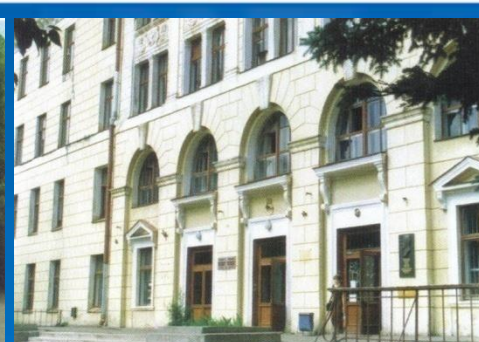
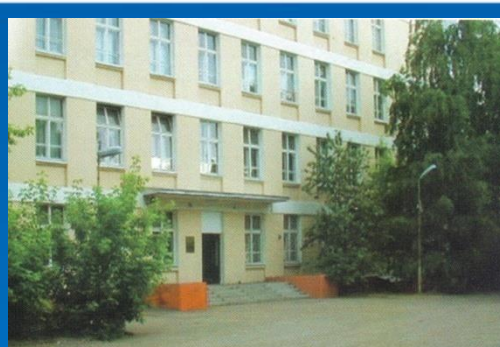
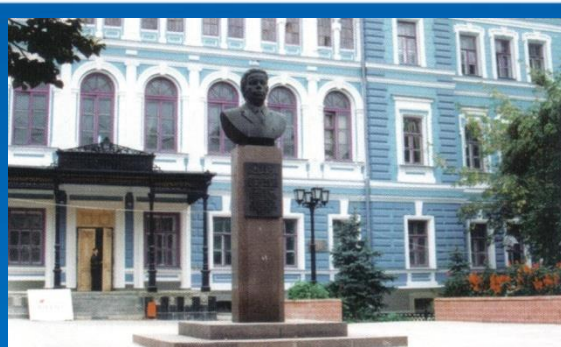




НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. Н.И. ЛОБАЧЕВСКОГО
ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МАТЕМАТИКИ И МЕХАНИКИ





Нижегородский государственный университет им. Н.И. Лобачевского
Институт информационных технологий, математики и механики

Об опыте разработки высокопроизводительных математических библиотек для процессоров RISC-V

Суперкомпьютерные дни
в России 2025
Москва, 30 сентября 2025

Валентин Волокитин
ст. преподаватель каф. ВВиСП
ИИТММ ННГУ им. Н.И. Лобачевского

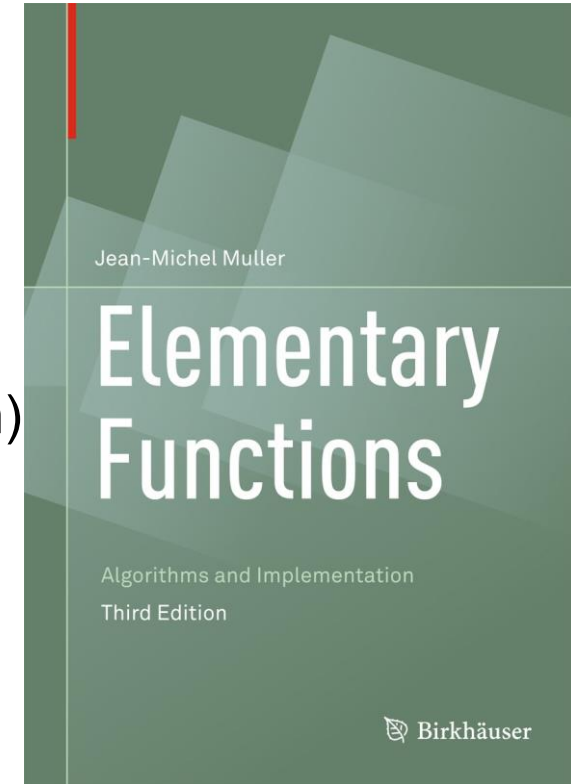
Математические функции

Математические функции

- ▶ Обычные операции (max, min, abs, fma)
- ▶ Экспоненциальные функции (exp, exp2, expm1, log, log2, log10, log1p)
- ▶ Степенные функции (pow, sqrt, cbrt, hypot)
- ▶ Тригонометрия (sin, cos, tan, asin, acos, atan, atan2)
- ▶ Гиперболические (sinh, cosh, tanh, asinh, acosh, atanh)
- ▶ Функции ошибок (erf, erfc)
- ▶ Гамма-функции (tgamma, lgamma)
- ▶ Округления (ceil, floor, trunc, round, rint)
- ▶ И другие (frexp, modf, copysign, ...)

