МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ФАКУЛЬТЕТ
ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ КАФЕДРА
СУПЕРКОМПЬЮТЕРОВ И КВАНТОВОЙ ИНФОРМАТИКИ

СИСТЕМЫ И СРЕДСТВА ПАРАЛЛЕЛЬНОГО ПРОГРАММИРОВАНИЯ

ЗАДАНИЕ 2:
АНАЛИЗ ВЛИЯНИЯ КЭША НА ОПЕРАЦИЮ БЛОЧНОГО МАТРИЧНОГО УМНОЖЕНИЯ

Выполнил:
Алёшин Н.А.

Москва 2020

Постановка задачи и формат данных.

Задача: реализовать последовательный алгоритм блочного матричного умножения и оценить влияние кэша на время выполнения программы. Дополнить отчёт результатами сбора информации с аппаратных счётчиков, используя систему PAPI.

Формат командной строки: <имя файла матрицы A> <имя файла матрицы B> <имя файла матрицы C> <режим, порядок индексов> <размер блока>

Режимы: ijk, ikj, jik, jki, kij, kji.

Формат файла-матрицы: матрицы представляются в виде бинарного файла следующего формата:

Тип	*Значение*	*Описание*
Число типа char	T – f (float)	Тип элементов
Число типа size_t	N – натуральное число	Число строк матрицы
Число типа size_t	M – натуральное число	Число столбцов матрицы
Массив чисел типа T	$N \times M$ элементов (хранятся построчно)	Массив элементов матрицы

Описание алгоритма.

Матрицы делятся на маленькие блоки и происходит блочное перемножение матриц. При этом размер блока подбирается так, чтобы все данные, нужные для вычисления блока матрицы C поместились в кэш.

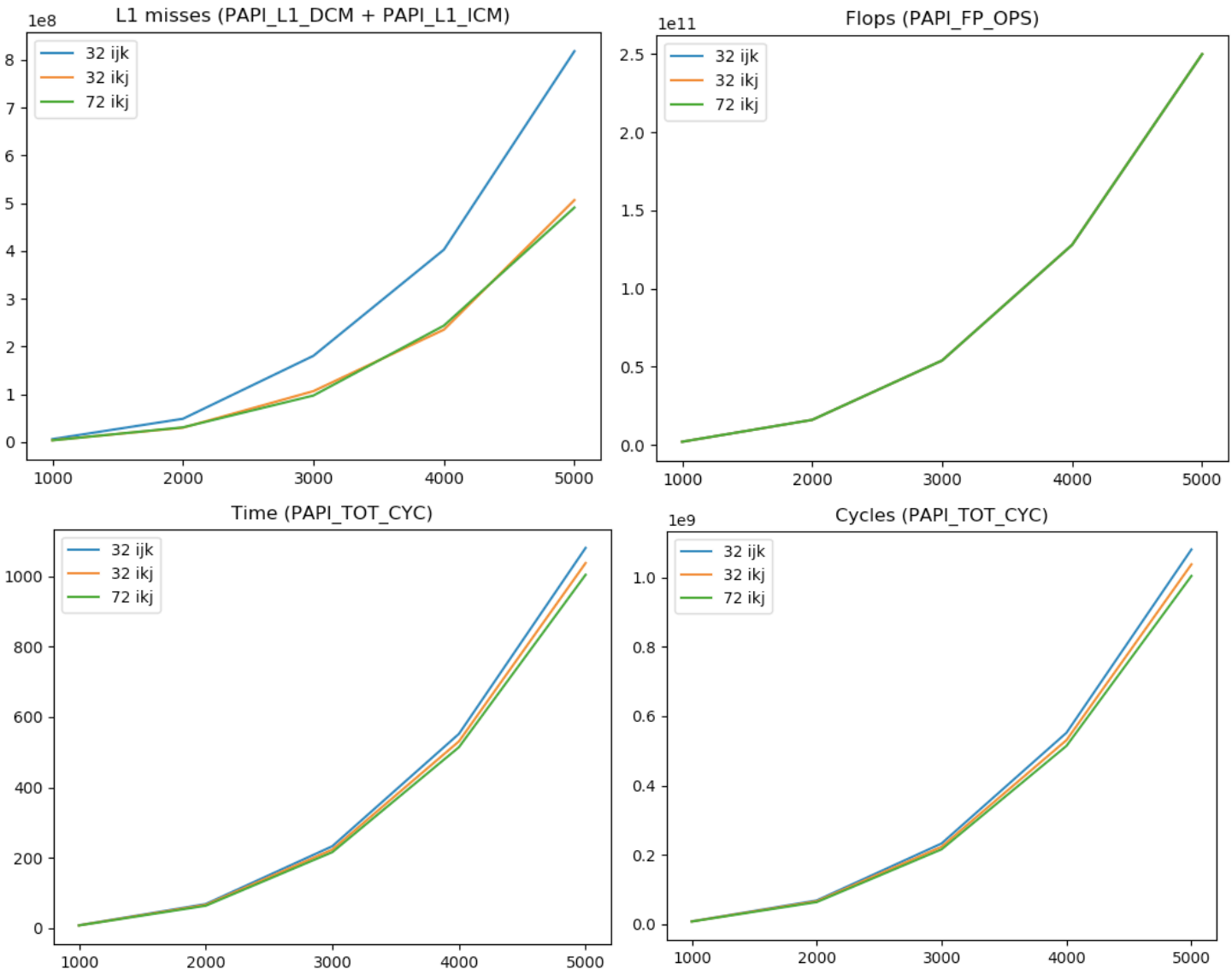
Результат выполнения.

Перемножения выполнялись для квадратных матриц размеров 1000x1000, 2000x2000, 3000x3000, 4000x4000, 5000x5000 типа данных float. Вычисления производились тремя разными способами:

1) с размером блока 32x32 и индексацией ijk;
2) с размером блока 32x32 и индексацией ikj;
3) с индексацией ikj и оптимальным размером блока 72x72 , посчитанным по формуле: $sizeBlock = \sqrt{\frac{cashSize}{3 \times sizeof(float)}} = \sqrt{\frac{65536 \text{ байт}}{3 \times 4 \text{ байт}}} = 72$

Была использована платформа Polus. Подсчет выполнялся с помощью PAPI. Не удалось посчитать промахи L2-кэша (PAPI_L2_DCM + PAPI_L2_ICM) и TLB (PAPI_TLB_TL) из-за недоступности соответствующих операций на POLUS.

Подсчитаны промахи L1-кэша, время выполнения программы, количество процессорных тактов, количество операций. Получены следующие результаты:



Выводы.

Время выполнения программы зависит от попаданий в кэш. Выбор оптимального размера блока приводит к уменьшению времени выполнения программы.