XDMoD METRICS ON DEMAND

NSF OAC Awards: 1927880 and 2137603

Benchmarking High-End ARM Systems with Scientific Applications. Performance and Energy Efficiency.

Nikolay Simakov*, Robert DeLeon*, Joseph White*, Matthew Jones*, Thomas Furlani*, Eva Siegmann + and Robert Harrison +

*Center for Computational Research, SUNY University at Buffalo
+Institute for Advanced Computational Science, Stony Brook University

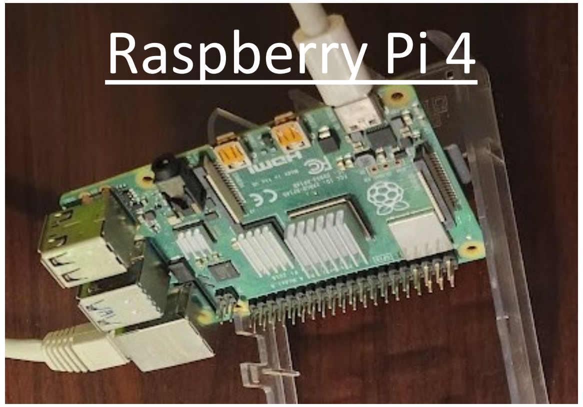
University at Buffalo Center for Computational Research

Stony Brook University

iACS INSTITUTE FOR ADVANCED COMPUTATIONAL SCIENCE

CUG 2023, Helsinki, Finland, May 7 – 11, 2023

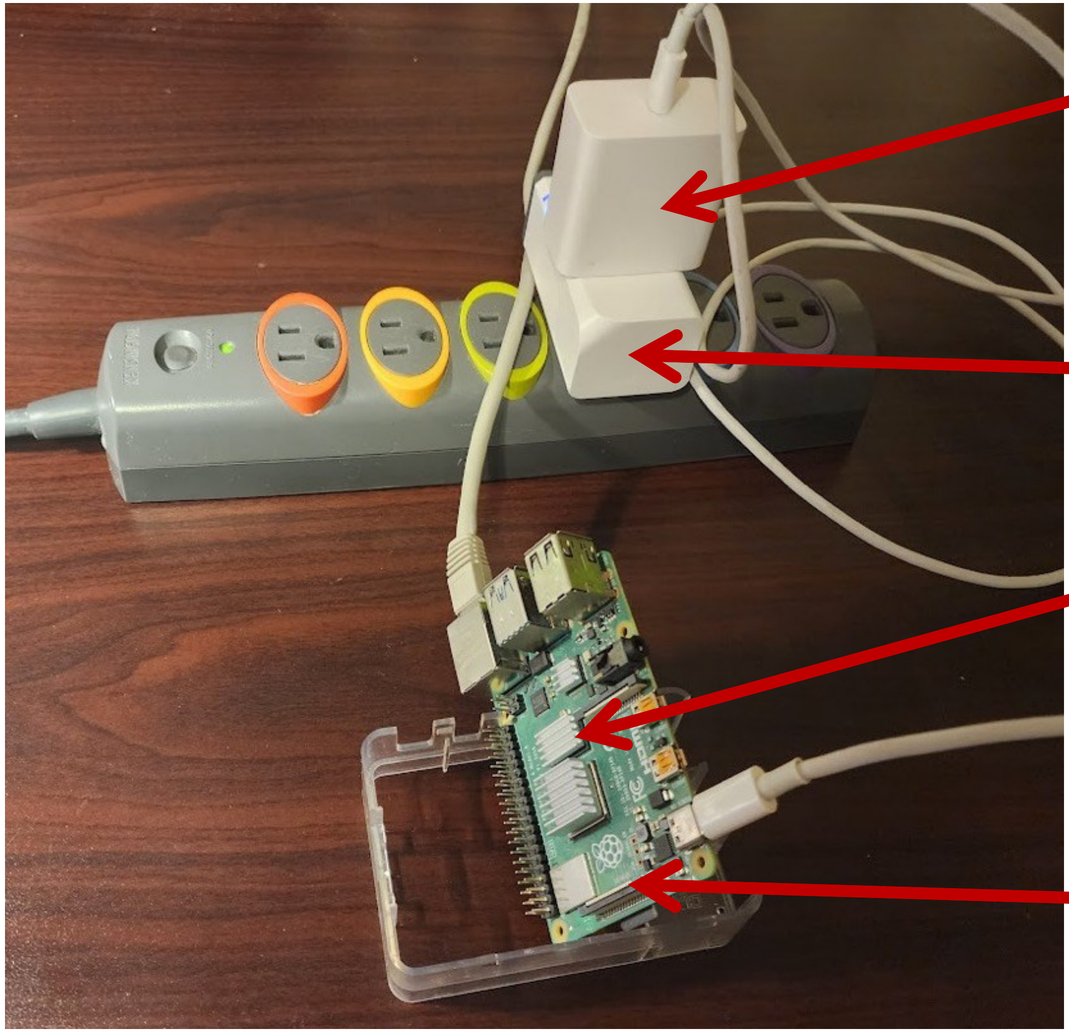
ARM Processors are going to HPC Market



- **Energy efficiency is crucial for our ever-increasing demand for compute power**
  - Information-Communications-Technologies (ICT) ecosystem uses about 10% of world electricity generation. Projected to reach 20% by 2030*
  - Exascale computing is not sustainable without adequate energy efficiency. Frontier, 1.1 Eflop/s machine, consumes 21 MW (17,000 households).
- **ARM CPUs are successfully used in many products, including energy consumption-sensitive products.**
  - Embedded systems and mobile computing devices, like smartphones and tablets
  - Linux server products such as file and web servers
- **More recently, ARM CPUs have been adapted to HPC workloads, and some are specifically designed for scientific calculations**
- **What is their performance for HPC workloads?**
  - What is the performance state of modern high-end ARM CPUs?
  - How do they compare in performance to x86 systems
  - Are we ready for broader adoption of ARM in the HPC community?

*[https://www.nature.com/articles/d41586-018-06610-y]

First Personal Experience with HPC Application on ARM



USB-C interface provides enough power

Smart outlet provides Power measurements

Raspberry Pi 4

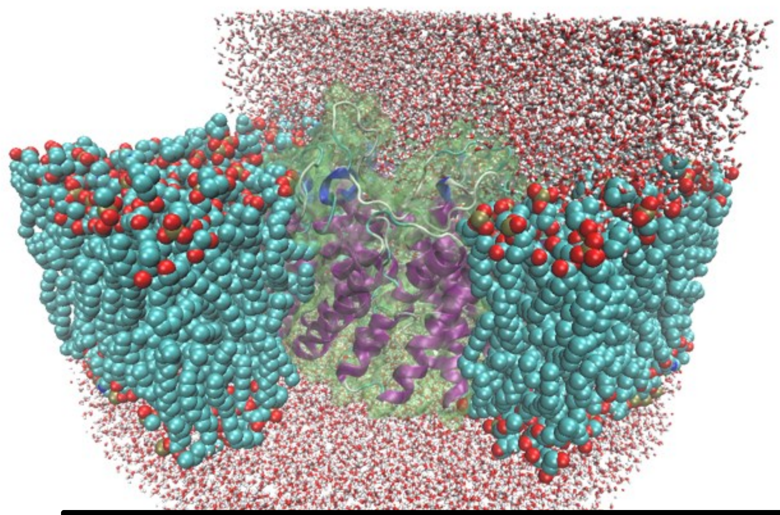
Vertical placement for Efficient cooling

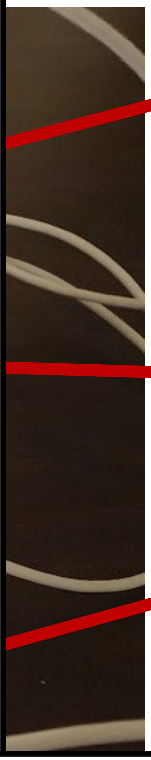
First Personal Experience with HPC Application on ARM

