



Анализ графов и разреженная линейная алгебра в экосистеме RISC-V

Рабочая группа “Развитие экосистемы ПО на RISC-V”

Семён Григорьев

Санкт-Петербургский Государственный Университет

3 октября 2025

Линейная алгебра и анализ графов: области применения

- **Анализ больших графов:** графовые БД, анализ кода, поиск уязвимостей, анализ трафика, анализ транзакции, банковская аналитика, социальные сети. . .
 - ▶ Важна производительность
 - ▶ Разнообразные алгоритмы
- Разреженная обобщённая линейная алгебра — путь к унифицированной параллельной обработке графов
- Примеры заинтересованных сторон
 - ▶ Sber и Montecristo
 - ▶ VK и Tarantool
 - ▶ ИТ-холдинг Т1 и Мирион
 - ▶ Т-Банк
 - ▶ ...

Линейная алгебра и анализ графов: особенности

- **Обобщённая:** матрицы и вектора параметризованы типом элемента, операции над ними могут быть заданы пользователем
 - ▶ Нет фиксированного набора типов и операций
- **Разреженная:** специализированные структуры для хранения матриц и векторов с малым числом значимых элементов, специализированные алгоритмы для их обработки
 - ▶ Понимание разреженности различается от области к области: для AI/ML типично заполнение десятков процентов ячеек, для графов — много меньше 1%
- **Не все решения, удачные для AI/ML и вычислительной математики окажутся удачными для анализа графов¹**

¹RVV, матричные расширения и т.д.

GraphBLAS API⁶

- API для создания алгоритмов анализа графов на основе линейной алгебры
 - ▶ Различные операции над матрицами и векторами (разреженными)
 - ▶ Параметризация алгебраическими структурами: полукольцами, моноидами и т.д.
- Выразимы различные алгоритмы (в LAGraph более 20 различных задач)
 - ▶ Обход в ширину, поиск кратчайших путей, достижимость, ...
 - ▶ Подсчёт треугольников, PageRank, остовные деревья, кластеризация, ...
 - ▶ Запросы с регулярными (RPQ) и контекстно-свободными (CFPQ) ограничениями ...
- Подробнее
 - ▶ The GraphBLAS C API Specification²
 - ▶ GraphBLAS Pointers³
 - ▶ Introduction to GraphBLAS⁴
 - ▶ LAGraph⁵

²https://graphblas.org/docs/GraphBLAS_API_C_v2.1.0.pdf

³<https://graphblas.org/GraphBLAS-Pointers/>

⁴<https://zenodo.org/record/4318870/files/graphblas-introduction.pdf>

⁵<https://github.com/GraphBLAS/LAGraph>

⁶<https://graphblas.org/>

Реализации GraphBLAS-подобных API

- **SuiteSparse:GraphBLAS⁷**: эталон на чистом C
 - ▶ Используется в [NetworkX](#), [FalkorDB](#), [OneSparse](#), [Open3d](#), [Eigen](#), [GNU Octave](#) и т.д.
 - ▶ Ведутся работы по использованию GPU через Cuda
 - ▶ Обёртки для различных языков: Python, Rust, ...
- Huawei's GraphBLAS⁸: частичная реализация на C++
- CombBLAS⁹: распределённая, частичная реализация на C++
- GraphBLAST¹⁰: поддержка GPGPU, Cuda C, частичная реализация
- **Spla¹¹**: поддержка GPGPU, OpenCL C, частичная реализация
- GraphLily¹²: подмножество GraphBLAS на FPGA

⁷<https://github.com/DrTimothyAldenDavis/GraphBLAS>

⁸<https://gitee.com/CSL-ALP/graphblas>

⁹<https://github.com/PASSIONLab/CombBLAS>

¹⁰<https://github.com/gunrock/graphblast>

¹¹<https://github.com/SparseLinearAlgebra/spla>

¹²GraphLily: Accelerating Graph Linear Algebra on HBM-Equipped FPGAs

- Linked Data Benchmark Council (LDBC)
 - ▶ LDBC Graphalytics Benchmark: алгоритмы анализа графов
 - ▶ LDBC Semantic Publishing Benchmark: семантические сети (RDF)
 - ▶ ...
- Suite Sparse Matrix Collection: набор разреженных матриц различной природы
 - ▶ В LAGraph есть своя инфраструктура для экспериментов
- Graph500: алгоритмы анализа графов для суперкомпьютеров
- GAP: алгоритмы анализа графов
- ...

