

Московский физико-технический институт
(государственный университет)

Факультет радиотехники и кибернетики

Кафедра информатики и вычислительной техники

Выпускная квалификационная работа бакалавра

Оптимизация времени компиляции путем автоматического распараллеливания попроцедурной части

СТУДЕНТ: СОКОЛОВ Б.Е., 213 ГРУППА

НАУЧНЫЙ РУКОВОДИТЕЛЬ: КАЛЯКИН А.А.

Актуальность

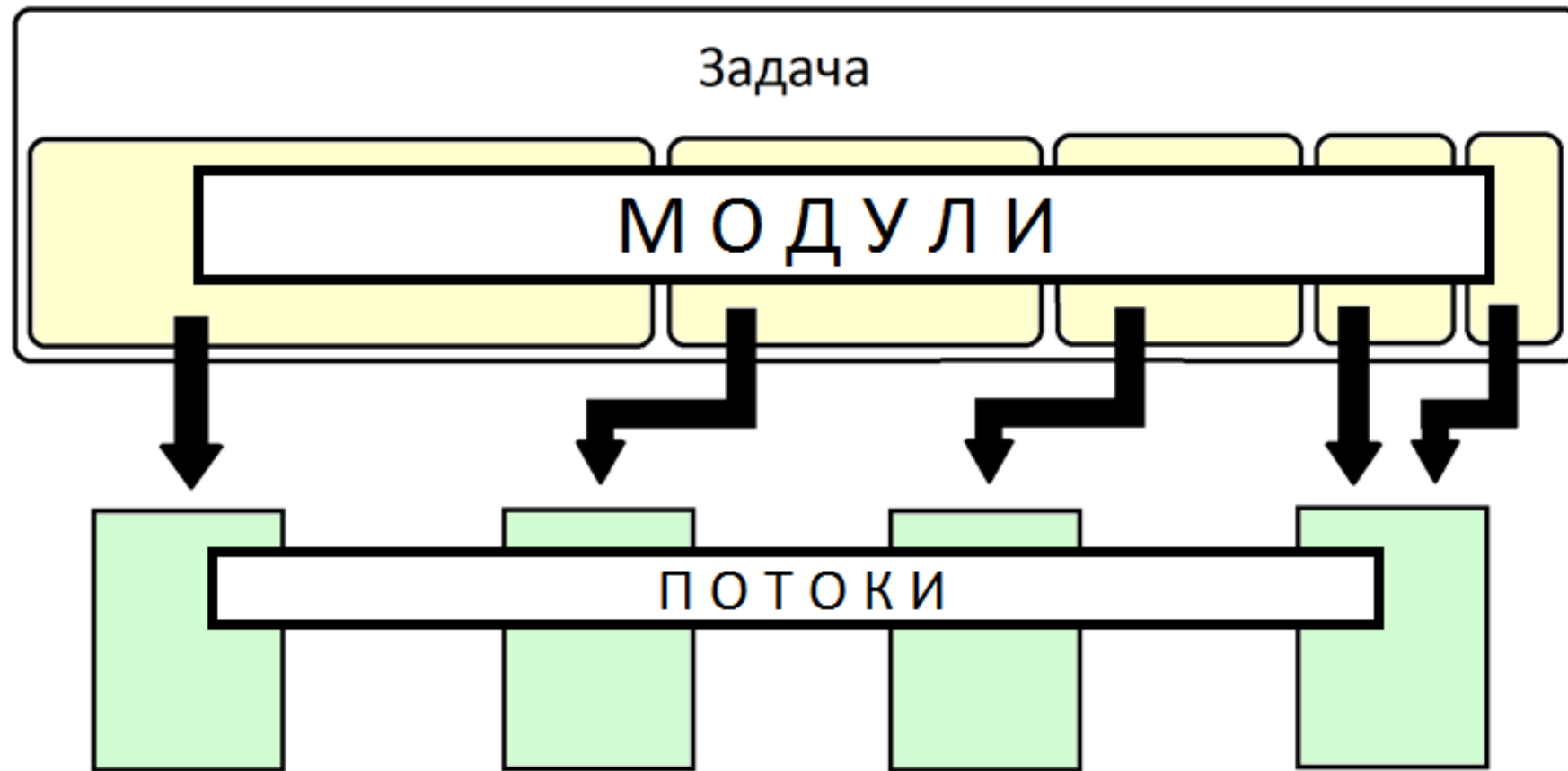
Рост размеров исходных кодов современных задач приводит к увеличению времени компиляции