Gromacs

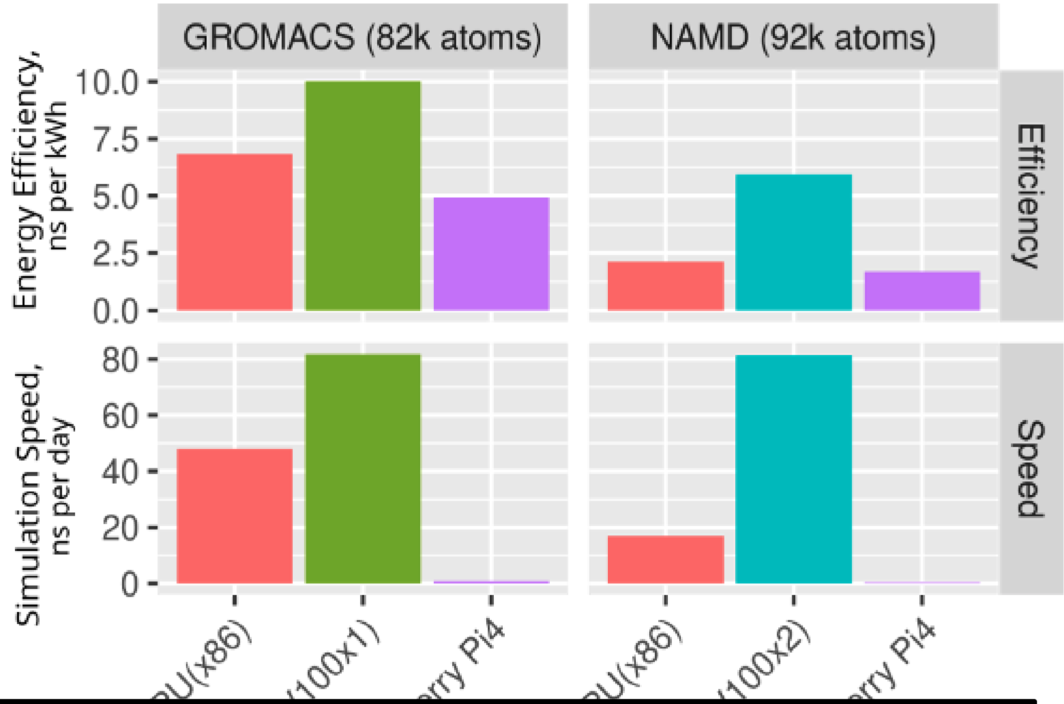
Membrane protein system

~82k atoms





Performance



nd
tel
i4

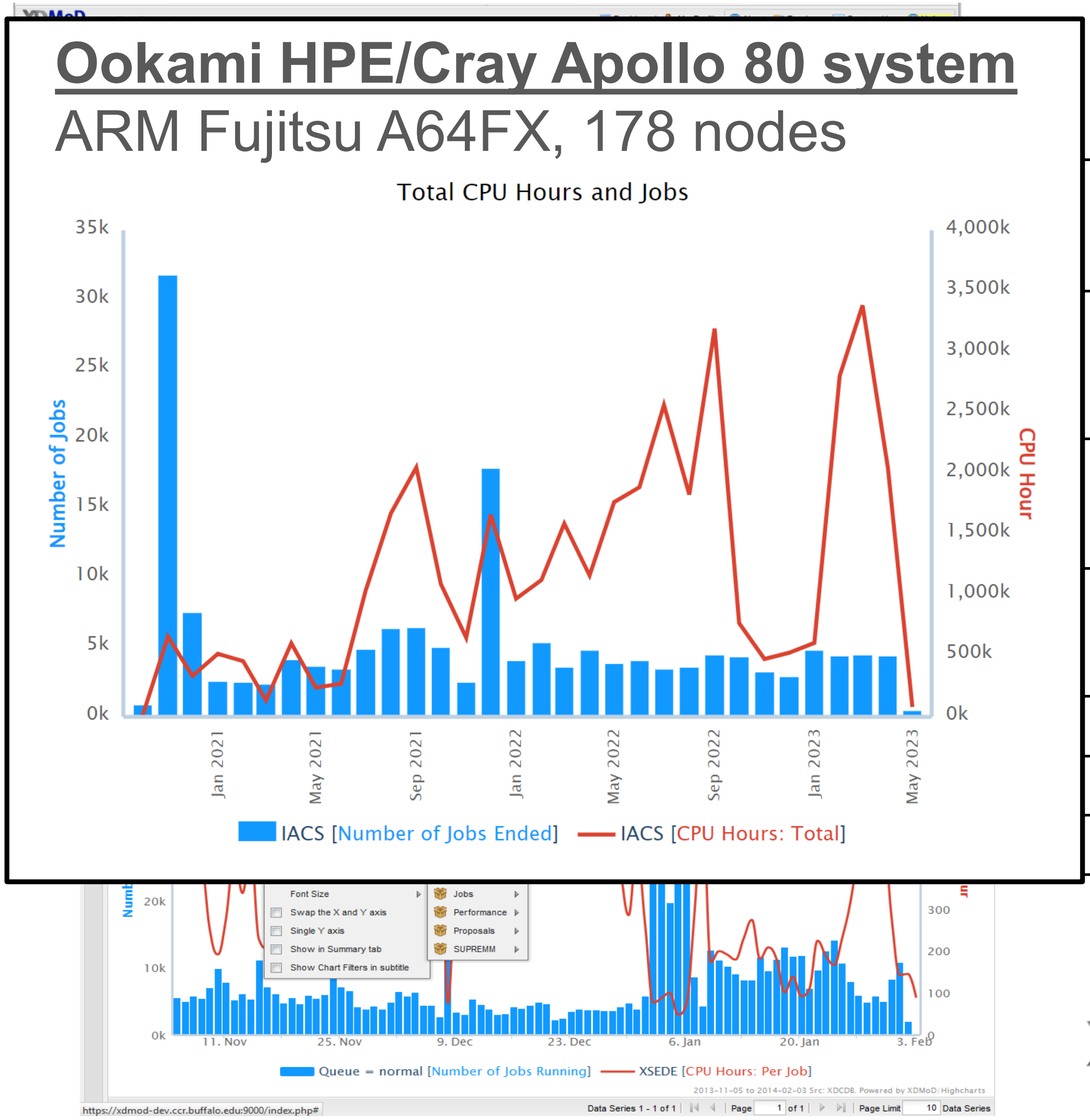
Results:

Good: everything compiled and ran

Bad: Raspberry Pie 4 is neither fast nor energy efficient in compute intensive application like molecular dynamics

What is the performance state of modern high end ARM CPUs?

XDMoD: A Comprehensive Tool for HPC System Management



NSF ACCESS Measurement and Metrics Service (MMS),

- Following XD Net Metrics Service (XMS) and prior 5 year TAS award
- Develop & deploy **XDMoD** (**XD** **M**etrics **o**n **D**emand) Tool

Open XDMoD: Open Source version for Data Centers

- Used to measure and optimize performance of HPC centers
- 300+ academic & industrial installations worldwide

Goal: Optimize Resource Utilization and Performance

- Provide detailed information on utilization
- Measure Quality of Service
- Enable data driven upgrades and procurements
- Measure and improve job and system level performance

