

О состоянии проекта RISC-V GPGPU Vortex

Рабочая группа “Развитие экосистемы ПО на RISC-V”

Семён Григорьев

Санкт-Петербургский Государственный Университет

13 ноября 2025

- **Spla**¹: библиотека анализа графов
 - ▶ На основе разреженной линейной алгебры
 - ▶ Использует графические ускорители через OpenCL
- **Vortex**²: открытый проект графического ускорителя
 - ▶ Набор инструкций, основанный на RISC-V ISA
 - ▶ Поддержка OpenCL через POCL³
 - ▶ Конфигурируемая архитектура: количество кластеров, ядер, потоков, наличие/отсутствие кэшей разного уровня и т.д.

¹<https://github.com/SparseLinearAlgebra/spla>

²<https://github.com/vortexgpgpu>

³<https://portablecl.org/>

Как Spla ведёт себя на Vortex?

? Работоспособность

- ▶ Проходят тесты
- ▶ Работают прикладные алгоритмы на основе Spla

? Масштабируемость

- ▶ Зависимость производительности на конкретных прикладных задачах от параметров архитектуры
- ▶ Потребление ресурсов ПЛИС

? Реальная производительность на ПЛИС

Проблемы со сбросом регистров

- Регистры сбрасываются в глобальную память
 - ▶ У «взрослых» ГПУ для этого специальные сегменты в районе локальной памяти
- Типичные оптимизации⁴ не работают^{5,6}
 - ▶ Не просто не улучшают производительность, а заметно её ухудшают
- В целом, есть подозрение, что мало регистров на нитку
 - ▶ Типичное для RISC-V 32 + 32
 - ▶ У «взрослых» ГПУ порядка 256
 - ▶ Ну и в целом с дизайном не всё гладко⁷

⁴ Например, в умножении плотных матриц

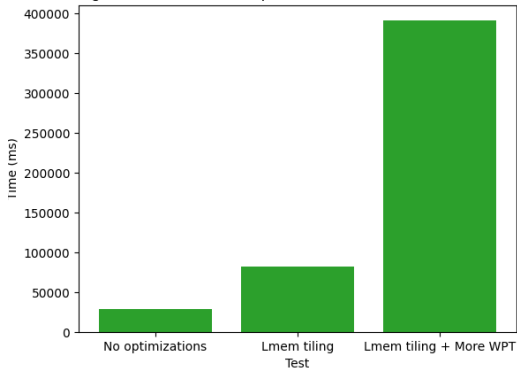
⁵ <https://github.com/vortexgpgpu/vortex/issues/251>

⁶ <https://github.com/vortexgpgpu/vortex/issues/205>

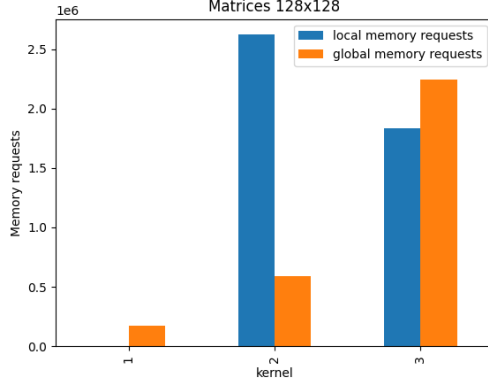
⁷ <https://github.com/vortexgpgpu/vortex/issues/286>

Эксперименты с оптимизациями

Average time of matrix multiplication of FPGA, Matrices 1024x1024



Number of memory requests, RTL simulation, Matrices 128x128



Не самая полная поддержка OpenCL

- Базовые функции драйвера
 - ▶ Не было поддержки работы с несколькими ядрами⁸
 - ▶ Не работает получение информации о скомпилированном ядре (`clGetProgramInfo`)^{9,10}
- Не все функции работы с памятью
 - ▶ Например, не работает `copyBuffer`¹¹
- Плохо поддержаны атомарные операции (хотя заявлена поддержка А-расширения)
 - ✓ Есть поддержка в функциональном симуляторе SimX
 - ⚙ Работаем над поддержкой в компиляторе¹²
 - ✗ Нет поддержки на уровне HDL¹³

⁸Теперь есть, мы поправили: <https://github.com/vortexgpgpu/pocl/pull/6>

⁹<https://github.com/vortexgpgpu/vortex/issues/258>

¹⁰<https://github.com/vortexgpgpu/vortex/issues/287>

¹¹<https://github.com/vortexgpgpu/vortex/issues/283>

¹²<https://github.com/vortexgpgpu/pocl/pull/8>

¹³<https://github.com/vortexgpgpu/vortex/issues/285>

- ✓ Существующие прикладные алгоритмы с использованием Spla запускаются в SimX (BFS, SSSP, PageRank, подсчёт треугольников), тесты проходят
- ✓ Подсчёт треугольников работает на уровне RTL симуляции¹⁴
- ⚙ Запустить подсчёт треугольников на ПЛИС
- ⚙ Подобрать оптимальную конфигурацию для прикладных алгоритмов, оценить масштабируемость
- ⌚ Поддержка атомарных операций
 - ▶ Мы, скорее всего, на стороне компиляторной инфраструктуры
 - ▶ Команда Vortex говорит, что работает над поддержкой на уровне RTL, но пока без конкретных сроков

¹⁴Так как не требует атомарных операций