Пакет SPEC CPU 2006, задача 483.xalancbmk

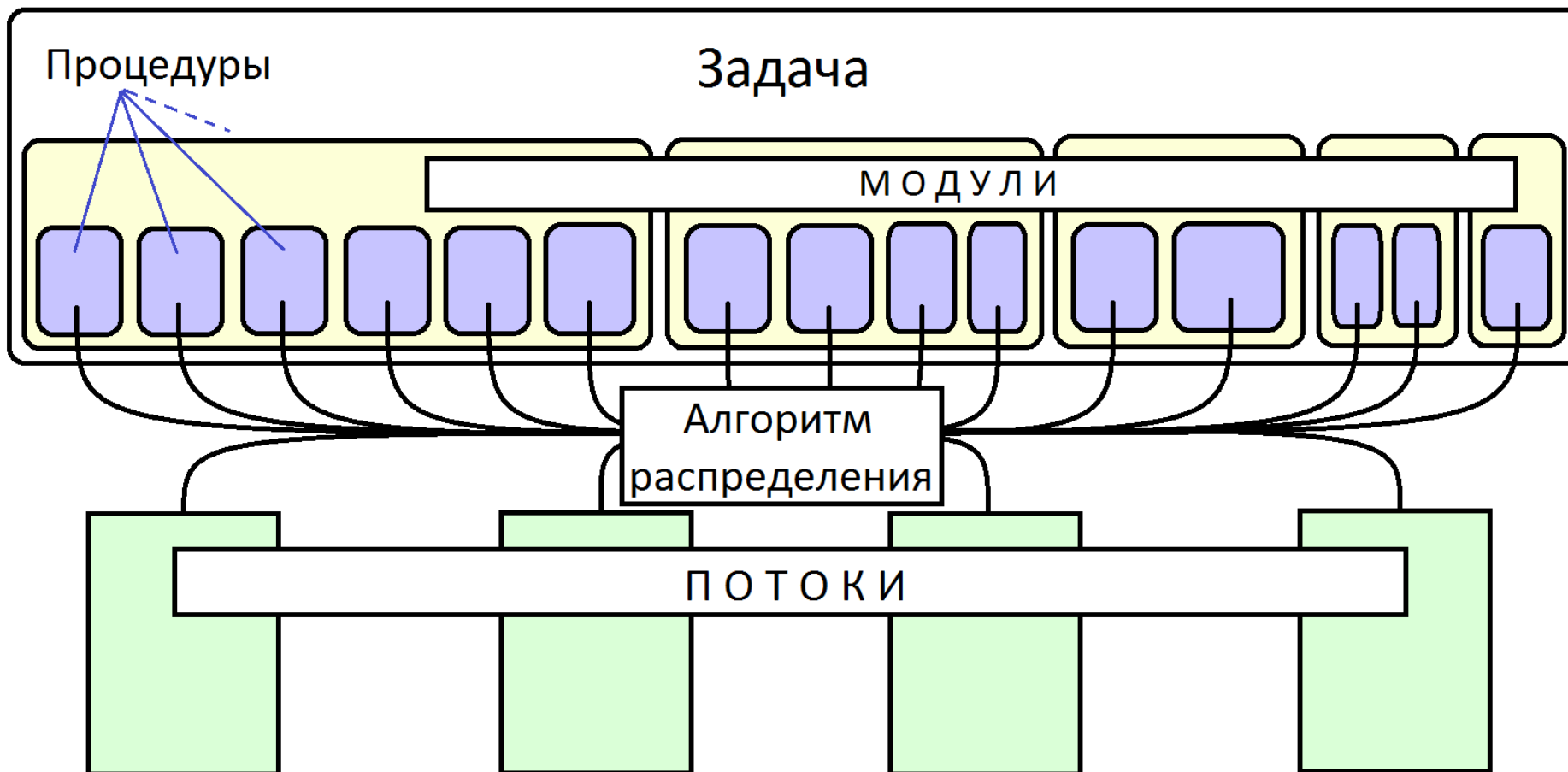
Размер кода – 22МБ

Время компиляции на машине с 4-х ядерным процессором x86 компилятором под архитектуру Elbrus – 8 минут

Распараллеливание на уровне модулей с помощью makefile



Предложенный метод



Постановка задачи

1. Оценить потенциальную эффективность распараллеливания компиляции на уровне процедур
2. Предложить алгоритм для распределения процедур по потокам
3. Проанализировать результаты применения алгоритма распределения на примере задач пакета SPEC CPU 2006