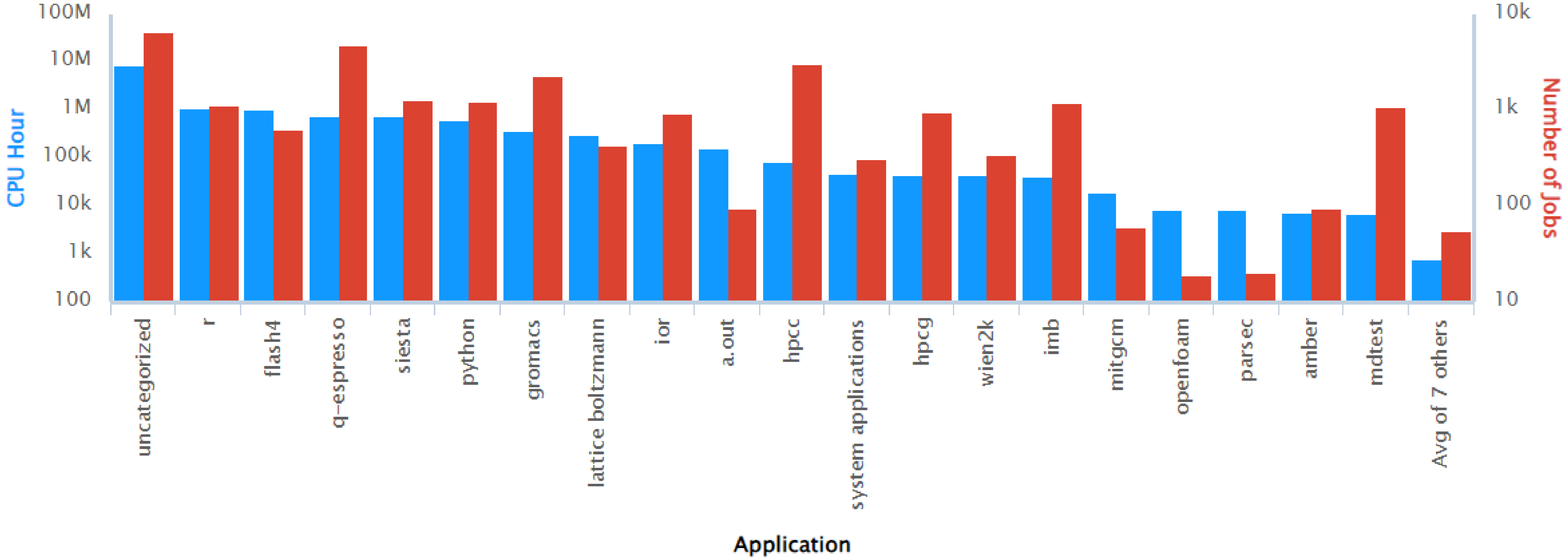
XDMoD METRICS ON DEMAND

University at Buffalo
Center for Computational Research

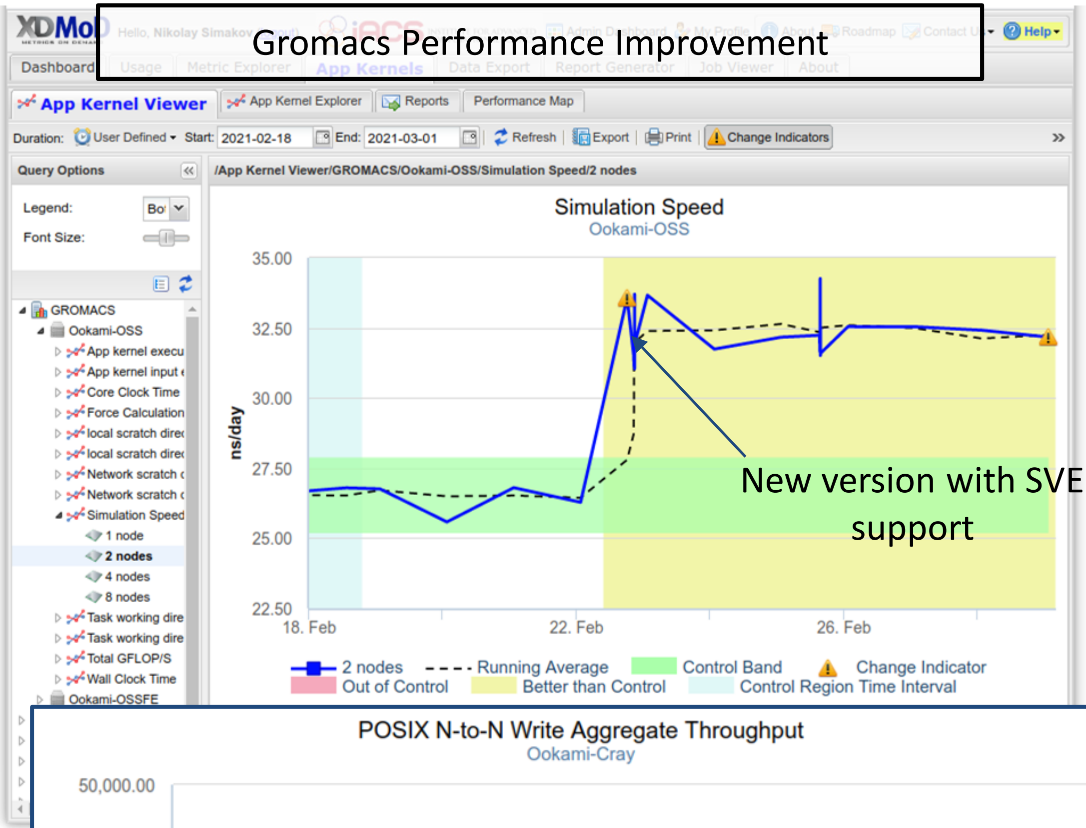
Application usage on Ookami (Fujitsu A64FX)

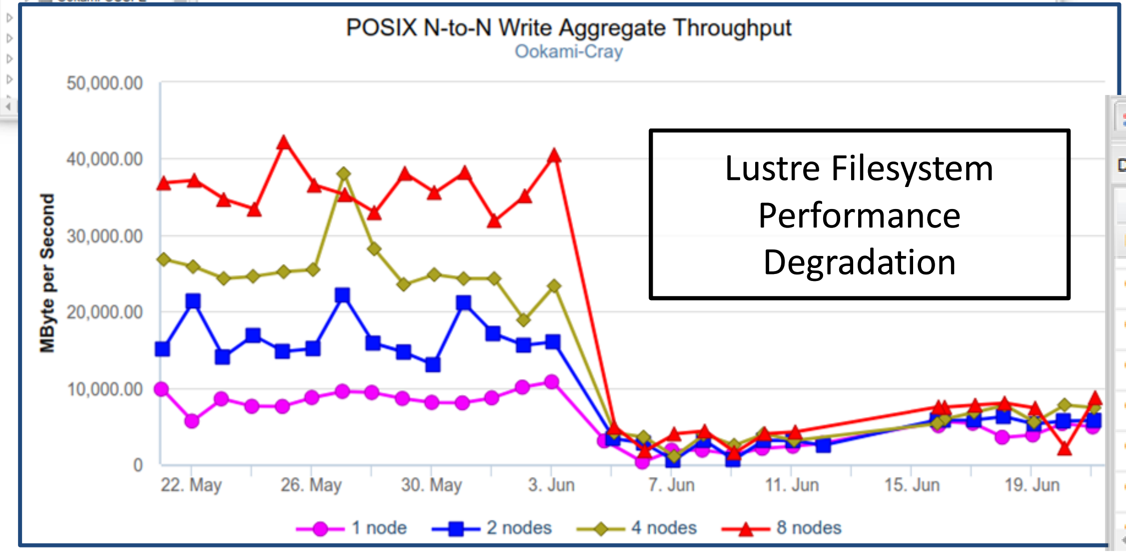
Ookami – an ARM Fujitsu A64FX machine with SVE support (512 bit wide)

Determine what are the mostly widely used applications (2021-01 application usage to 2022-09-30 shown)



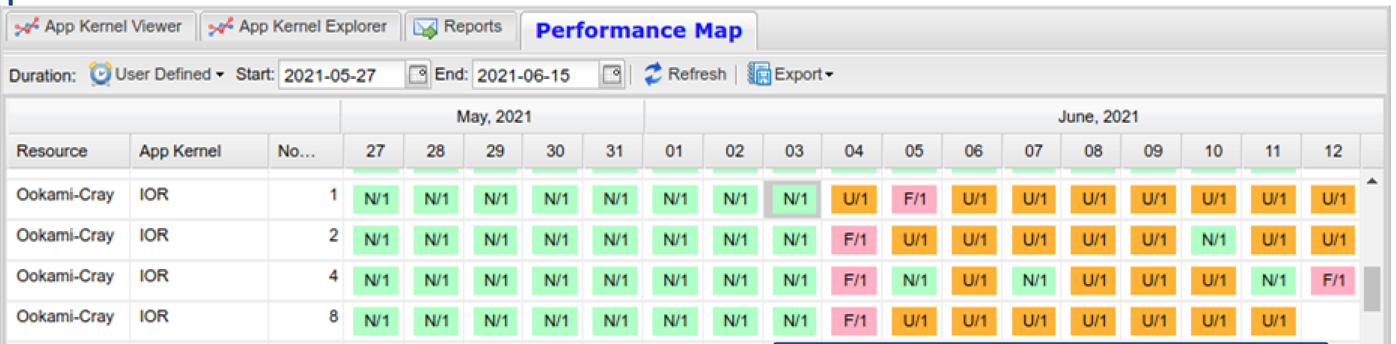
QoS and Performance Monitoring with Application Kernels





Application kernels module allows **continuous performance monitoring** by periodic execution of applications and benchmarks.

