

Об эффективном выборе
аппаратных средств для
НРС-расчетов вплоть до
суперкомпьютерного
уровня.

Кузьминский М.Б.,
Институт органической химии РАН

Лидеры списка суперЭВМ (TOP500 – июнь 2024)

СуперЭВМ	Число GPU в узлах (Модель GPU)
1. Frontier	4 x (AMD MI250X)
2. Aurora	6 x (Intel GPU Max 1550)
3. Eagle	8 x (Nvidia H100)
4. Fugaku	Нет
5. LUMI	4 x (AMD MI250X)

TOP500 по тесту HPL-MxP

СуперЭВМ	Число ядер	Достигнуто, EFLOPS
1. Aurora	8,159,232	10.6
2. Frontier	8,699,904	10.2
- Sunway	41,140,224	5.048
3. LUMI	2,752,704	2.35
4. Fugaku	7,630,848	2.000
5. Leonardo	1,824,768	1.842

Данные из: <https://www.top500.org/lists/top500/2024/06/highs/>; Lin, Rongfen, et al. "5 ExaFlop/s HPL-MxP Benchmark with Linear Scalability on the 40-Million-Core Sunway Supercomputer." Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. 2023.

Важные особенности современных НРС

- более быстрое развитие аппаратуры, чем программных средств
- быстрый рост числа процессорных ядер
- применение GPU
- стремление к энергоэффективности
- применение виртуализации и облачных технологий для экономически эффективного проведения НРС-расчетов
- проблемы пропускной способности памяти
- актуальность использования смешанной точности

Но тенденции диктует ИИ, и идет «слияние» НРС и ИИ

Серверные процессоры x86-64 (старшие модели)

	Intel Xeon			AMD EPYC	
Кодовое имя	Sapphire Rapids-SP		Emerald Rapids-SP	Bergamo	Genoa
Семейство	4th generation scalable	Xeon Max HBM	5th generation scalable	Zen 4	Zen 4
Номера моделей	Xeon 84xx	Xeon 94xx	Xeon 85xx	EPYC 9xx4	
Старшая модель	Xeon 8490H	Xeon Max 9480	Xeon 8592+	EPYC 9754	EPYC 9654
Число ядер	60	56	64	128	96
Частота, ГГц	1.9-3.5	1.9-3.5	2.2-3.9	2.25-3.1	2.4-3.7
FLOPS/такт на ядро	32	32	32	24	24
GFLOPS	3648	3405	4506	6912	5530
Цена	\$17000	\$12980	\$11600	\$11900	\$11805
TDP, Вт	350	350	350	360	360
Технол., нм	10	10	10	5	5
Кэш L3, МБ	112.5	112.5	320	256	384
В иерархии памяти есть		HBM2e 64ГБ			Кэш L3: 1152 МБ ¹

¹ В 96-ядерной модели EPYC 9684X/2.25-3.7 ГГц (\$14756)

Производительность и цены старших моделей 4-го и 5-го поколений серверных процессоров x86-64 от Intel и AMD.

Модель	n _c	Час- тота, ГГц	Цена	Чис- ло CPUs	SPEC CPU2017	
					fp_speed (base)	fp_rate (base)
EPYC 9754	128	2.25- 3.1		2	430	1460
			\$11900	1	316	733
EPYC 9654	96	2.4- 3.7		2	449	1480
			\$11805	1	324	746
EPYC 9684X	96	2.55- 3.7		2	476	1630
			\$14756	1	357	820
Xeon 8490H	60	1.9- 3.5		2	378	1040
			\$17000	1	255	508
Xeon 8480+	56	2-3.8		2	371	1020
			\$10710	1	242	465
Xeon 8592+	64	1.9- 3.9		2	428	1260
			\$11600	1	286	619
Xeon 8580	60	2-4		2	419	1160
			\$10710	1	286	570

Данные от 28.08.2024 . n_c — число ядер. Приведены максимально достигнутые показатели из <https://www.spec.org/cpu2017/results/>

Данные о **производительности серверов** со старшими моделями **AMD EPYC** и **Intel Xeon** в тестах **SPECint 2011**

Модель	Число ЦП	Tiny base; peak	Small base; peak
EPYC 9654	1	6.99; 6.99	0.735; 0.735
	2	13.9; 14.2	1.45; 1.45
EPYC 9754	1	7.32;	0.823;
	2	16.4;	1.59;
Xeon 8480+	2	7.98; 8.35	0.945; 0.949
Xeon 8490H	2	9.00;	1.00;
	4	17.2; 17.6	1.88; 1.89
Xeon 8592+	1	4.88;	0.553
	2	10.8;	1.15

Приведены максимальные достигнутые на 27.08.2024 показатели из <https://www.spec.org/hpc2021/results/>

Разные числа получены на разных серверах разными производителями.

