла и оптимизации есть ограничение на рост кода, которое зависит от важности цикла, оптимизации и резерва общего роста кода процедуры.

IsLoopNotDead(loop) – проверка, что цикл был удалён при оптимизации.

SetLoopState(loop, state) - установка состояния цикла.

Reconstr.AllStructures() - перестроение всех аналитических структур.

GetLoopPredInTree(loop) – получение предшественника цикла в дереве циклов.

IsAllLoopSucclsClosed(loop) – проверка, что все преемники цикла в дереве циклов имеют состояние CLOSED.

InsertLoopInLoopListAfterLoopByProfile(loops_list, loop1, loop2) - вставка в список цикла *loop1* в соответствии с приоритетом, основанным на профильной информации, но не раньше цикла *loop2*.

RemoveLoopFromList(loops_list, loop) - удаление цикла из списка; возврат к циклу, следующему за удаляемым.

Обработка дерева циклов происходит от листьев к корню. Список циклов представляет собой сечение дерева циклов, состоящее из независимых циклов. Вначале он состоит из самых вложенных циклов. В случае, когда ни одна из оптимизаций пакета не применилась к циклу, он удаляется из списка и происходит попытка включить в список его предшественника в дереве. Для сохранения независимости сечения необходимо убедиться, что все преемники кандидата на включение были уже обработаны. Если к циклу была применена некоторая оптимизация, он переводится в состояние ожидания. Необходимо скорректировать глобальные структуры данных, которые стали некорректными для этого цикла, прежде чем можно будет снова запускать для него оптимизации. В силу независимости сечения циклы обрабатываются параллельно, то есть не происходит перестроения всех структур данных после применения оптимизирующего преобразования к одному циклу. При этом оптимизации могут быть применены к другим циклам сечения. Это существенно сокращает число перестроений глобальных структур данных, особенно для невысоких и широких деревьев циклов.

В представленной схеме невозможно бесконтрольное увеличение размера процедуры, так как менеджер пакета не допускает применения оптимизаций, приводящих к увеличению кода выше предельно допустимого. В то же время это ограничение сглаживается тем, что вначале обрабатываются самые важные циклы, затем менее значимые, согласно профилю. Таким образом, при обработке циклов, имеющих высокие приоритеты согласно профильной информации, резервов для дублирования кода должно быть ещё достаточно. Это верно для самых вложенных циклов. Для часто исполняемого, но не самого вложенного цикла возможна более ранняя обработка менее значимых циклов, когда один из преемников цикла находится в состоянии ожидания. В этом случае, как отмечалось выше, менеджер пакета не допускает увеличения размера процедуры при обработке цикла, если это приводит к исчерпанию резервов, отведенных для других, ещё не обработанных, циклов с более высокими приоритетами.

Конфликтующие оптимизации, выполняющие схожие преобразования, упорядочены по возрастанию их общности. Сначала запускаются более специализированные и, следовательно, более эффективные оптимизации, затем рассчитанные на самый общий случай. Основным критерием при разрешении конфликтов между оптимизациями является их эффективность. Как и в случае с пакетом низкоуровневных оптимизаций, запуск пакета макрооптимизаций также зациклен. Это решает проблему связанных оптимизаций (применимость одной оптимизации зависит от результата работы другой).

5. Макропакеты

Выше была предложена простейшая реализация цикловых макрооптимизаций в виде управляемого пакета оптимизаций. Рассмотрим более сложную организацию оптимизаций, позволяющую увеличить эффективность их применения. Многие цикловые оптимизации при принятии решений используют ресурсные оценки. Это означает, что они наиболее эффективны на оптимизированном промежуточном представлении. На представлении, содержащем мёртвый код и дублирующие вычисления, интеллектуальная часть оптимизации может отработать ошибочно, т.е. оптимизирован-

ное позже представление, на котором не применялась данная цикловая макрооптимизация, может оказаться более эффективным, чем оптимизированное позже представление, на котором она применялась. Поскольку многие оптимизации пакета используют технику дублирования представления, после их применения в управляющем графе могут возникать пути, которые никогда не исполняются, т.е. возникать мёртвый код на уровне управления. Присутствие избыточных вычислений может приводить и к отказу от оптимизации менеджером пакета на основании ограничения роста кода процедуры.