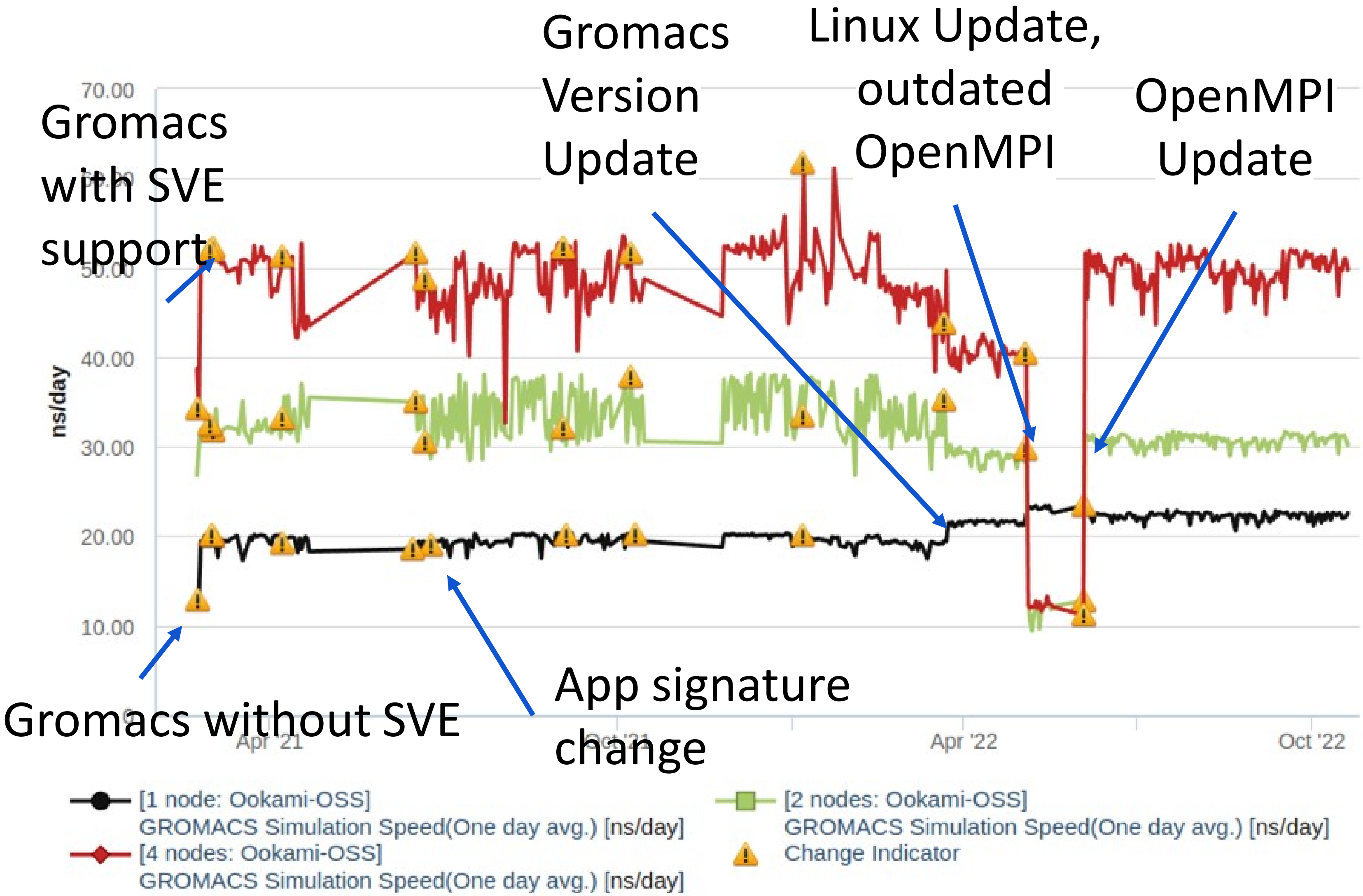
- Computationally lightweight benchmarks or applications
- Run periodically or on demand to actively measure performance
- Measure system performance from User's perspective
- Proactively identify underperforming hardware and software



			May, 2021					June, 2021											
Resource	App Kernel	No...	27	28	29	30	31	01	02	03	04	05	06	07	08	09	10	11	12
Ookami-Cray	IOR	1	N/1	N/1	N/1	N/1	N/1	N/1	N/1	N/1	U/1	F/1	U/1	U/1	U/1	U/1	U/1	U/1	U/1
Ookami-Cray	IOR	2	N/1	N/1	N/1	N/1	N/1	N/1	N/1	N/1	F/1	U/1	U/1	U/1	U/1	U/1	N/1	U/1	U/1
Ookami-Cray	IOR	4	N/1	N/1	N/1	N/1	N/1	N/1	N/1	N/1	F/1	N/1	U/1	N/1	U/1	U/1	U/1	N/1	F/1
Ookami-Cray	IOR	8	N/1	N/1	N/1	N/1	N/1	N/1	N/1	N/1	F/1	U/1	U/1	U/1	U/1	U/1	U/1	U/1	
Ookami-Cray	MDTest	1	N/1	N/1	N/1	N/1	N/1	N/1	N/1										
Ookami-Cray	MDTest	2	N/1	N/1	N/1	N/1	N/1	N/1	N/1										

Code	Description
N	Application kernel was executed within control interval
U	Application kernel was under-performing
O	Application kernel was over-performing
F	Application kernel failed to run

Performance over Time: GROMACS on Ookami (ARM Fujitsu A64FX)



Tested Compute Resources

Resource	CPU	Arch/Core Name	Proc., nm	SIMD	Release Date	Cores / Node	Freq, GHZ base/turbo
ARM							
SBU-Ookami	Fujitsu A64FX	v8.2-A	7	SVE 512b	~2019	48	1.8
SBU-Ookami	Cavium ThunderX2	v8.1	14	NEON 128b	2018	64	2.0-2.5
Amazon	Amazon Graviton 2	v8.2, Neov. N1	7	128b	Nov-19	48	2.6
Amazon	Amazon Graviton 3	v8.5, Neov. V1	5	SVE 512b	Nov-21	48	2.5
Amazon	Amazon Graviton 3	v8.5, Neov. V1	5	SVE 512b	Nov-21	64	2.5
Google	Ampere Altra	v8.2+, Neov. N1	7	128b	Mar-21	48	Up to 3.0
Azure	Ampere Altra	v8.2+, Neov. N1	7	128b	Mar-21	48	Up to 3.0
Azure	Ampere Altra	v8.2+, Neov. N1	7	128b	Mar-21	64	Up to 3.0
x86 AMD							
PSC-Bridges 2	EPYC 7742	Zen2(Rome)	14	AVX2 256b	Mid-2019	128	2.25/3.4
SDSC-Expanse	EPYC 7742	Zen2(Rome)	14	AVX2 256b	Mid-2019	128	2.25/3.4
Purdue-Anvil	EPYC 7763	Zen3(Milan)	7+	AVX2 256b	Mar-21	128	2.45/3.5
x86 Intel							
TACC-Stampede 2	Xeon Phi 7250	Knights Landing	14	AVX512	Q2 2016	68	1.4/1.6
TACC-Stampede 2	Xeon Platinum 8160	Skylake-X	14	AVX512	Q3 2017	48	2.1/3.7
TACC-Stampede 2	Xeon Platinum 8380	Ice Lake	10	AVX512	Q2 2021	80	2.3/3.4
UB-HPC	Xeon Gold 6130	Skylake-X	14	AVX512	Q3 2017	32	2.1/3.7
UB-HPC	Xeon Gold 6330	Ice Lake	10	AVX512	Q2 2021	56	2/3.7
x86 Intel and NVIDIA GPU							
UB-HPC V100x2	Xeon Gold 6130	Skylake-X	14	AVX512	Q3 2017	32	2.1/3.7
UB-HPC A100x2	Xeon Gold 6330	Ice Lake	10	AVX512	Q2 2021	56	2/3.7

Overall, eighteen different hardware configurations were used from nine different resource providers, including the clouds (accessed through CloudBank) and traditional HPC services (ACCESS allocations, UB, SBU).

All calculations were performed on a single node or single virtual machine instance.