Жирным шрифтом отмечены самые сопоставимые показатели.

Главные показатели современных GPU AMD и Nvidia (применяемых в суперЭВМ)

Показатель	MI210	MI250	MI250X	A100 PCIe	A100 SXM	H100 PCIe	H100 SXM
Пиковая производительность FP64 (TFLOPS)	22.6	45.3	47.9	9.7	9.7	25.6	33.5
Пиковая производительность FP64 с тензорными ядрами (TFLOPS)	45.3	90.5	95.7	19.5	19.5	51.2	66.9
Емкость памяти, ГБ	64	2x64	2x64	80	80	80	80
Ее пропускная способность, Гбайт/с	1638	3277 (2x 1638)	3277 (2x 1638)	1935	2039	2039	3352
Энергопотребление, Вт	300	500	500	300	400	350	700
Цена, тысяч \$	10	15	~15	15-16		30	

Цены: июль — сентябрь 2023. Цена MI250X — для сервера с 4 GPU

Квантовая химия в гауссовском базисе.

N - число базисных функций (пропорционально числу атомов).

Методы	Масштабирование
Полуэмпирические (NDO)	$O(N^3)$: диагонализация
Неэмпирические:	
HF, DFT	$O(N^4)$: интегралы
MP2	$O(N^5)$
MP3, CISD, QCISD, CCSD	$O(N^6)$
MP4, QCISD(T), CCSD(T)	$O(N^7)$
MP5, CISDT, CCSDT	$O(N^8)$
MP7, CISDTQ, CCSDTQ	$O(N^{10})$
FCI	$O(N!)$

Данные из https://spindynamics.org/documents/cqc_lecture_5.pdf

Черное жирное — использовалось ранее

Синее — основное сегодня

Красное — дает достаточную точность 2 ккал/моль

Борьба за снижение p в $O(N^p)$:

- умножение матриц уже $O(N^{2.4})$
- уменьшение p за счет отбрасывания маленьких интегралов
- разработка новых, в т.ч. линейно-масштабируемых методов: другие базисы; специальные варианты расчета
- специализированные методы расчета больших молекул



Вычислительные показатели молекулярной динамики (МД)

I. Масштабирование с числом атомов N

- классическая МД $O(N^2)$
- классическая сильно-связанная МД (ТВМД) $O(N^3)$
- квантовая МД (DFT в плоских волнах) $O(N^3)$
(диагонализация)
- квантовая МД Кар-Парринелло (CPMD) $O(N^3)$
(умножение матриц)
- Понижение степени N в реальных специальных случаях

II. Является на GPU обычно вычислительно-связанной

: Квантовая МД - вычислительно связанная

: Классическая МД на GPU : вычислительно связанная

: Ядра (kernels), связанные памятью, забирают 15-20% общего времени расчета

Anderson J. A., Lorenz C. D., Travesset A. General purpose molecular dynamics simulations fully implemented on graphics processing units //Journal of computational physics. – 2008. – Т. 227. – №. 10. – С. 5342-5359.

III. Для классической МД обычно не требуется очень большой емкости памяти

Вычислительная гидродинамика (CFD), $O(N^2)$ обычно связана памятью (умножение матрицы на вектор)