SuiteSparse:GraphBLAS на RISC-V

- SuiteSparse:GraphBLAS (и SuiteSparse целиком) готов к использованию на RISC-V
 - ✓ Кросс-сборка: 12мин. на Intel Core i5-1340P, 4x4.6ГГц, 8x3.4ГГц
 - ✓ Нативно: 1ч.48мин. на Milk-V (8 ядер 1.6 ГГц)
 - ✓ Тесты проходят
 - ✓ Поддержка RISC-V в CI есть¹³
- Есть простор для оптимизаций
 - ▶ Ручная векторизация¹⁴
 - ▶ Изучение возможностей автоматической векторизации¹⁵
 - ▶ Применимость матричных расширений¹⁶

¹³Соответствующий реквест: <https://github.com/DrTimothyAldenDavis/SuiteSparse/pull/949>

¹⁴Есть поддержка AVX и Neon, а с RVV1.0 только эксперименты:

<https://github.com/DrTimothyAldenDavis/GraphBLAS/pull/381>

¹⁵В проекте много кода, написанного с расчётом на автоматическую векторизацию

¹⁶На текущий момент они нацелены на AI/ML, а это другое

Результаты векторизации умножения матриц в SuiteSparse:GraphBLAS¹⁷

№	Matrix name	Rows number	Nonzeros	AVX2 (ms.)	No AVX2 (ms.)	RVV (ms.)	No RVV (ms.)	AVX speedup (%)	RVV speedup (%)
1	olafu	16146	515651	5327.7	6629.7	43080.7	52940.1	19.6	18.6
2	fd18	16428	63406	476.4	482.0	2212.6	2181.2	1.2	-1.4
3	sme3Da	12504	874887	4236.9	5124.9	32008.0	42763.8	17.3	25.2
4	stokes64	12546	74242	508.8	564.4	2629.7	2814.1	9.8	6.6
5	sinc12	7500	294986	632.6	864.0	5970.1	8593.8	26.8	30.5
6	fd12	7500	28462	90.4	92.3	484.1	555.3	2.0	12.8
7	bcsstk15	3948	60882	87.8	117.9	1271.5	1770.8	25.6	28.2
8	tols4000	4000	8784	17.1	18.2	184.0	203.5	5.9	9.6
9	ex36	3079	53843	28.5	41.0	574.2	584.8	30.5	1.8
10	iprob	3001	9000	25.2	34.7	279.3	344.9	27.5	19.0
11	MISKnowledgeMap	2427	28511	31.3	38.5	401.6	490.0	18.8	18.0
12	LeGresley_2508	2508	16727	10.4	12.1	106.5	97.7	14.3	-8.9
13	reorientation_2	1544	9408	5.6	9.8	117.9	125.4	42.7	6.0
14	netscience	1589	2742	1.5	2.8	31.4	28.5	47.0	-10.0
15	mcfe	765	24382	2.3	5.5	51.1	65.3	58.8	21.8
16	orbitRaising_3	761	3256	0.6	1.6	10.5	13.1	63.0	19.5

