



Санкт-Петербургский государственный университет
Кафедра системного программирования

Исследование производительности умножения матриц на графических ускорителях от Imagination Technologies

Пучкин Станислав Андреевич, группа 23.Б08-мм

Научный руководитель: С. В. Григорьев, доцент кафедры системного программирования

Санкт-Петербург
2025

- GEMM — ключевая операция линейной алгебры
- Используется в ML, CV, численных методах
- Производительность зависит от GPU
- OpenCL — кроссплатформенный стандарт
- Актуальность для архитектуры RISC-V

Постановка задачи

Целью работы является исследование производительности различных реализаций умножения матриц из библиотек MyGEMM и CLBlast на одноплатных компьютерах с архитектурой RISC-V с графическим ускорителем от Imagination Technologies

Задачи:

- Обновление системных пакетов и программного обеспечения
- Провести серию тестов производительности различных ядер библиотеки MyGEMM
- Выполнить процедуру автоматического тюнинга библиотеки CLBlast и провести измерения производительности.
- Провести сравнительный анализ производительности MyGEMM и CLBlast

- GEMM
- CUDA vs OpenCL
- RISC-V и Imagination Technologies
- MyGEMM и CLBlast

MyGEMM и CLBlast

MyGEMM	CLBlast
Учебная библиотека, демонстрирующая пошаговую оптимизацию	Отпимизированная библиотека для использования в производительных решениях
11 отдельных ядер с нарастающей сложностью оптимизаций	Единое универсальное ядро с адаптивными параметрами
Ручная настройка параметров для каждой платформы	Автоматический поиск оптимальных параметров
Фокус на понимании принципов оптимизации	Фокус на максимальной производительности

Условия эксперимента

- **Платформы:**

- ▶ Banana Pi BPI-F3, GPU IMG BXE-2-32
- ▶ StarFive VisionFive 2, GPU IMG BXE-4-32 MC1
- ▶ Intel Core i9-12900H, GPU Intel Iris Xe Graphics

- **Условия тестирования:**

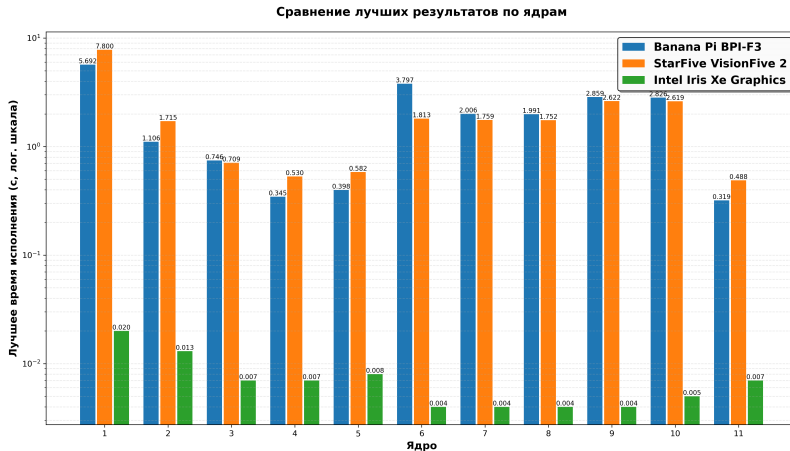
- ▶ Форк репозитория (<https://github.com/stanislawgexrby/myGEMM/tree/develop>) со скриптом `run_all_kernels.sh` для MyGEMM
- ▶ Репозиторий (<https://github.com/vkutuev/matrix-benchmark/tree/main>) со скриптами `benchmark.sh` и `conf.sh` для CLBlast

- **Параметры тестирования:**

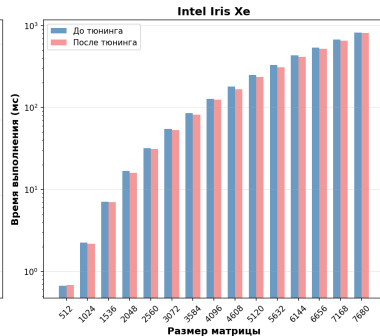
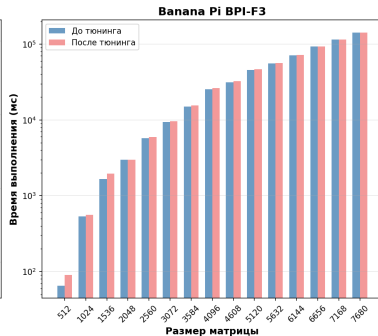
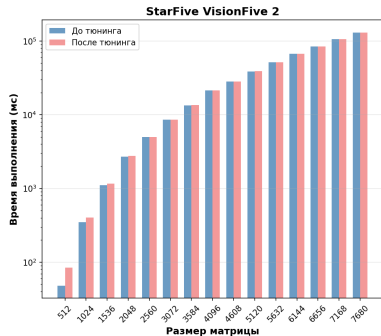
- ▶ Наборы параметров TS, TSM, TSN в диапазоне 8--32, 32--128 для MyGEMM
- ▶ Атоматический подбор параметров и размер матриц из набора 512--7680 для CLBlast

- **Измеряемые метрики:**

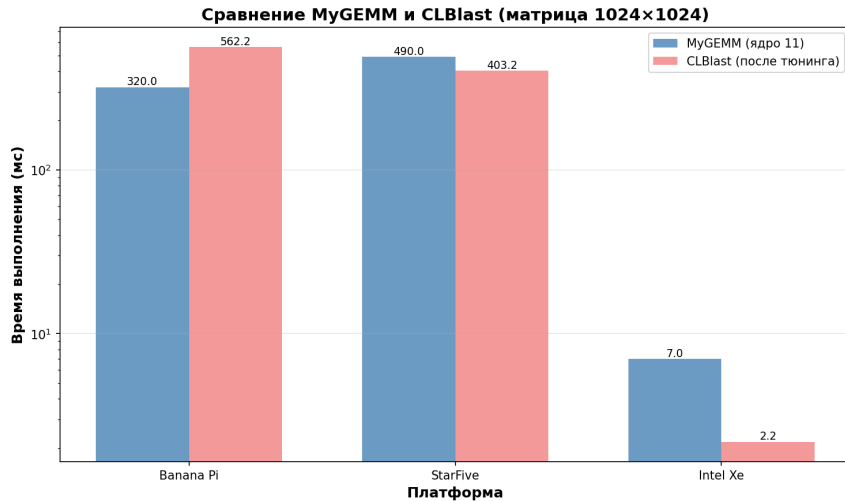
- ▶ Время исполнения каждого ядра без учета задержек загрузки



Результаты CLBlast



Сравнение MyGEMM и CLBlast



- Библиотека MyGEMM адаптирована для работы на платформах Imagination Technologies с учётом аппаратного ограничения размера рабочей группы в 512 потоков
- Проведено тестирование с 100 прогонами для каждой конфигурации на матрицах 1024×1024 элементов для библиотеки MyGEMM, результаты проанализированы
- Для библиотеки CLBlast выполнен автоматический тюнинг с анализом результатов до и после на тестируемых платформах
- Проведен сравнительный анализ этих двух библиотек