Стандарты

- ▶ Дополнительные разделы IEEE-754, требуется корректное округление во всех точках функции
- ▶ Основной стандарт ISO 10967-2, для каждой функции прописаны ограничения допустимых ошибок



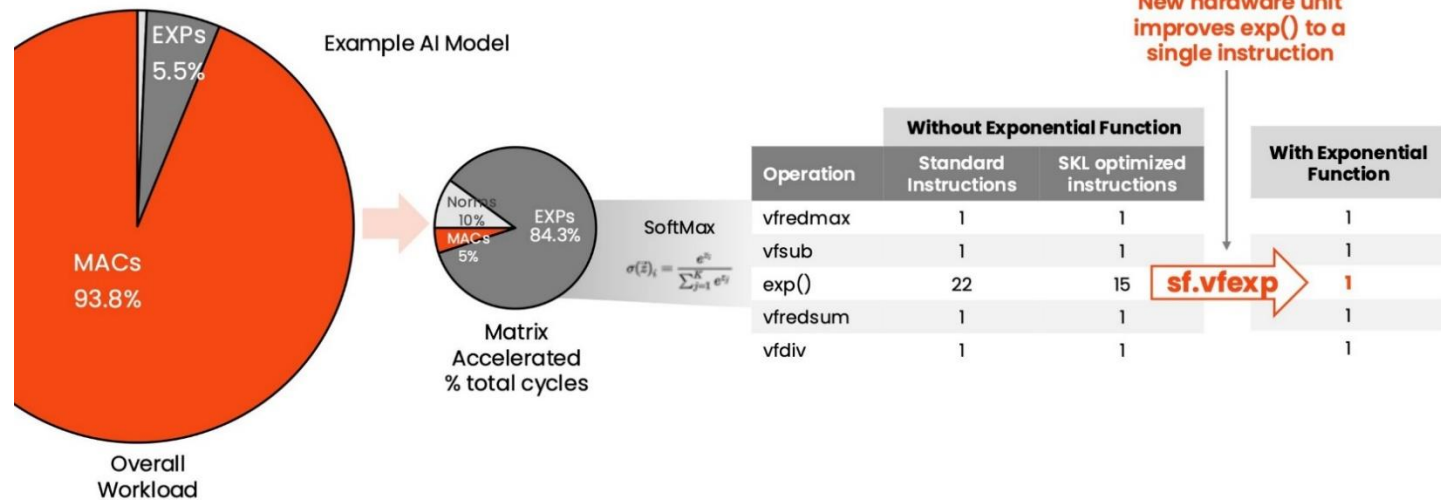
Математические функции

Векторизация математических функций

- ▶ Ускорение вычислений:
 - ▶ При работе компилятора с циклами
 - ▶ При обработке данных с использованием математических функций
- ▶ Основные требования: скорость и точность
- ▶ Один из последних примеров:

New: Exponential Function

Optional hardware pipelined exponential unit



Аппаратно-программное обеспечение

Аппаратное обеспечение:

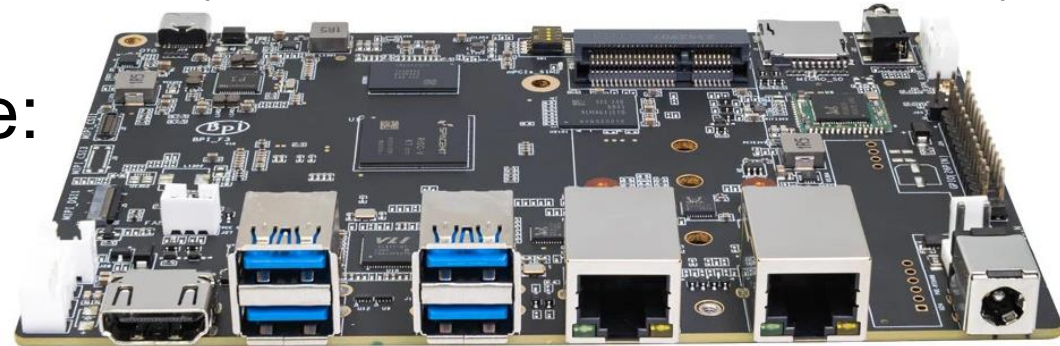
- ▶ Banana Pi BPI-F3:
 - ▶ SpacemiT K1 8x X60(RV64GCVB, RVA22, RVV1.0)
 - ▶ RVV 1.0 (fp64, fp32, fp16), VLEN = 256 (FPU предположительно 128бит)
 - ▶ 4-16ГБ LPDDR4