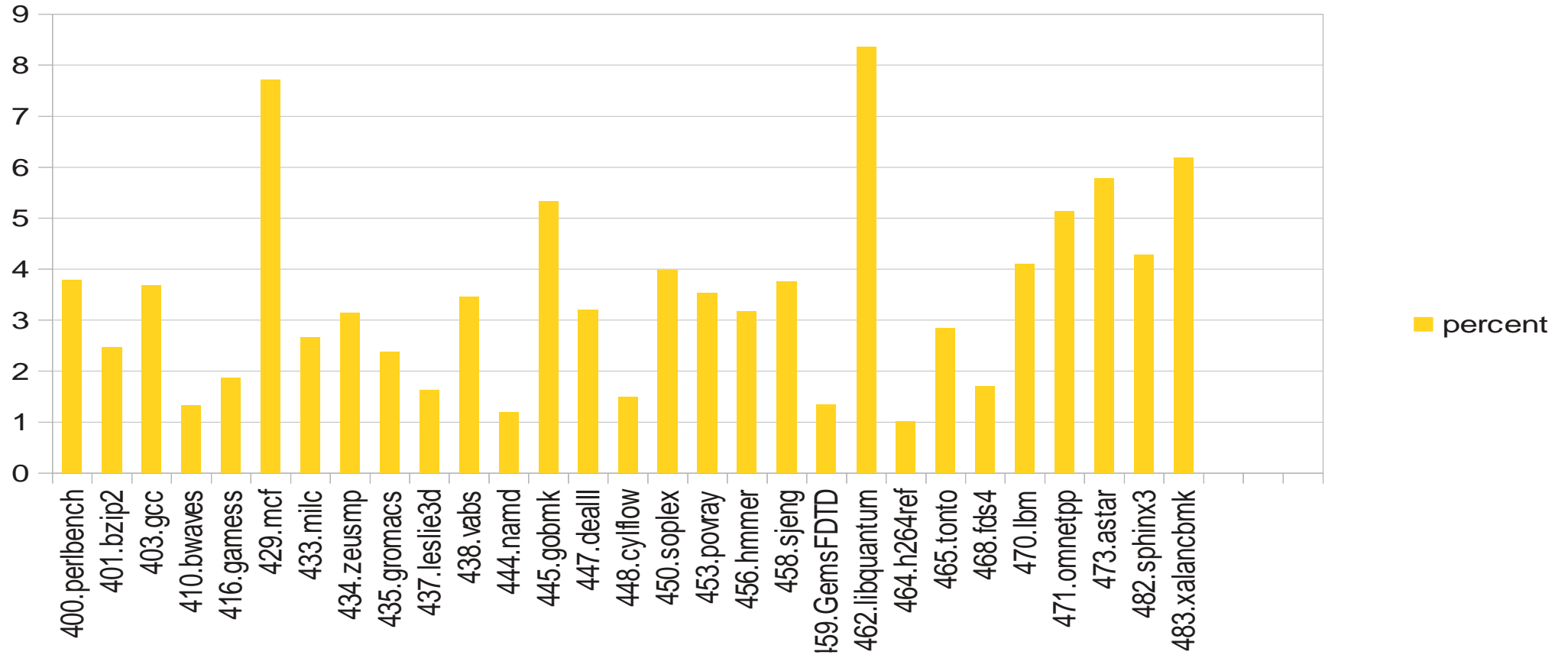
Структура процесса компиляции

1. frontend (AST)
2. Межпроцедурная часть
3. Попроцедурная часть
4. backend

frontend+backend – 2-5% от общего времени компиляции задачи (SPEC CPU 2006)

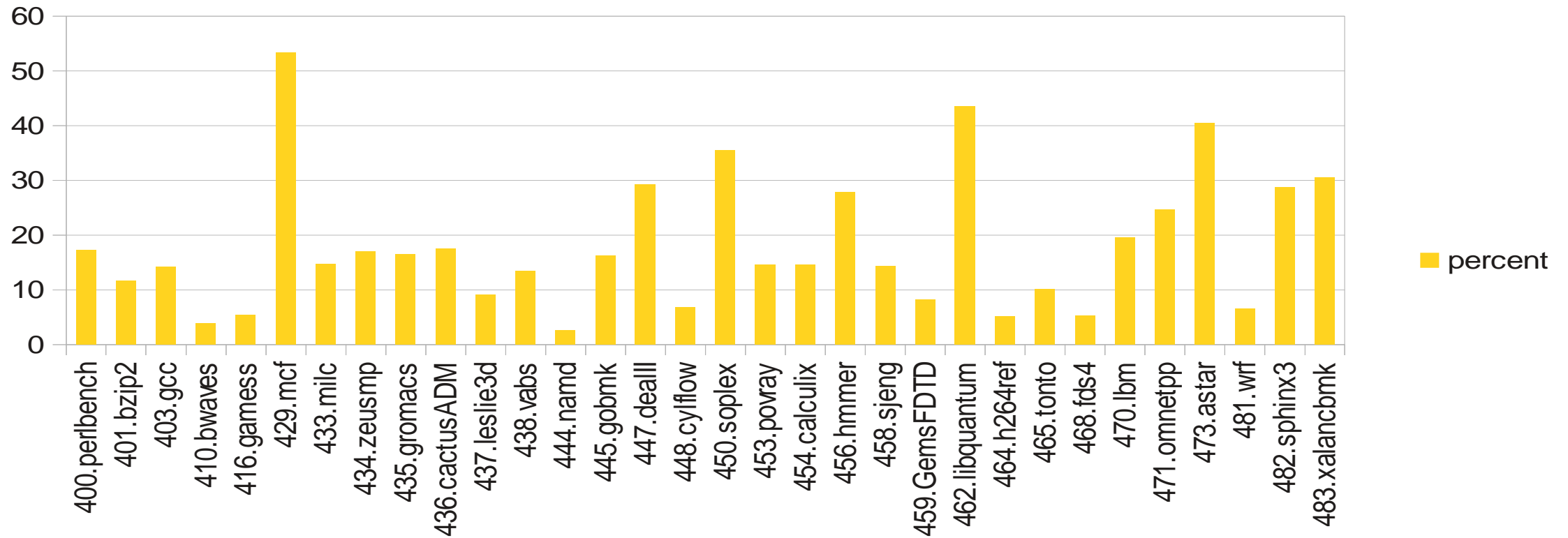
Межпроцедурная часть компиляции не может быть распараллелена в компиляторе lcc, является ограничением на минимальное время компиляции задачи

Процент времени компиляции межпроцедурной части от общего времени компиляции SPEC CPU 2006 в помодульном режиме



Среднее геометрическое = 2.96%

Процент времени компиляции межпроцедурной части от общего времени компиляции SPEC CPU 2006 в режиме «вся программа»