Tested Applications and Benchmarks

- The application kernels used for this study span a variety of computational domains and paradigms:
  - HPCC – multiple benchmarks, including LINPACK and FFT
  - NWChem - ab initio chemistry
  - Open Foam - partial differential equation solver
  - GROMACS - biomolecular simulation
  - AI Benchmark Alpha - AI benchmark
  - Enzo - adaptive mesh refinement

HPCC: HPC challenge benchmark

HPC Challenge Benchmark combine multiple benchmarks together

- High Performance LINPACK, which solves a linear system of equations and measures the floating-point performance
- Matrix-matrix multiplication
- Fast Fourier Transform
- Stream: memory bandwidth
- Parallel Matrix Transpose
- MPI Random Access

HPCC: HPC challenge benchmark

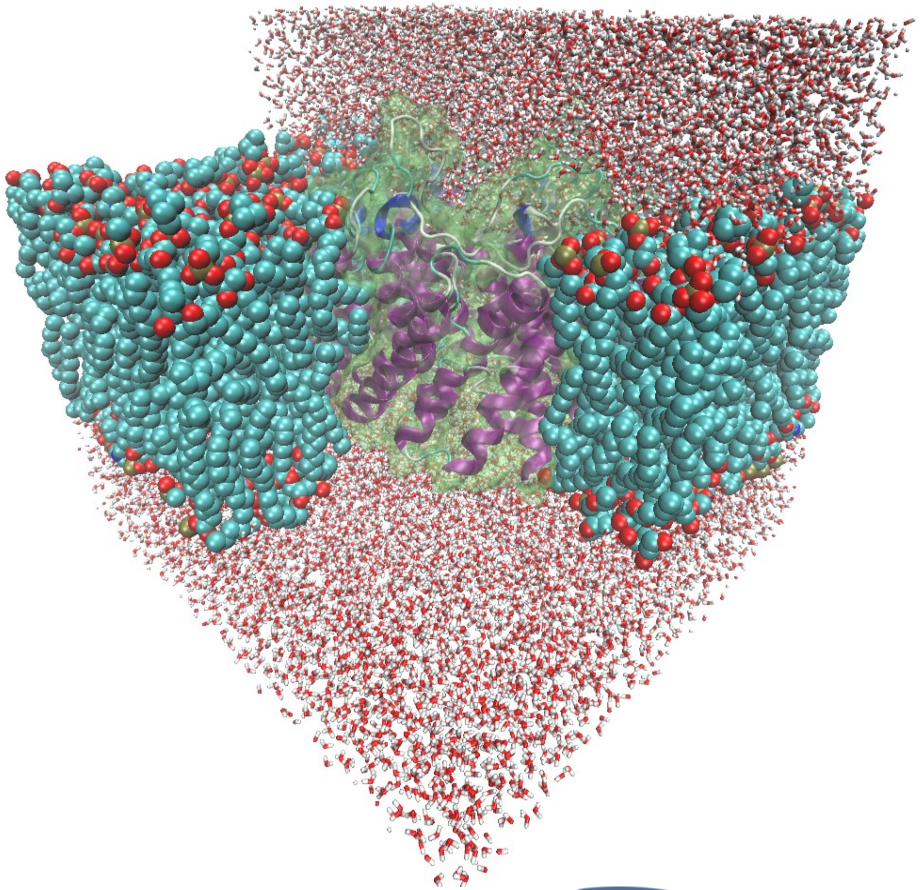
CPU/System	Cores	Matrics Multiplication		LINPACK		FFT		Power, W	Energy Eff., Jobs per kWh	N
		GFLOPS	GFLOPS/Core	GFLOPS	GFLOPS/Core	GFLOPS	GFLOPS/Core			
ARM Fujitsu A64FX, SVE 512b (SBU-Ookami, GCC)	48	1363	28.4 ± 0.1	828 ± 27	17.3	6.2 ± 0.7	0.13	110**	560	60
ARM Fujitsu A64FX, SVE 512b (SBU-Ookami, FJ)	48	1978	41.2 ± 0.2	1177 ± 19	24.5	24.4 ± 0.9	0.51	110**	185	60
ARM Fujitsu A64FX, SVE 512b (SBU-Ookami, ARM)	48	1651	34.4 ± 1.8	884 ± 72	18.4	0.3 ± 0.0	0.01	110**	335	60
ARM Fujitsu A64FX, SVE 512b (SBU-Ookami, Cray)	48	917	19.1 ± 3.7	758 ± 149	15.8	6.8 ± 0.6	0.14	110**	204	60
ARM Cavium ThunderX2 (SBU-Ookami)	64	742	11.6 ± 2.1	522 ± 106	8.2	33.5 ± 3.7	0.52			14
ARM Amazon Graviton 2, Neoverse N1 (AWS)	48	816	17 ± 0.0	682 ± 1	14.2	27.1 ± 0.6	0.56			20
ARM Amazon Graviton 3, Neoverse V1, SVE 256b (AWS)	48	907	18.9 ± 0.3	776 ± 10	16.2	55.4 ± 0.5	1.15			20
ARM Amazon Graviton 3, Neoverse V1, SVE 256b (AWS)	64	1158	18.1 ± 0.0	965 ± 1	15.1	71 ± 0.7	1.11			20
ARM Ampere Altra, Neoverse N1 (Azure)	48	816	17 ± 0.0	675 ± 17	14.1	26.5 ± 0.4	0.55			11
ARM Ampere Altra, Neoverse N1 (Azure)	48	826	17.2 ± 0.0	691 ± 18	14.4	26.8 ± 0.7	0.56			11
ARM Ampere Altra, Neoverse N1 (Azure)	64	1037	16.2 ± 0.0	850 ± 4	13.3	33.1 ± 1.1	0.52	270*	314	20
x86 AMD EPYC 7742 Zen2(Rome), AVX2 (PSC-Bridges-2)	128	2624	20.5 ± 0.7	1895 ± 42	14.8	50.3 ± 0.3	0.39			20
x86 AMD EPYC 7742 Zen2(Rome), AVX2 (SDSC Expanse)	128	3200	25 ± 1.4	1721 ± 47	13.4	71.8 ± 2.0	0.56			20
x86 AMD EPYC 7763 Zen3(Milan), AVX2 (Purdue Anvil)	128	3046	23.8 ± 1.6	2176 ± 100	17.0	54.7 ± 4.8	0.43			20
x86 Intel Xeon Phi 7250, KNL, AVX512 (TACC-Stampede 2)	68	340	5 ± 0.3	986 ± 8	14.5	46.5 ± 0.7	0.68			20
x86 Intel Xeon Plat. 8160, Skylake-X, AVX512 (TACC-Stampede 2	48	2122	44.2 ± 1.7	1158 ± 34	24.1	35.8 ± 1.9	0.75			20
x86 Intel Xeon Plat. 8380, Ice Lake, AVX512 (TACC-Stampede 2)	80	3824	47.8 ± 0.6	1713 ± 5	21.4	76.4 ± 2.0	0.96			12
x86 Intel Xeon Gold 6130, Skylake-X, AVX512 (UB-HPC)	32	1536	48 ± 2.0	997 ± 54	31.2	50.9 ± 1.9	1.59	345±31	74	53
x86 Intel Xeon Gold 6330, Ice Lake, AVX512 (UB-HPC)	56	2761	49.3 ± 1.1	1396 ± 37	24.9	47.9 ± 0.7	0.86	588±64	109	12
x86 Intel Xeon Gold 6330, Ice Lake, AVX512 (UB-HPC, ICC)	56	2845	50.8 ± 1.0	1399 ± 13	25.0	28.2 ± 0.3	0.50	501±107	299	12

HPCC: HPC challenge benchmark

CPU/System	Cores	Matrics Multiplication		LINPACK		FFT		Power, W	Energy Eff., Jobs per kWh	N
					GFLOPS/C		GFLOPS/			
ARM Fujitsu A								*	560	60
ARM Fujitsu A								*	185	60
ARM Fujitsu A								*	335	60
ARM Fujitsu A								*	204	60
ARM Cavium T										14
ARM Amazon										20
ARM Amazon										20
ARM Amazon										20
ARM Ampere										11
ARM Ampere										11
ARM Ampere									314	20
x86 AMD EPY										20
x86 AMD EPY										20
x86 AMD EPY										20
x86 Intel Xeon										20
x86 Intel Xeon										20
x86 Intel Xeon										12
x86 Intel Xeon								31	74	53
x86 Intel Xeon								64	109	12
x86 Intel Xeon Gold 6330, Ice Lake, AVX512 (UB-HPC, ICC)	56	2845	50.8 ± 1.0	1399 ± 13	25.0	28.2 ± 0.3	0.50	501±107	299	12

- **Matrix-Matrix Multiplication**
  - Dual modality: 512 bit SIMD offer best performance
  - Fujitsu A64FX has highest performance in ARM camp
  - Overall ARM performance is comparable to x86 counterparts
- **LINPACK**
  - Similar to DGEMM there is dual mode behaiviour
  - Fujitsu A64FX is fastest in ARM camp
  - Per core performance is competitive with x86 counterpart with similar SIMD widths
- **FFT**
  - Graviton 3 is fastest in ARM camp. It is also has highest per core performance overall.
  - Overall performance is competitive with x86 counterpart with similar SIMD widths