Программное обеспечение:

- ▶ GCC 14.2 – кросс-компиляция
- ▶ OS: Bianbu 2.1

Детали экспериментов:

- ▶ В тестах производительности производился запуск на массиве в LLC
- ▶ Время измеряется на 1 элемент вектора
- ▶ ULPs – учитывает статистику ошибок 1 *ulp* на N случайных FP:
 - ▶ Наша цель $0.501 \text{ ulp} - 0.5 \text{ ulp}$, кроме быть может 1 числа из 1000 случайных FP



Обзор библиотек. LibM

LibM:

- ▶ LibM – стандартная библиотека, которая поставляется вместе с компилятором

- ▶ Преимущества:

- ▶ Общепринятая и кроссплатформенная
- ▶ В компиляторе может быть специализированная векторная версия (для RISC-V пока нет)
- ▶ Для некоторых версий известны максимальные ошибки (Glibc LibM: https://sourceware.org/glibc/manual/2.41/html_node/Errors-in-Math-Functions.html)
- ▶ Есть исследования по точности: Gladman B. et al. Accuracy of mathematical functions in single, double, double extended, and quadruple precision. – 2025.

- ▶ Недостатки:

- ▶ Алгоритмы почти всегда слабо векторизуемые
- ▶ Часто отсутствует половинная точность

RISC-V 64-bit

expf	1
exp	1
expl	1
expf128	-
exp10f	-
exp10	2
exp10l	2
exp10f128	-

Обзор библиотек. CRLibm

CRLibm:

- ▶ CRLibm – библиотека, обеспечивающая корректное округление во всех точках функции
- ▶ Преимущества:
 - ▶ Гарантия точности
 - ▶ Работает автоматическая векторизация
 - ▶ Есть векторная реализация (x86 - SSE)
- ▶ Недостатки:
 - ▶ Из-за проверок точности и алгоритмов уточнения работает в 2-10 раз дольше высокопроизводительных библиотек
 - ▶ Проект заброшен в исходном виде и продолжается как множество разрозненных библиотек

Обзор библиотек. Intel SVML (VML)

Intel SVML и VML:

- ▶ Intel SVML – библиотека с векторными математическими функциями для Intel x64
- ▶ Преимущества:
 - ▶ Высокопроизводительная
 - ▶ Содержит множество режимов точности
- ▶ Недостатки:
 - ▶ Не получится запустить на RISC-V Vector Extension

Обзор библиотек. `veclibm`

`veclibm`:

- ▶ `veclibm` – библиотека, написанная для RISC-V Vector Extension
- ▶ Преимущества:
 - ▶ Высокопроизводительная
 - ▶ Реализована на RVV Intrinsics
 - ▶ Достаточно точная (заявляется $< 1\ ulp$, хотя мы находили функции с $2\ ulp$)
- ▶ Недостатки:
 - ▶ Реализовано только для *double*
 - ▶ Библиотека использует интерфейс с вызовом на массивах с оптимально подобранным LMUL
 - ▶ Судя по документации, тестируется на QEMU, а не на реальном железе

Обзор библиотек. Sleef

Sleef:

- ▶ Sleef – кроссплатформенная векторная библиотека математических функций
- ▶ Преимущества:
 - ▶ Универсальная библиотека, написанная на псевдоинтринсиках
- ▶ Недостатки:
 - ▶ Реализовано только для LMUL=1,2
 - ▶ Часто наблюдаются проблемы в точности математических функций по сравнению с заявленными
 - ▶ Обязательно требует `-ffp-contract=off` (отключение сокращения выражений), что может снизить производительность

Библиотека RVVMF*. Точность

ULPs	sin			cos			tanh		
	FP64	FP32	FP16	FP64	FP32	FP16	FP64	FP32	FP16
rvvmf	0,5009	0,5005	0,5008	0,5009	0,5006	0,5008	0,5002	0,5003	0,5007
sleef	0,5088	0,5261	-	0,5096	0,5302	-	2,0000	2,0000	-
veclibm	0,5092	-	-	0,5090	-	-	0,5002	-	-
glibc	0,5007	-	-	0,5007	-	-	2,0000	-	-