¹⁷SoC: SPACEMIT K1/M1, Octa-core X60™(RV64GCVB), 1.6GHz, RVA22, RVV1.0

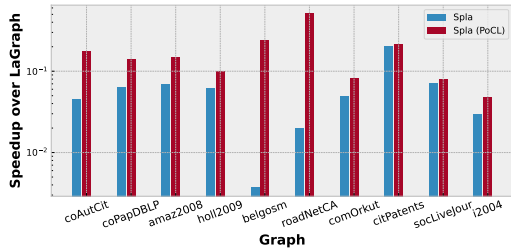
- **RISC-V CPU + Imagination Technologies GPU:** самая распространённая (из доступных) конфигурация
- **RISC-V CPU + AMD GPU:** утверждается, что работает и есть официальная поддержка ([Milk-V Pioneer](#), [Milk-V Megrez](#), [RuyiBook](#))
- **RISC-V CPU + Intel GPU:** [ходят слухи, что можно](#), но официальной поддержки пока нет
- **RISC-V CPU + Nvidia GPU:** [анонсировано](#)
- **RISC-V GPU:** [Vortex](#) (только на ПЛИС)

- ✓ **RISC-V CPU + Imagination Technologies GPU:** собирается, запускается, тесты проходят
- ⚙ **RISC-V GPU Vortex:** собирается, запускается, не все тесты проходят
- ⌚ **RISC-V CPU + AMD GPU, Intel GPU, Nvidia GPU:** требуется проверка
 - ▶ Серьёзных проблем не ожидается, так как на аналогичных конфигурациях с x86 CPU всё нормально¹⁸
- Есть простор для оптимизаций
 - ▶ Performance Portability Problem никто не отменял
 - ▶ Эксперименты с различными конфигурациями Vortex
- Vortex пока относительно “сырой”
 - ▶ Проблемы со сбросом регистров
 - ▶ Проблемы с работой с памятью (не все функции для clBuffer работают)
 - ▶ Проблемы с “сервисными” функциями (типа getProgramInfo)
 - ▶ ...

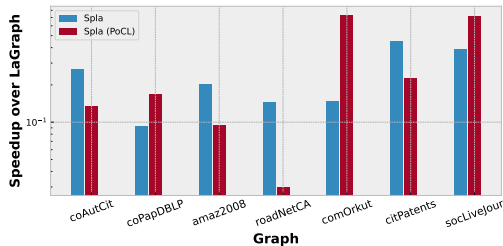
¹⁸Разве что на уровне драйвера или другого системного ПО, не относящегося к библиотеке

Spla на SpacemiT M1 CPU, IMG BXE-2-32 GPU

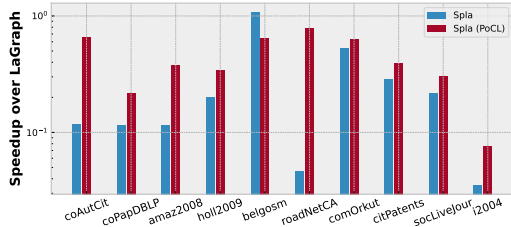
BFS



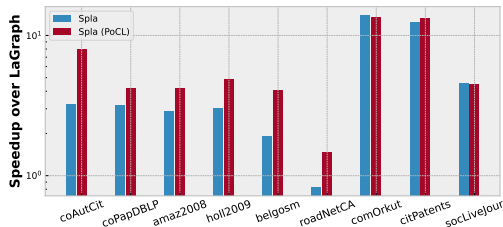
Triangle Count (TC)