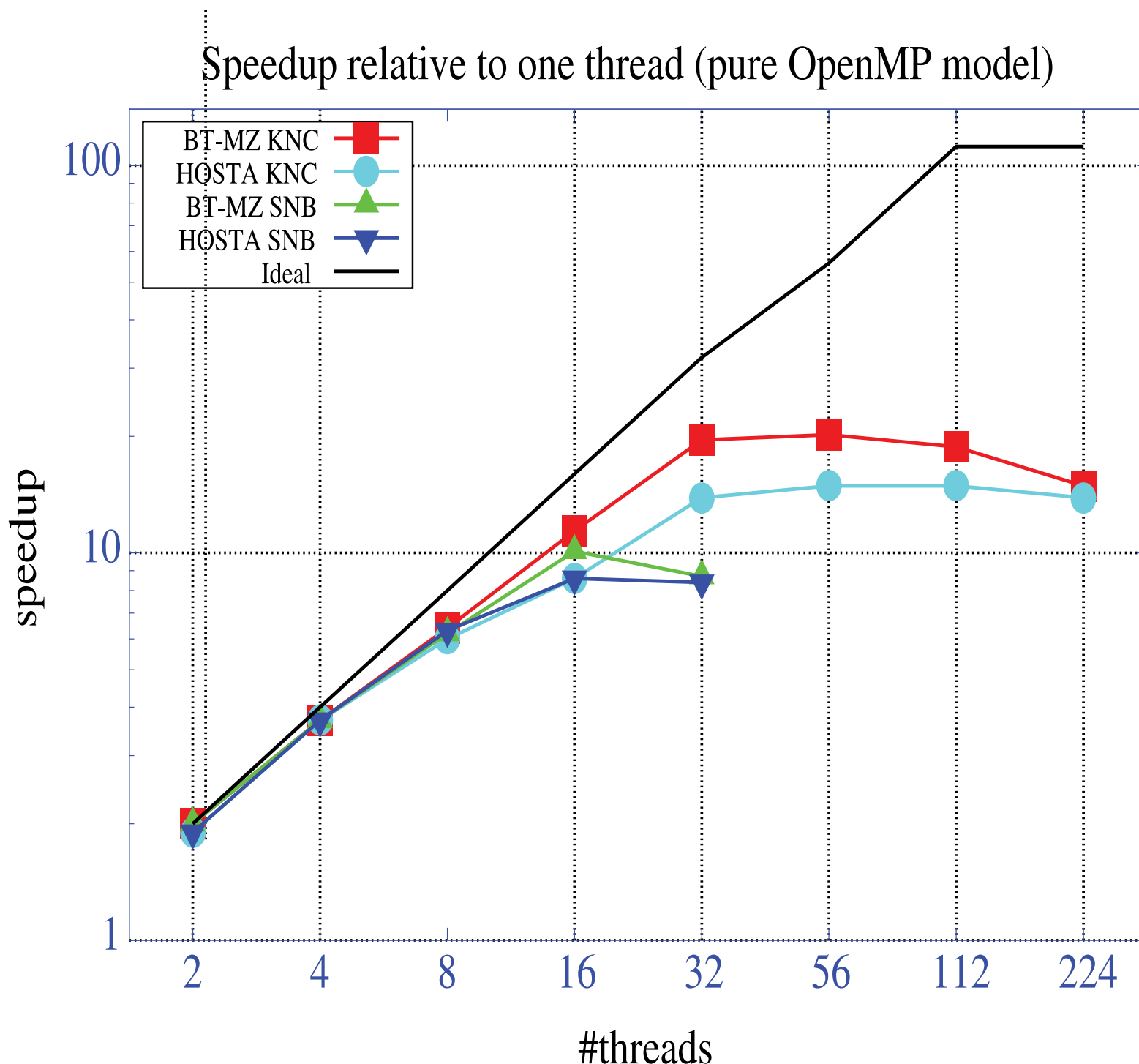


Рисунок из: Che Y. et al. Realistic performance characterization of CFD applications on intel many integrated core architecture //The Computer Journal. – 2015. – Т. 58. – №. 12. – С. 3279-3294.

KNC: Xeon Phi 3110P (57 ядер); SNB: 2xXeon E5-2670, 16 cores
Prichard R., Strasser W. **When Fewer Cores Is Faster: A Parametric Study of Under-subscription in High-Performance Computing** //Cluster Computing. – 2024. – С. 1-14.

Производительность (нс/день) программы молекулярной динамики **AMBER22** на разных GPU (чем больше — тем лучше).

Тест	A100-PCIe	MI250	H100-PCIe
JAC Production NVE 4fs	1199.22	1871	1479.32
JAC Production NPT 4fs	1194.5	1794	1424.90
STMV Production NPT 4fs	52.02	80.65	

<https://www.exxactcorp.com/blog/Molecular-Dynamics/RTX3090-Benchmarks-for-HPC-AMBER22-A100-vs-RTX3080-vs-RTX3070-vs-RTX6000>; <https://www.amd.com/en/graphics/server-accelerators-benchmark>

Время расчета поправки T (в секундах) по квантовохимическому методу CCSD(T) на GPU (программа **NWChemEX**).