ULPs	exp			exp2			expm1		
	FP64	FP32	FP16	FP64	FP32	FP16	FP64	FP32	FP16
rvvmf	0,5000	0,5000	0,5002	0,5000	0,5000	0,5002	0,5005	0,5001	0,5009
sleef	0,5008	0,5076	-	0,5009	0,5051	-	0,5002	0,5021	-
veclibm	0,5002	-	-	2,0000	-	-	2,0000	-	-
glibc	0,5000	-	-	0,5000	-	-	0,5006	-	-

*Проект библиотеки RVVMF поддержан ООО «КНС ГРУПП»

Библиотека RVVMF. Точность

ULPs	log			log2			log10		
	FP64	FP32	FP16	FP64	FP32	FP16	FP64	FP32	FP16
rvvmf	0,5000	0,5000	0,5004	0,5000	0,5000	0,5002	0,5000	0,5000	0,5002
sleef	0,5002	0,5004	-	0,5001	0,5003	-	0,5001	0,5003	-
veclibm	0,5000	-	-	0,5000	-	-	0,5000	-	-
glibc	0,5000	-	-	0,5000	-	-	2,0000	-	-

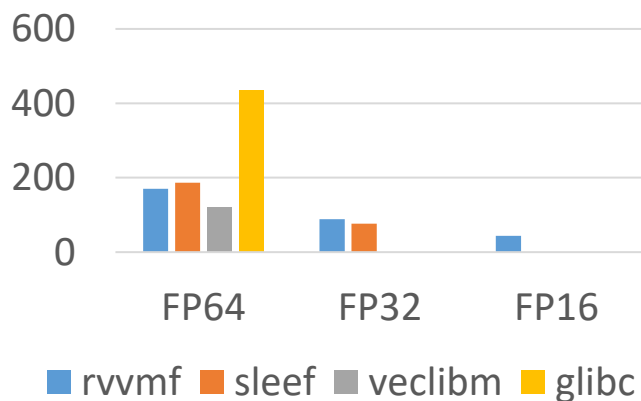
ULPs	log1p			sqrt		
	FP64	FP32	FP16	FP64	FP32	FP16
rvvmf	0,5007	0,5005	0,5006	0,5	0,5	0,593
sleef	0,5006	3,0000	-	0,5	0,5	-
veclibm	0,5001	-	-	-	-	-
glibc	0,5005	-	-	0,5	-	-

Другие функции в RVVMF:

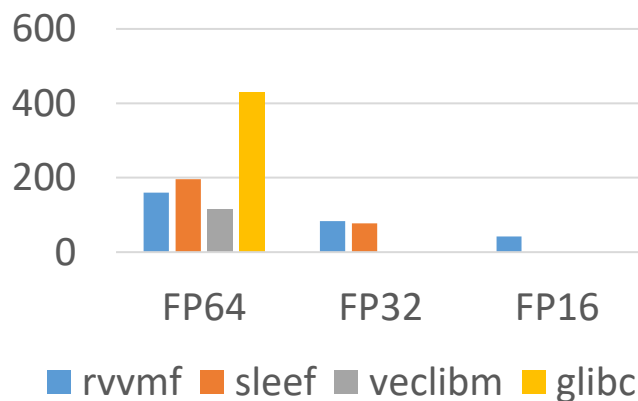
- ▶ hypot – 1 ULPs
- ▶ floor, ceil, trunc, round, rint, fabs - CR