Среднее геометрическое = 14.23%

Алгоритм распределения процедур по потокам

1. Создать список всех процедур задачи, отсортированный по убыванию предсказанного времени компиляции
2. Заполнить потоки процедурами, имеющими самое большое время компиляции
3. По мере освобождения потоков взять следующую по очереди процедуру из списка и передать на свободный поток
4. Повторить п.3 до тех пор, пока не будет скомпилирована вся задача

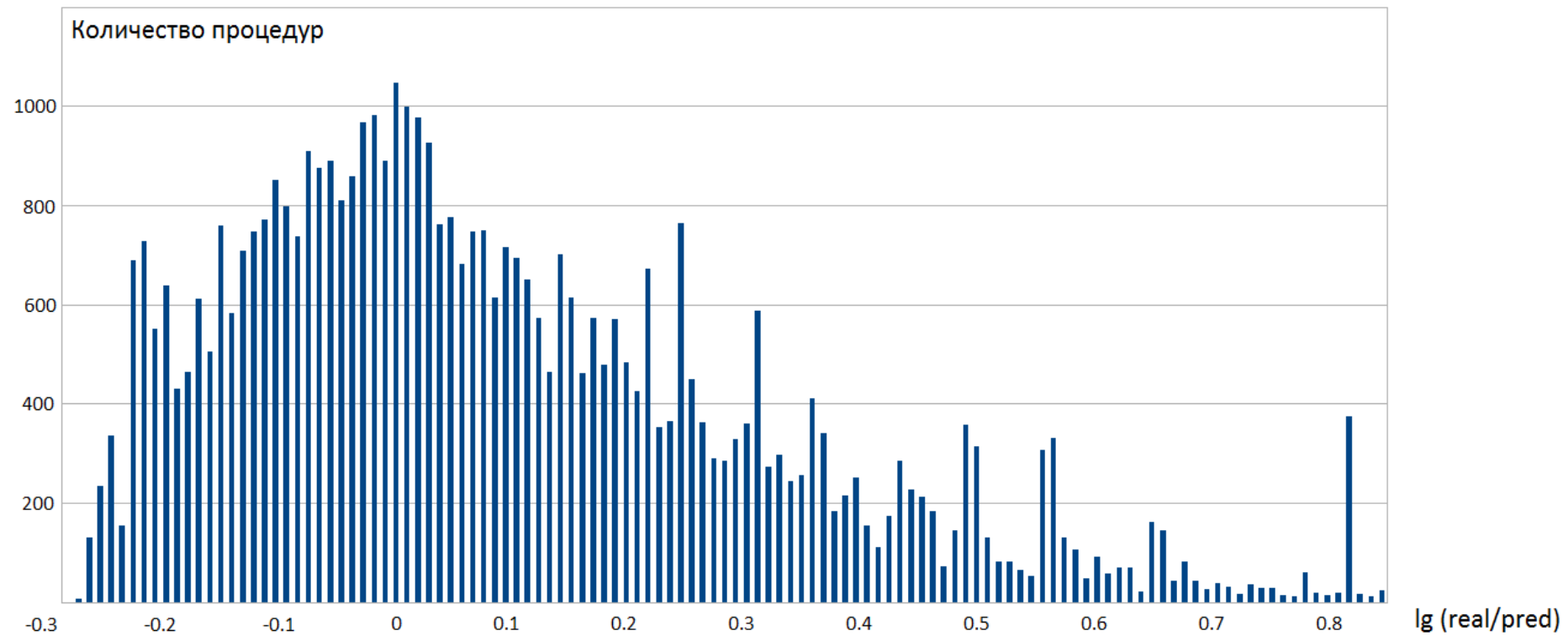
Предсказание времени компиляции

Реальное время компиляции – отношение времени компиляции процедуры к общему времени компиляции задачи.

Предсказанное время компиляции – отношение количества команд в промежуточном представлении процедуры к количеству команд в промежуточном представлении всей задачи.