GROMACS: Molecular Dynamics of Biomolecular Systems



FAST. FLEXIBLE. FREE.
GROMACS

GROMACS is molecular dynamics simulation of biomolecular systems

Application computational characteristics:

- Solve ODE (second Newton law)
- Particle interactions
  - Short range/long range
- FFT

Test case:

- Membrane protein
- 82k atoms system

GROMACS: Molecular Dynamics of Biomolecular Systems

CPU/System	Cores	Simulation Speed, ns/day	Simulation Speed per Core, ns/day/core	Power, W	Energy Efficiency, ns/kWh	N
ARM Fujitsu A64FX, SVE 512bit (SBU-Ookami, GCC)	48	22.3 ± 0.2	0.46	111 ± 8	8.43 ± 0.6	32
ARM Fujitsu A64FX, SVE 512bit (SBU-Ookami, Fujitsu)	48	22.8 ± 0.3	0.48	105 ± 5	9.06 ± 0.4	12
ARM Cavium ThunderX2 (SBU-Ookami)	64	28.8 ± 4.2	0.45			14
ARM Amazon Graviton 2, Neoverse N1 (AWS)	48	37.8 ± 0.1	0.79			20
ARM Amazon Graviton 3, Neoverse V1, SVE 256bit (AWS)	48	57.0 ± 0.4	1.19			20
ARM Amazon Graviton 3, Neoverse V1, SVE 256bit (AWS)	64	71.4 ± 1.0	1.12			20
ARM Ampere Altra, Neoverse N1 (Google)	48	39.0 ± 1.8	0.81			11
ARM Ampere Altra, Neoverse N1 (Azure)	48	41.0 ± 2.2	0.85			11
ARM Ampere Altra, Neoverse N1 (Azure)	64	56.5 ± 0.6	0.88	270 *	8.71	20
x86 AMD EPYC 7742 Zen2(Rome), AVX2 (PSC Bridges-2)	128	109.6 ± 4.8	0.86			20
x86 AMD EPYC 7742 Zen2(Rome), AVX2 (SDSC Expanse)	128	99.8 ± 8.6	0.78			20
x86 AMD EPYC 7763 Zen3(Milan), AVX2 (Purdue Anvil)	128	169.9 ± 4.4	1.33			20
x86 Intel Xeon Plat. 8160, Skylake-X, AVX512 (TACC-Stampede 2)	48	70.4 ± 0.8	1.47			11
x86 Intel Xeon Plat. 8380, Ice Lake, AVX512 (TACC-Stampede 2)	80	133.3 ± 6.0	1.67			20
x86 Intel Xeon Gold 6130, Skylake-X, AVX512 (UBHPC)	32	37.6 ± 0.9	1.18	379 ± 33	4.17 ± 0.4	21
x86 Intel Xeon Gold 6130, Skylake-X, AVX512 (UBHPC ICC)	32	39.3 ± 0.9	1.23	367 ± 35	4.5 ± 0.5	11
x86 Intel Xeon Gold 6330, Ice Lake, AVX512 (UBHPC)	56	81.7 ± 6.9	1.46	633 ± 28	5.38 ± 0.4	12
x86 Intel Xeon Gold 6330, Ice Lake, AVX512 (UBHPC ICC)	56	103.0 ± 2.0	1.84	619 ± 17	6.94 ± 0.2	17
x86 Intel Xeon Gold 6130, NVIDIA V100x2 (UBHPC)	32	145.1 ± 2.8		435 ± 7	13.91 ± 0.3	12
x86 Intel Xeon Gold 6330, NVIDIA A100x2 (UBHPC)	56	236.5 ± 10.8		707 ± 9	13.94 ± 0.8	11

- **Graviton 3 is the fastest in ARM camp**
- **Per core performance of Graviton 2 and Ampere Alta is similar to Zen 2 (Rome) and Graviton 3 is approaching zen3 (Millan) and Skylake-X**

NWChem: Quantum Chemistry

- NWChem is an ab initio computational chemistry software package developed by Pacific Northwest National Laboratory.
- The input to the benchmark runs is the Hartree-Fock energy calculation of Au+ with MP2 and Coupled Cluster corrections

NWChem: Open Source High-Performance Computational Chemistry

EMSL

NWChem: Quantum Chemistry

CPU/System	Cores	Wall Clock Time, Seconds	Power, W	Energy Efficiency, Jobs per kWh	N
ARM Fujitsu A64FX, SVE 512b (SBU-Ookami, GCC)	48	62.7 ± 0.7	110 ± 0	522 ± 6	60
ARM Amazon Graviton 2, Neoverse N1 (AWS)	48	61.1 ± 0.9			12
ARM Amazon Graviton 3, Neoverse V1, SVE 256b (AWS)	48	36.6 ± 0.7			11
ARM Amazon Graviton 3, Neoverse V1, SVE 256b (AWS)	64	29.8 ± 0.4			20
ARM Ampere Altra, Neoverse N1 (Azure)	48	56.5 ± 2.7			11
ARM Ampere Altra, Neoverse N1 (Azure)	64	42.8 ± 0.5	270*	312	20
x86 AMD EPYC 7742 Zen2(Rome), AVX2 (PSC-Bridges-2)	128	32.4 ± 4.4			20
x86 AMD EPYC 7742 Zen2(Rome), AVX2 (SDSC Expanse)	128	28.6 ± 7.8			20
x86 AMD EPYC 7763 Zen3(Milan), AVX2 (Purdue Anvil)	128	26.7 ± 0.3			20
x86 Intel Xeon Phi 7250, KNL, AVX512 (TACC-Stampede 2)**	68	262.1 ± 22.1			20
x86 Intel Xeon Plat. 8160, Skylake-X, AVX512 (TACC-Stampede 2)**	48	50.3 ± 0.3			12
x86 Intel Xeon Plat. 8160, Skylake-X, AVX512 (TACC-Stampede 2)	48	31.2 ± 0.2			8
x86 Intel Xeon Plat. 8380, Ice Lake, AVX512 (TACC-Stampede 2)	80	19.2 ± 1.2			11
x86 Intel Xeon Gold 6130, Skylake-X, AVX512 (UB-HPC)	32	90 ± 1.6	332 ± 50	124 ± 25	27
x86 Intel Xeon Gold 6330, Ice Lake, AVX512 (UB-HPC)	56	46.9 ± 0.6	376 ± 2	204 ± 3	11

- We need bigger input
- Graviton 3 is fastest in ARM Camp, 18% slower than Stampede 2 (Skylake-X)

AI-Benchmark-Alpha (Tensorflow)

Images are taken from https://ai-benchmark.com/alpha

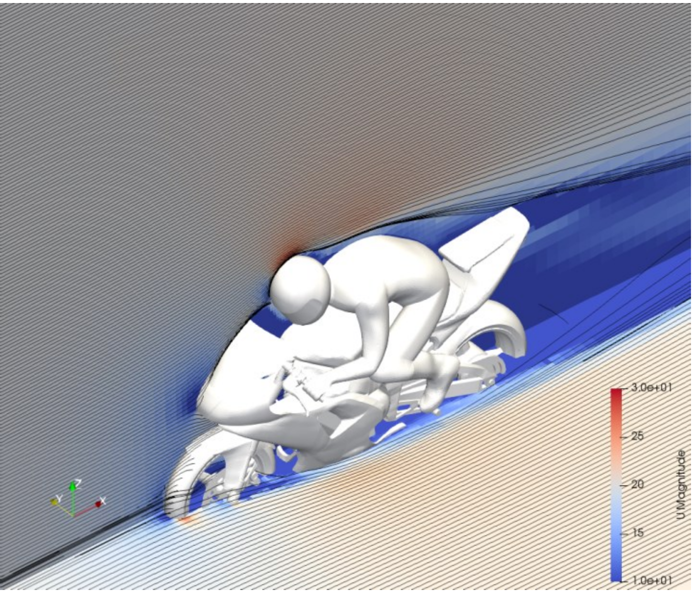
- AI-Benchmark-Alpha includes multiple machine learning tasks utilizing deep neuron networks. Tests includes classification, image to image mapping, image segmentation, image inpainting, sentence sentiment analysis and text translation.
- It is relatively light-weight
- Utilize Tensorflow for computation

AI-Benchmark-Alpha (Tensorflow)