Библиотека RVVMF. Производительность

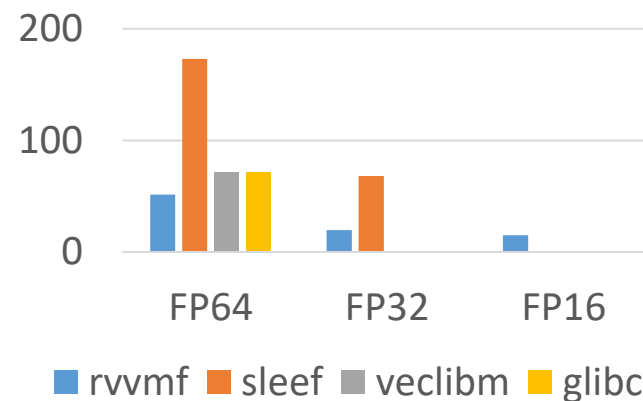
sin



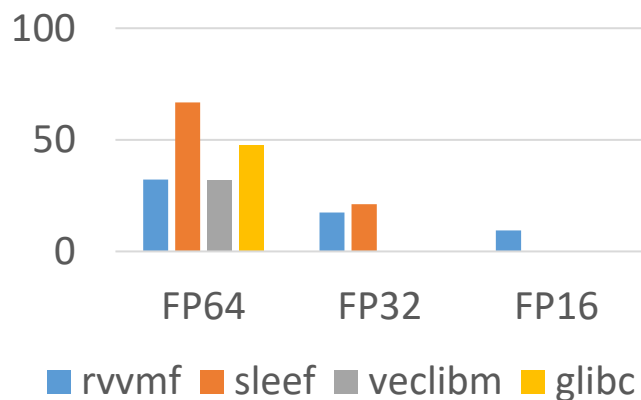
cos



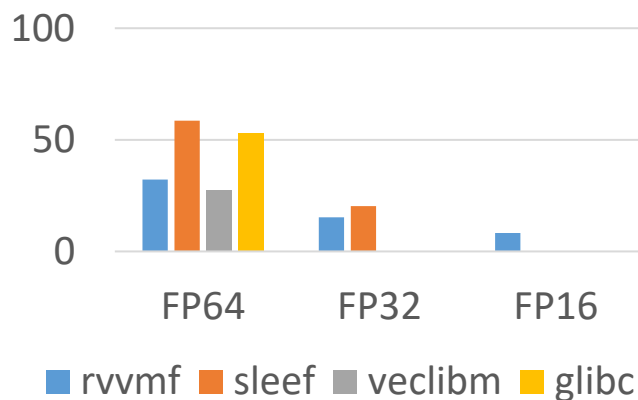
tanh



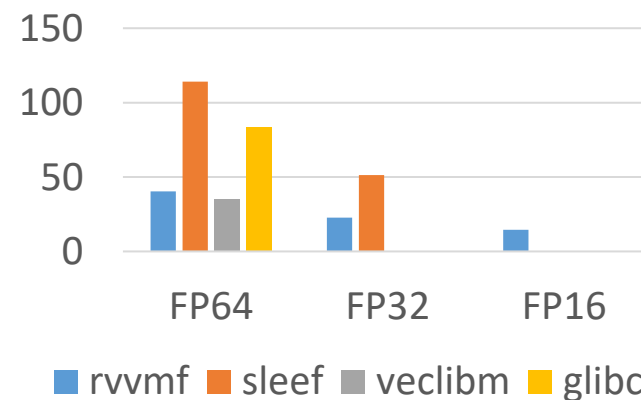
exp



exp2

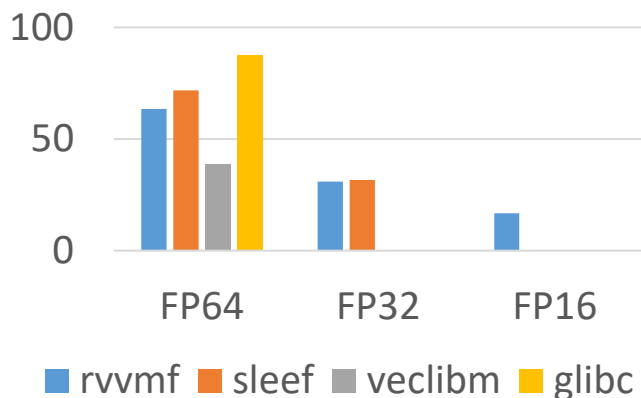


expm1

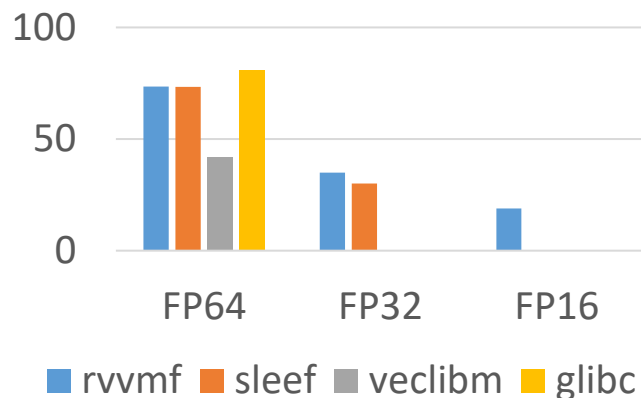


Библиотека RVVMF. Производительность

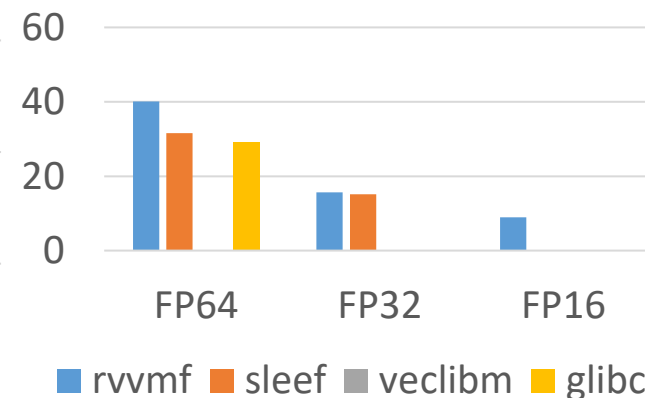
log