Предсказанное время не обязано точно совпадать с реальным. Достаточно было бы иметь монотонную зависимость, но даже с нарушениями монотонности возможно получить неплохой результат.

Распределение отношений реального и предсказанного времен



Среднее геометрическое по всем отношениям времен $\text{real/pred} = 0.8796$

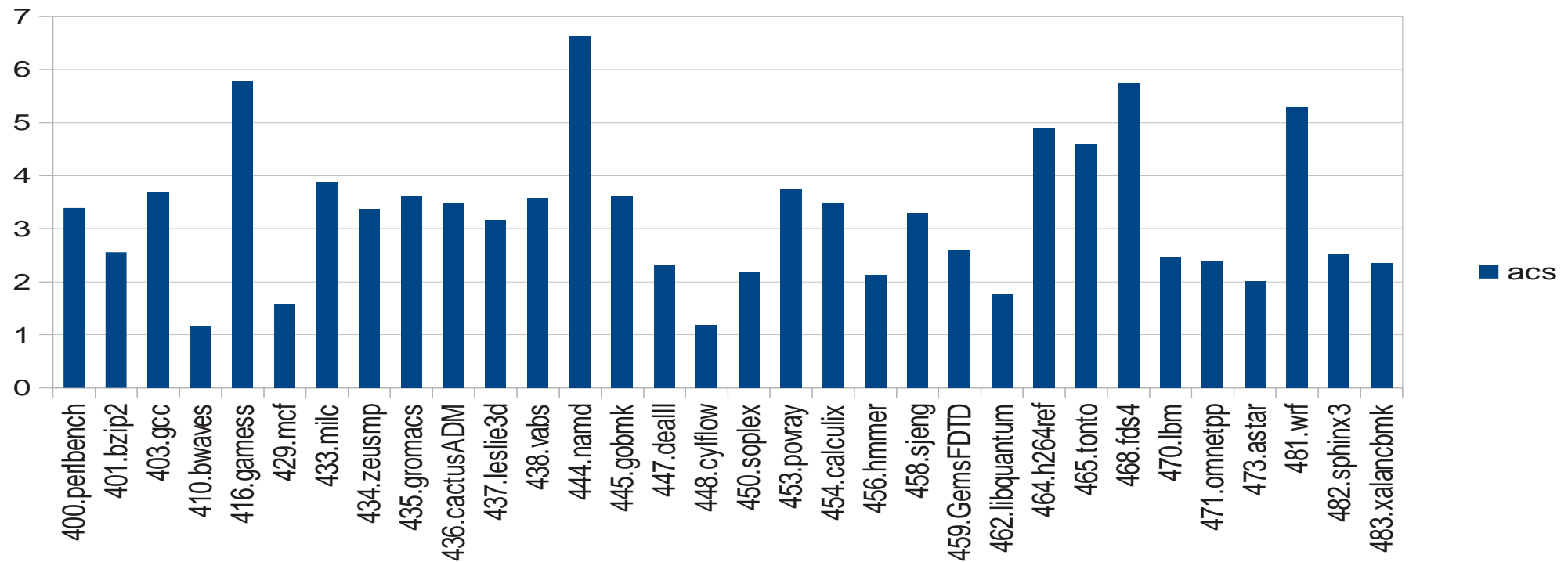
Факторы, ограничивающие эффективное применение метода

- большая по времени компиляции межпроцедурная часть, frontend и backend
- большие по времени компиляции процедуры по сравнению с другими процедурами
- ошибки в оценке времен компиляции процедур