CPU/System	Cores	Larger Better AI Score	Larger Better Inference Score	Larger Better Training Score	Power, W	AI Score per W	N
ARM Fujitsu A64FX, SVE 512b (SBU-Ookami)	48	1034 ± 3	535 ± 2	499 ± 2	111 ± 7	9.4 ± 0.6	20
ARM Amazon Graviton 2, Neoverse N1 (AWS)	48	3030 ± 12	1676 ± 7	1355 ± 6			12
ARM Amazon Graviton 3, Neoverse V1, SVE 256b (AWS)	48	4581 ± 12	2407 ± 10	2174 ± 8			11
ARM Amazon Graviton 3, Neoverse V1, SVE 256b (AWS)	64	4850 ± 31	2708 ± 21	2143 ± 13			20
ARM Ampere Altra, Neoverse N1 (Azure)	48	3177 ± 15	1803 ± 10	1375 ± 6			11
ARM Ampere Altra, Neoverse N1 (Azure)	64	3214 ± 20	1977 ± 15	1238 ± 6	270*	11.9	20
x86 AMD EPYC 7742 Zen2(Rome), AVX2 (SDSC Expanse)	128	2696 ± 17	1761 ± 14	936 ± 9			11
x86 AMD EPYC 7763 Zen3(Milan), AVX2 (Purdue Anvil)	128	3079 ± 26	1992 ± 16	1087 ± 13			11
x86 Intel Xeon Plat. 8160, Skylake-X, AVX512 (TACC-Stampede 2)	48	3606 ± 20	2292 ± 18	1314 ± 4			11
x86 Intel Xeon Plat. 8380, Ice Lake, AVX512 (TACC-Stampede 2)	80	8805 ± 27	3725 ± 20	5081 ± 14			11
x86 Intel Xeon Gold 6130, Skylake-X, AVX512 (UB-HPC)	32	3233 ± 253	1941 ± 165	1292 ± 88	403 ± 14	8 ± 0.5	11
x86 Intel Xeon Gold 6330, Ice Lake, AVX512 (UB-HPC)	56	10197 ± 53	4398 ± 31	5799 ± 29	543 ± 33	18.9 ± 1.2	12
x86 Intel Xeon Gold 6130, NVIDIA V100x2 (UB-HPC)	32	32628 ± 433	15656 ± 278	16972 ± 163	379 ± 34	86.8 ± 8.3	11
x86 Intel Xeon Gold 6330, NVIDIA A100x2 (UB-HPC)	56	59323 ± 378	29691 ± 290	29631 ± 152	561 ± 69	107.2 ± 13.6	11

- NVIDIA GPU A100x2 is the fastest
- Intel Ice Lake show fastest performance among CPUs and 4.2 times slower than V100 GPU
- ARM Graviton 3 is second fastest CPU outperformed only by Intel Icelake X
- ARM Graviton 2 and Ampre Alta show results comparable to older x86

OpenFOAM: Toolbox for numerical solvers (CFD)



- OpenFOAM is a library and collection of applications for the numerical solution of PDE. Used often in computation fluid dynamics.
- Test case incompressible airflow around motorcycle
- Application computational characteristics:
  - Unstructured grid

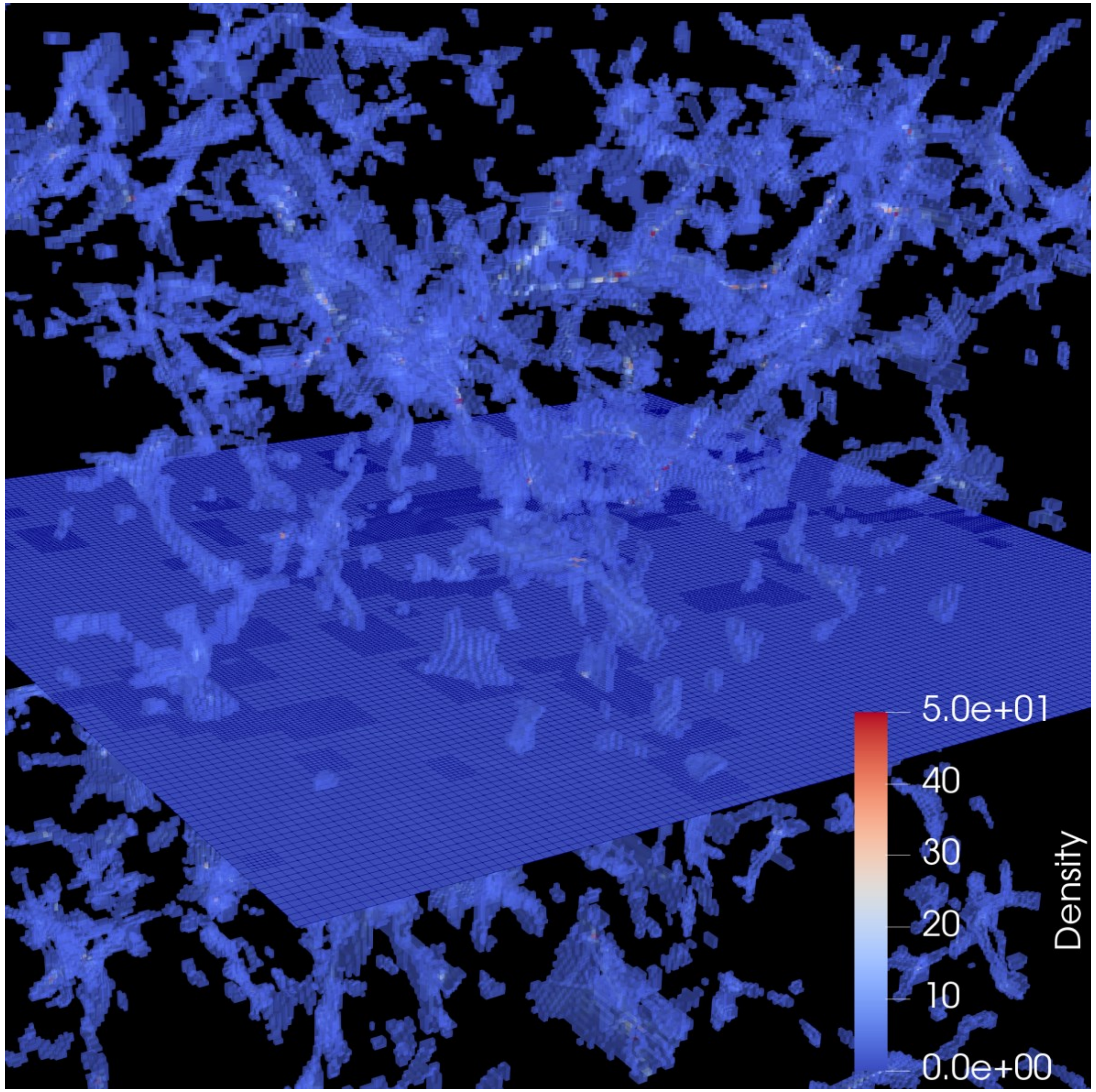
OpenFOAM®

OpenFOAM

CPU/System	Cores	Wall Clock Time, Minutes	Meshing Time, Minutes	Solver Time, Minutes	Power, W	Energy Efficiency, Jobs per kWh	N
ARM Fujitsu A64FX, SVE 512b (SBU-Ookami, GCC)	48	28.4 ± 0.9	14.6 ± 0.9	12.4 ± 0.1	110 ± 7	19.3 ± 1.6	21
ARM Fujitsu A64FX, SVE 512b (SBU-Ookami, FJ)	48	22.4 ± 0.3	8.5 ± 0.1	10.9 ± 0.2	111 ± 7	24.1 ± 1.6	21
ARM Amazon Graviton 2, Neoverse N1 (AWS)	48	11.9 ± 0.3	3.5 ± 0.2	8 ± 0.1			10
ARM Amazon Graviton 3, Neoverse V1, SVE 256b (AWS)	48	7.1 ± 0.2	2.2 ± 0.2	4.7 ± 0.0			5
ARM Amazon Graviton 3, Neoverse V1, SVE 256b (AWS)	64	6.8 ± 0.1	2.2 ± 0.1	4.4 ± 0.1			20
ARM Ampere Altra, Neoverse N1 (Azure)	48	11.1 ± 0.2	3.2 ± 0.2	7.6 ± 0.1			10
ARM Ampere Altra, Neoverse N1 (Azure)	64	10.9 ± 0.4	3.2 ± 0.2	7.2 ± 0.2	270*	20.4	20
x86 AMD EPYC 7742 Zen2(Rome), AVX2 (SDSC Expanse)	128	9.5 ± 1.9	5.6 ± 1.4	3.2 ± 1.1			20
x86 AMD EPYC 7763 Zen3(Milan), AVX2 (Purdue Anvil)	128	6.6 ± 0.2	3.1 ± 0.5	2.9 ± 0.5			19
x86 Intel Xeon Plat. 8160, Skylake-X, AVX512 (TACC-Stampede 2)	48	10.7 ± 0.4	3.7 ± 0.3	6.4 ± 0.1			10
x86 Intel Xeon Plat. 8380, Ice Lake, AVX512 (TACC-Stampede 2)	80	6.8 ± 0.3	2.6 ± 0.2	3.7 ± 0.3			20
x86 Intel Xeon Gold 6130, Skylake-X, AVX512 (UB-HPC)	32	13.2 ± 0.8	4.1 ± 0.4	7.7 ± 0.1	375 ± 35	12.3 ± 1.0	23
x86 Intel Xeon Gold 6330, Ice Lake, AVX512 (UB-HPC)	56	8.9 ± 0.5	2.8 ± 0.3	4.7 ± 0.2	505 ± 34	13.4 ± 0.9	20