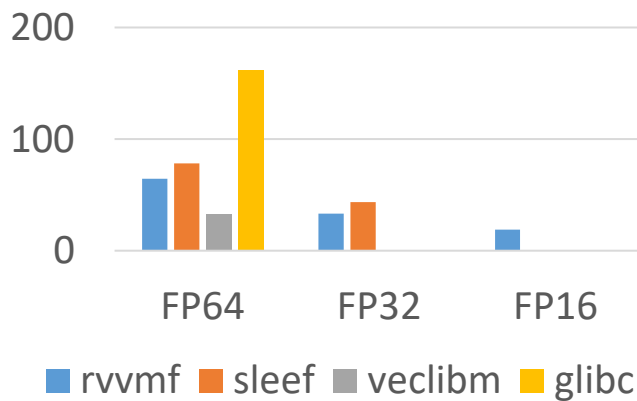
log2



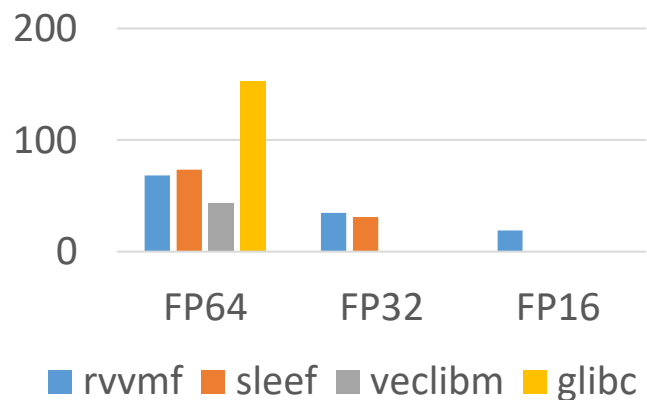
sqrt



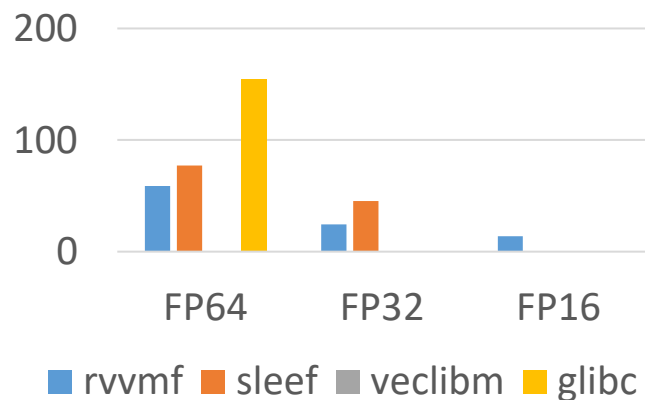
log1p



log10



hypot



Контакты и благодарности

Нижегородский государственный университет

<http://www.unn.ru>

Институт информационных технологий, математики и механики

<http://www.itmm.unn.ru>

Волокитин В.Д.

volokitin@itmm.unn.ru

Авторы благодарят за поддержку программу Приоритет-2030