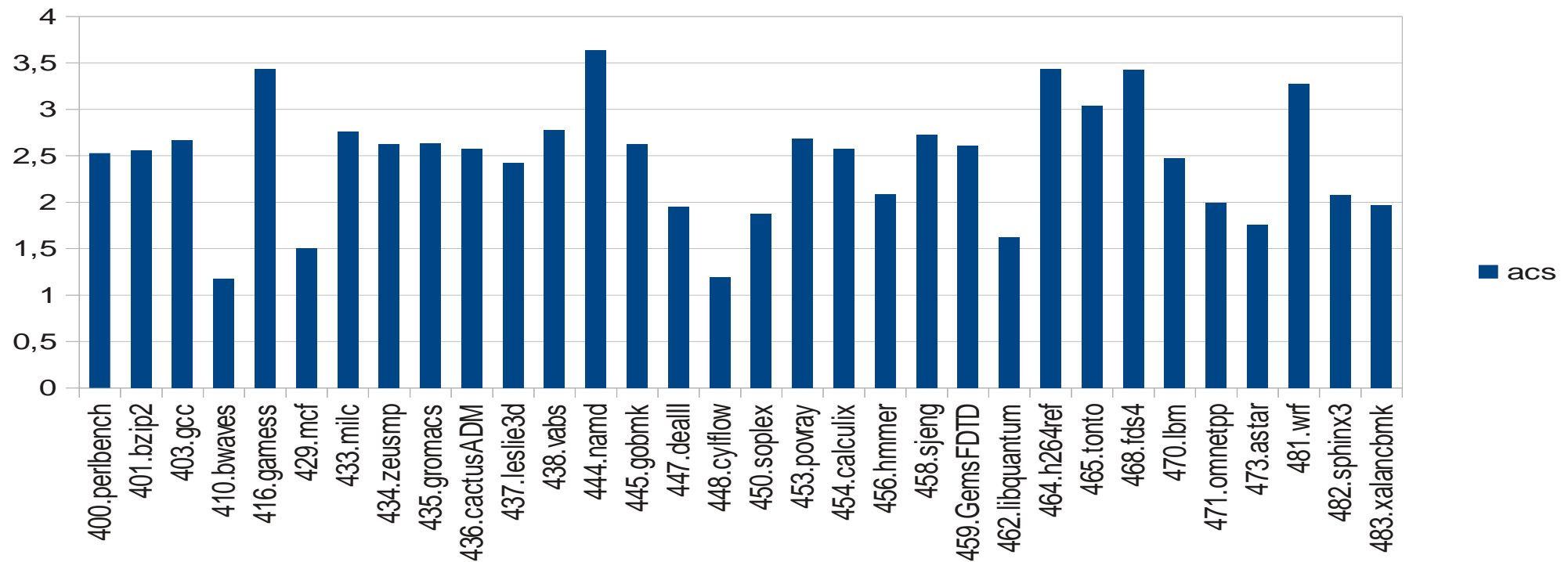
Теоретический расчет моделирования распараллеливания компиляции

- составляется список процедур, отсортированный по убыванию предсказанного времени компиляции
- процедуры распределяются по потокам согласно этому списку
- время работы каждого потока вычисляется как сумма реальных времен компиляции процедур, распределенных в этот поток
- временем компиляции задачи является максимальное из времен работы всех потоков

Ускорения компиляции задач SPEC CPU 2006 при распараллеливании на 8 потоков (среднее геометрическое = 2.9997)



Ускорения компиляции задач SPEC CPU 2006 при распараллеливании на 4 потока (среднее геометрическое = 2.3695)