- Graviton 3 is fastest in ARM Camp, comparable to modern x86 and faster than older generations
- Mashing is fastest in Graviton 3
- OpenFOAM doesn’t scale well within node. Possibly saturated memory bandwidth earlier.

Enzo: Astrophysics and Cosmology



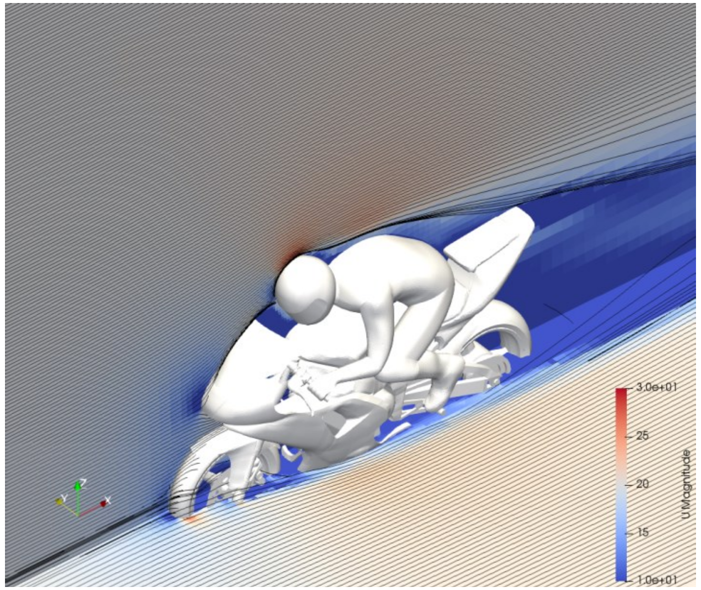
- Enzo is an Adaptive Mesh Refinement Code for Astrophysics
- Application computational characteristics:
  - Unstructured grid
  - Adaptive Mesh Refinement
- Test case: cosmology simulation that simulates reionization using the ray tracing radiation transfer method with radiating star particles and a Haardt & Madau background.

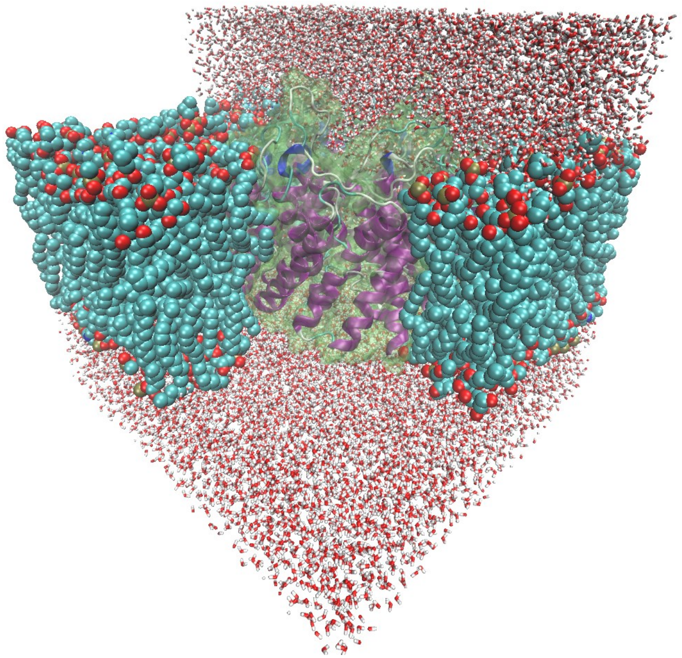
Enzo: Astrophysics and Cosmology

CPU/System	Cores	Wall Clock Time, Minutes	Power, W	Energy Efficiency, Jobs per kWh	N
ARM Fujitsu A64FX, SVE 512b (SBU-Ookami, GCC)	48	115.7 ± 17.7	112 ± 7	4.7 ± 0.5	10
ARM Amazon Graviton 2, Neoverse N1 (AWS)	48	23.6 ± 1.1			12
ARM Amazon Graviton 3, Neoverse V1, SVE 256b (AWS)	48	17 ± 1.2			11
ARM Amazon Graviton 3, Neoverse V1, SVE 256b (AWS)	64	13.2 ± 0.7			20
ARM Ampere Altra, Neoverse N1 (Azure)	48	21 ± 1.0			11
ARM Ampere Altra, Neoverse N1 (Azure)	64	15.9 ± 0.8	270*	14	20
x86 AMD EPYC 7742 Zen2(Rome), AVX2 (PSC-Bridges-2)	128	7.1 ± 0.4			20
x86 AMD EPYC 7742 Zen2(Rome), AVX2 (SDSC Expanse)	128	6.6 ± 0.4			20
x86 AMD EPYC 7763 Zen3(Milan), AVX2 (Purdue Anvil)	128	6.9 ± 0.3			20
x86 Intel Xeon Phi 7250, KNL, AVX512 (TACC-Stampede 2)	68	14.7 ± 0.3			20
x86 Intel Xeon Plat. 8160, Skylake-X, AVX512 (TACC-Stampede 2)	48	4.2 ± 0.1			20
x86 Intel Xeon Gold 6130, Skylake-X, AVX512 (UB-HPC)	32	4.8 ± 0.3	338 ± 31	37.1 ± 3.9	50
x86 Intel Xeon Gold 6130, Skylake-X, AVX512 (UB-HPC)	32	25.8 ± 0.9	379 ± 26	6.2 ± 0.4	11
x86 Intel Xeon Gold 6330, Ice Lake, AVX512 (UB-HPC)	56	15.5 ± 0.6	559 ± 34	6.9 ± 0.4	11

- Failed to compile enzo on intel Ice Lake, Skylake binary also failed.
- Graviton 3 is fastest in ARM Camp, still significantly slower than x86 counterparts (2.5 to 4.0 times)

Conclusions





- **The building and compiling experience on ARM platforms is very similar to that of traditional HPC systems**
- As tested by HPCC benchmark, numerical libraries implementing linear algebra and FFT routines support ARM CPUs well and the latter exhibit a solid performance.
- ARM machines shows good performance in Gromacs, OpenFOAM, Tensorflow and NWChem applications.
- The ARM performance is comparable to x86 counterparts, and they often outperform previous generations of x86 CPUs (largely Amazon Graviton3).
- In ENZO, Amazon Graviton3 and Ampere Altra are within the x86 systems performance
- Fujitsu A64FX and Ampere Altra are more energy efficient in GROMACS, NWChem and OpenFOAM than x86 CPUs.
- In cases where GPU performance was tested, the GPU systems showed the fastest speed and the highest energy efficiency.
- Compiler and libraries can greatly affect the simulation speed and energy efficiency.
- Intel Skylake-X is a very robust architecture for scientific calculations, more than five years later since the initial release, it still competes well with modern CPUs.

- **From the performance, energy efficiency and software building point of view as of now for all tested applications, modern ARM CPUs provide a viable alternative to their x86 counterparts and not only as a cheaper option for the GPU gateway (look at one with SVE support for future support).**

Acknowledgements

NSF OAC Awards: 1927880 and 2137603

- This work is supported by the National Science Foundation under awards OAC 1927880 and 2137603.
- This work used compute resources at Stony Brook University, SUNY UB CCR, the XSEDE/ACCESS (CCR120014) and CloudBank .

UB Slurm Simulator

University at Buffalo Center for Computational Research

Stony Brook University

iACS Institute for Advanced Computational Science