GPU	Модель программирования		
	SYCL	CUDA	HIP
MI250X (1 GCD)	17.41	-	15.56
MI250X (2 GCD)	8.97	-	8.12
A100	18.23	16.14	-

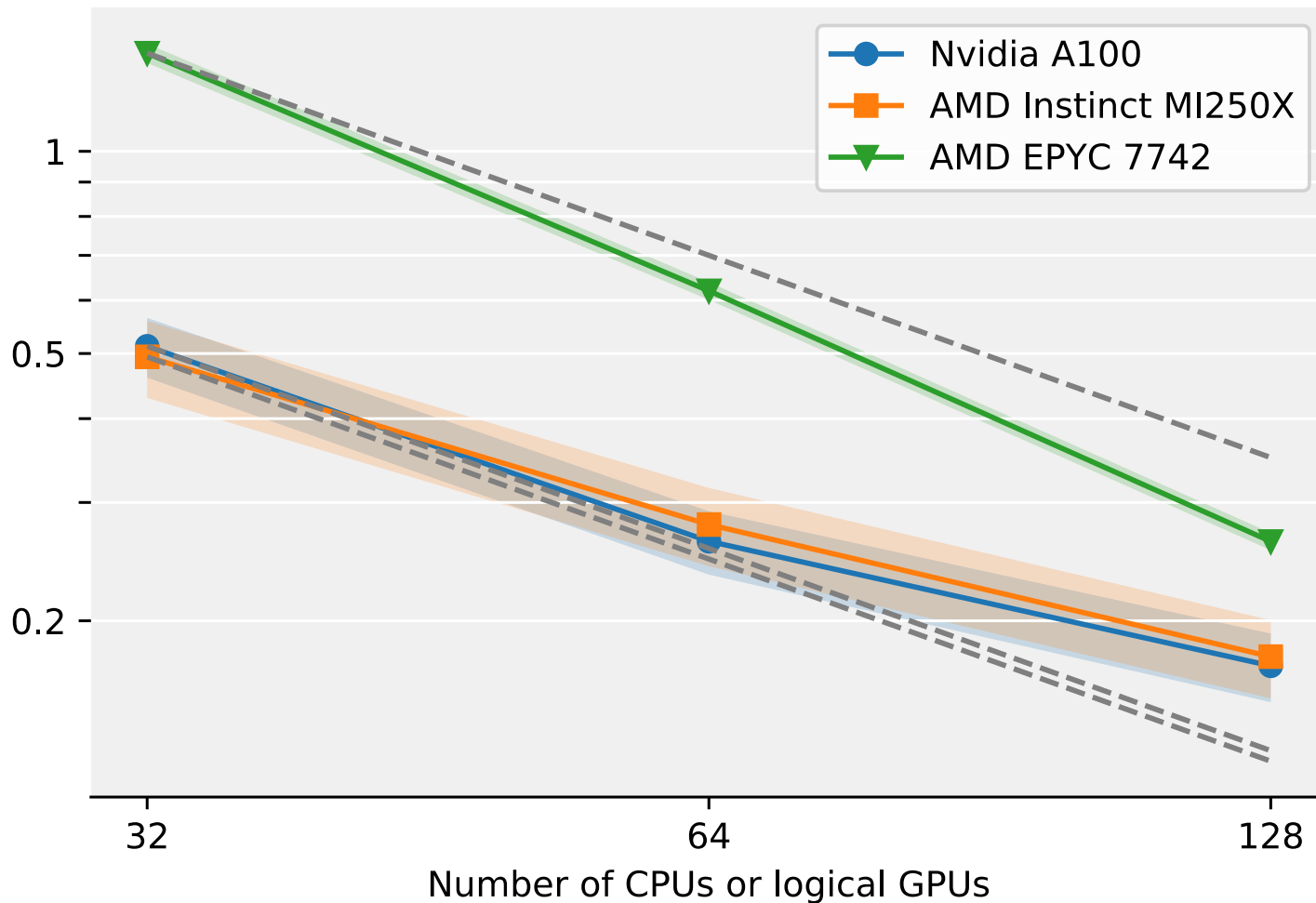
Bagusetty A. et al. Towards Cross-Platform Portability of Coupled-Cluster Methods with Perturbative Triples using SYCL //2022 IEEE/ACM International Workshop on Performance, Portability and Productivity in HPC (P3HPC). - IEEE, 2022. - С. 81-88.