Анализ результатов моделирования

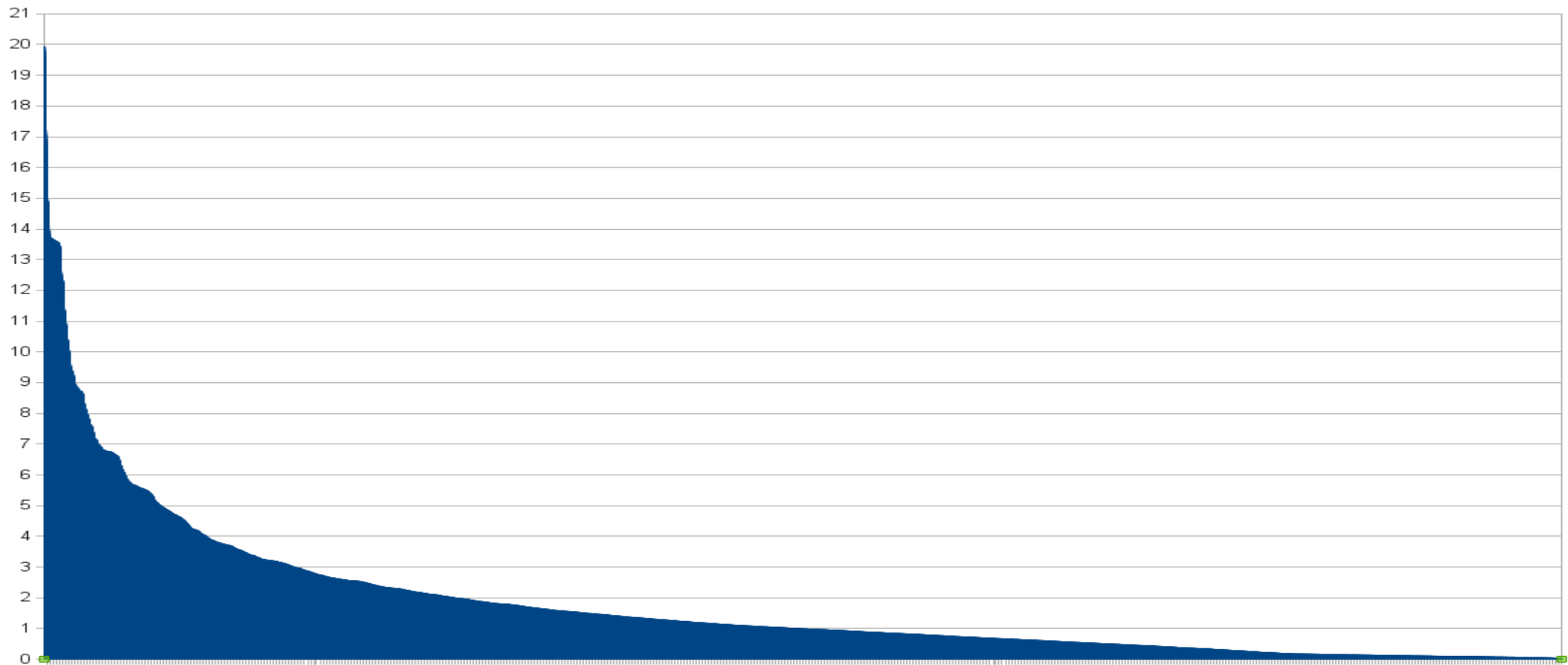
1. Задачи, ускорившиеся меньше всего (при распараллеливании на 8 потоков в 1-2 раза; 410.bwaves, 429.mcf, 448.cylflow), состоят из небольшого количества процедур (10-20), из которых 2-3 занимают 70-80% времени.
2. Основная группа задач (при распараллеливании на 8 потоков ускорившиеся в 2-4 раза).
3. Задачи, ускорившиеся больше всего (при распараллеливании на 8 потоков в 4-7 раз; 416.games, 444.namd, 468.fds4), состоят из большого количества мелких процедур, благодаря этому возможно наиболее эффективное распределение процедур по потокам.

Результаты

1. Выявлена потенциальная эффективность распараллеливания попроцедурной части компиляции
2. Разработан алгоритм распределения компиляции процедур по потокам
3. Вычислено и проанализировано теоретическое ускорение компиляции задач пакета SPEC CPU 2006

Спасибо за внимание!

Распределение реального и предсказанного времени в процедурах задач пакета SPEC CPU 2006. По вертикальной оси отношение реального времени компиляции процедуры к предсказанному, по горизонтальной оси процедуры (~54k; не учтены 80 процедур с отношениями времен от 20 до 215).
AVERAGE = 1.8296, GMEAN = 0.8796



Межпроцедурная часть компиляции

Не параллелится, представляет собой ограничение на минимальное время компиляции задачи