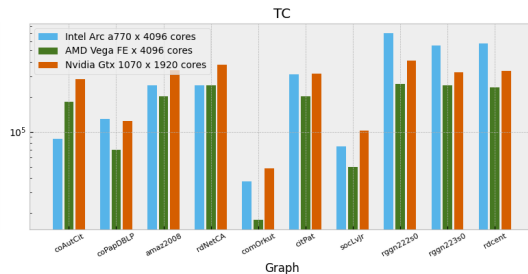
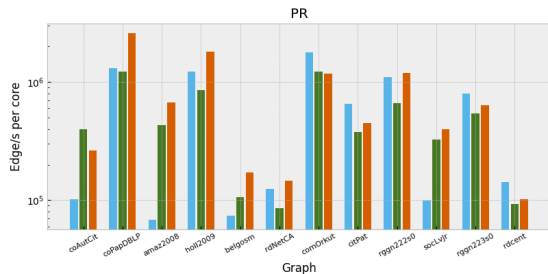
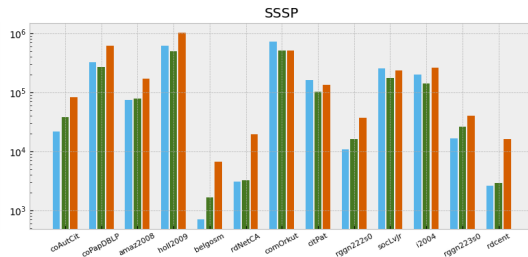
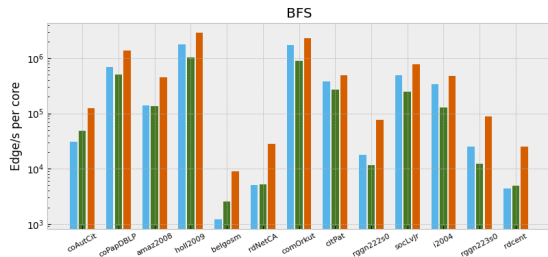
Single Source Shortest Path (SSSP)



PageRank (PR)



Срѣна на ГПУ од Nvidia, Intel и AMD

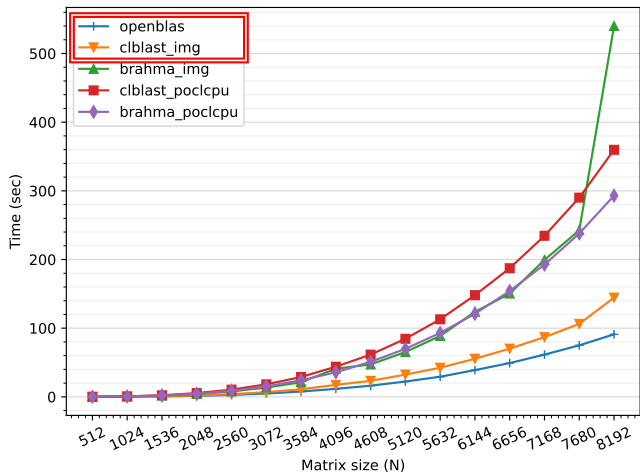


- Есть решения для высокопроизводительного анализа графов с использованием линейной алгебры, пригодные для использования в экосистеме RISC-V
 - ✓ CPU — SuiteSparse:GraphBLAS
 - ★ Стабильно работает, можно брать и пользоваться
 - ★ Требуются оптимизации¹⁹
 - ⚙️ GPGPU — Spla
 - ★ Много работает, можно пробовать, ставить эксперименты
 - ★ Требуемая инфраструктура “сырая”: не до конца отлажены драйвера, “узкие” PCI-шины, слабые ГПУ в доступных конфигурациях ...
- Нужны инструменты для профилирования
 - ▶ В том числе, аппаратная поддержка (счётчики, ...)

¹⁹ Впрочем, как и не на RISC-V

- Проверка связок RISC-V CPU + AMD GPU и RISC-V CPU + Intel GPU
 - ▶ Замеры Spla на них
- Запуск работоспособных бенчмарков Spla на Vortex
- Эксперименты на более разнообразном “железе”
- Анализ векторизации в SuiteSparse:GraphBLAS: что имеет смысл векторизовывать, какими средствами это лучше делать
- Исследование применимости матричных расширений
- Исследование производительности графовых СУБД на RISC-V

Результаты умножения плотных матриц на SparceiT M1 с IMG GPU



(Некоторые) GPGPU от Imagination Technologies (пока) не совсем для вычислений