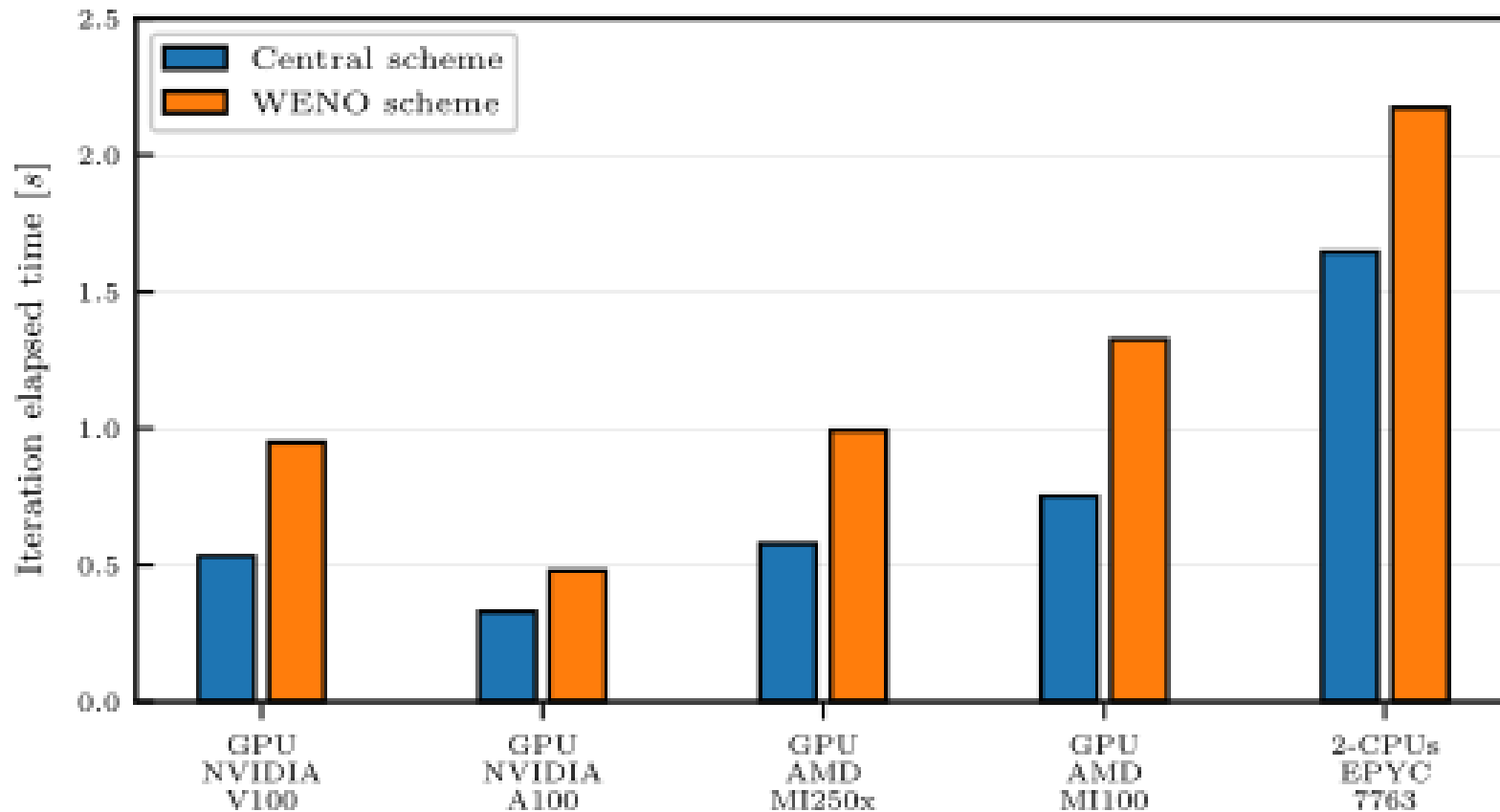
Масштабирование производительности CFD-приложения Неко в зависимости от числа логических GPU.

Karp M. et al. Large-Scale Direct Numerical Simulations of Turbulence Using GPUs and Modern Fortran //arXiv preprint arXiv:2207.07098. – 2022.

Strong Scaling

[s / time step]





CFD: Время на итерацию (сек.) для приложения STREAMS-2 для двух расчетных схем: центральной схемы оценки потока и схемы WENO. Для MI250X — один логический GPU (1 GCD). Sathyanarayana S. et al. High-speed turbulent flows towards the exascale: STREAMS-2 porting and performance //arXiv preprint arXiv:2304.05494.-2023

Производительность классической МД на разных GPU Nvidia



Данные для **AMBER24** из <https://www.exxactcorp.com/blog/molecular-dynamics/amber-molecular-dynamics-nvidia-gpu-benchmarks>. Размер системы: 1,067,095 атомов. **RTX 4090** быстрее еще и в **GROMACS**.

Если есть возможность выбора,
можно определять, где лучше считать:

- на ноутбуке с GPU, на ПК или в кластере ПК (проблема возможного отсутствия ECC-кодов в памяти)
- на своем сервере или кластере
- использовать облачную технологию для HPC: тогда не нужно своих серверов
- на суперкомпьютере: надо учитывать возможности распараллеливания и цены