

Теория параллелизма

Задание 8 – CUDA

Цель работы

Целью данной работы является исследование производительности “Задачи о распределении тепла в пластинке”. Сравнение проводится между реализацией на OpenACC и cuBLAS и CUDA с использованием GPU.

Используемый компилятор

Компилятор: nvcc (NVIDIA CUDA Compiler)

Используемый профилировщик

Профилировка производительности выполнялась с использованием замеров времени выполнения на основе стандартной библиотеки C++ (<chrono>).

Также для GPU-реализаций может использоваться pgprof или nsight-systems, однако в данной работе акцент сделан на количественное сравнение по времени выполнения.

Как производили замер времени работы

Замеры времени проводились с использованием высокоточного таймера из стандартной библиотеки C++11:

```
using namespace std::chrono;
auto start = high_resolution_clock::now();
auto end = high_resolution_clock::now();
duration<double> diff = end - start;
cout << "Время выполнения: " << diff.count() << " секунд." << endl;
```

Это позволяет получать точные оценки времени исполнения независимо от реализации.

Выполнение на GPU: Сравнение OpenACC, cuBLAS и CUDA

Для задач малого и среднего масштаба (например, 128×128 и 256×256), OpenACC и cuBLAS показывают сопоставимую производительность. Например:

- 128×128: OpenACC – 1.84 сек, cuBLAS – 3.23 сек, CUDA – 2.20 сек
- 256×256: OpenACC – 16.70 сек, cuBLAS – 15.82 сек, CUDA – 16.32 сек

Однако при увеличении размера задачи до 512×512 время исполнения возросло у всех реализаций:

- 512×512: OpenACC – 199.64 сек, cuBLAS – 201.21 сек, CUDA – 248.86 сек

Важно отметить, что замеры CUDA проводились на загруженном сервере, что могло существенно повлиять на итоговое время выполнения из-за конкуренции за ресурсы GPU. Это объясняет аномально высокое время исполнения CUDA-версии при размере 512×512 по сравнению с OpenACC и cuBLAS.

Качество результатов

Сравнение финальных матриц размером 10×10 показывает, что реализации на OpenACC/cuBLAS и CUDA дают почти идентичные результаты, что говорит о корректности подходов и численной стабильности. Все три реализации показывают почти идентичную численную сходимость (Final norm ≈ 0.0633 при 1000 итерациях).