Проблема решается применением управляемого пакета низкоуровневых оптимизаций и канонизации управления. Она заключается в удалении избыточных ветвлений в управляющем графе на основании неизменности предикатов переходов с последующим вызовом сборщика мусора, удалении опустевших узлов, слиянии эквивалентных по управлению соседних узлов. Канонизацию управления следует запускать после работы низкоуровневых потоковых оптимизаций. Вызывать оптимизации можно каждый раз перед перестроением глобальных аналитических структур данных. Это не приводит к заметному увеличению времени работы всего алгоритма, так как низкоуровневые оптимизации и канонизация управления имеют меньшую сложность по сравнению с обновлением глобальных аналитических структур данных, однако значительно повышает его эффективность.

Пакет оптимизирующих преобразований, некоторые из которых представляют собой управляемые пакеты оптимизаций, называется макропакетом. Макропакеты с управлением называются управляемыми макропакетами.

Рассмотренный выше способ организации цикловых макрооптимизаций с применением оптимизаций низкого уровня и канонизации управления является примером управляемого макропакета.

6. Экспериментальные результаты

Как уже отмечалось выше, одним из преимуществ организации оптимизаций в виде управляемых пакетов является сокращение времени обработки программы компилятором. Особенно важен этот показатель для макрооптимизаций, занимающих основную часть времени работы компилятора. При этом самым трудоёмким процессом является построение глобальных аналитических структур данных, необходимых для их работы. С помощью организации макрооптимизаций в виде управляемых пакетов удаётся во многих случаях сократить число перестроений глобальных структур данных. В таблице приведена статистика соотношения числа срабатываний макрооптимизаций и числа перестроений аналитических структур на задачах пакета `spcs92`. Лучшие показатели достигаются на задачах, где достаточно большое число применений оптимизаций и где процедуры имеют невысокие, но широкие деревья циклов. Действительно, поскольку при применении оптимизаций пакета на процедуру необходимо, по крайней мере, одно перестроение структур данных, число таких применений должно быть больше единицы. Широкие деревья циклов обеспечивают большую параллельность при применении оптимизаций. Из таблицы следует, что «хорошую» структуру, позволяющую существенно сократить число перестроений глобальных структур данных, имеют задачи, содержащие вычисления над числами с плавающей точкой, `039.wave5` и `090.hydro2d`.

Заключение

В данной работе предложен способ организации оптимизаций в виде управляемых пакетов. Он обеспечивает более эффективное использование оптимизаций по сравнению с классической организацией, представляющей собой последовательный запуск преобразований. Организация оптимизаций в виде управляемых пакетов позволяет уменьшить затраты при реализации новых оптимизаций и развитии компилятора, сократить время компиляции программы.

Сравнение числа срабатываний макрооптимизаций и числа перестроений глобальных аналитических структур данных

Тест	Число срабатываний (n)	Число перестроений (r)	n/r
026.compress	1	1	1.000
023.eqntott	20	13	1.538
008.espresso	189	109	1.734
022.lisp	28	28	1.000
072.sc	51	39	1.308
085.gcc	202	126	1.603
052.alvinn	8	6	1.333
047.tomcatv	1	1	1.000
056.ear	12	8	1.500
094.fppp	8	5	1.600
078.swm256	7	5	1.400
093.nasa7	32	21	1.524
039.wave5	141	42	3.357
089.su2cor	29	17	1.706
090.hydro2d	77	22	3.500

Литература

1. Steven S. Muchnick, "Advanced Compiler Design and Implementation", Morgan Kauffman, San Francisco, 1997
2. J.R. Ellis, "Bulldog: A Compiler for VLIW Architectures", Doctoral Dissertation, MIT Press Cambridge MA 1985.
3. D.F. Bacon, S. L. Graham, O.J. Sharp, "Compiler Transformations for High-Performance Computing", ACM Computing Surveys, Vol. 26, No 4, December 1994.
4. Clifford N. Click, "Combining Analyses, Combining Optimizations", A thesis submitted in partial fulfillment of the requirements for the degree Doctor of Philosophy. Houston, Texas, February 1995.
5. Babayan B. A. E2k Technology and Implementation. // Proceedings of the Euro-Par 2000 – Parallel Processing: 6th International. – V. 1900/2000. – January, 2000. – P. 18-21.
6. K. Dieffendorf. The Russians Are Coming. Supercomputer Maker Elbrus Seeks to Join x86/IA-64 Melee // Microprocessor Report, V. 13, № 2. February 15, 1999. P. 1-7.