Early results from Isambard 3, one of the first NVIDIA Grace CPU-based systems

Simon McIntosh-Smith, Thomas Green

Bristol Centre for Supercomputing (BriCS)

University of BRISTOL

13th June 2025
Arm HPC User Group Workshop @ ISC25

BriCS
Bristol Centre for Supercomputing

Brief history

- **2018** – First Isambard system in production with Arm-based ThunderX2 processors
- **2020** – Isambard 2 launched with increased capacity of ThunderX2, added a 72-node A64FX cluster
  - Hosted at the **Met Office** in Exeter, UK
  - **328** nodes or 20,992 ThunderX2 cores (featured 4-way SMT)
- **Sep 2024** – Isambard 2 decommissioned
  - Funded by **EPSRC** (UK research agency)
  - In collaboration with **GW4** universities

University of BRISTOL

BriCS
Bristol Centre for Supercomputing

What is Isambard 3?

- A new general purpose air-cooled CPU HPC machine, ~300kW
- Based on **NVIDIA Grace CPU Superchip**
- Delivered by HPE
- 384 nodes, 55,236 cores
- 2 PBytes storage, Slingshot 11 network
- In production since Jan 25
- Funded by UKRI
- Collaboration with GW4 universities
- Hosted at Bristol (alongside **Isambard-AI**)

UK Research and Innovation

GW4

University of BRISTOL

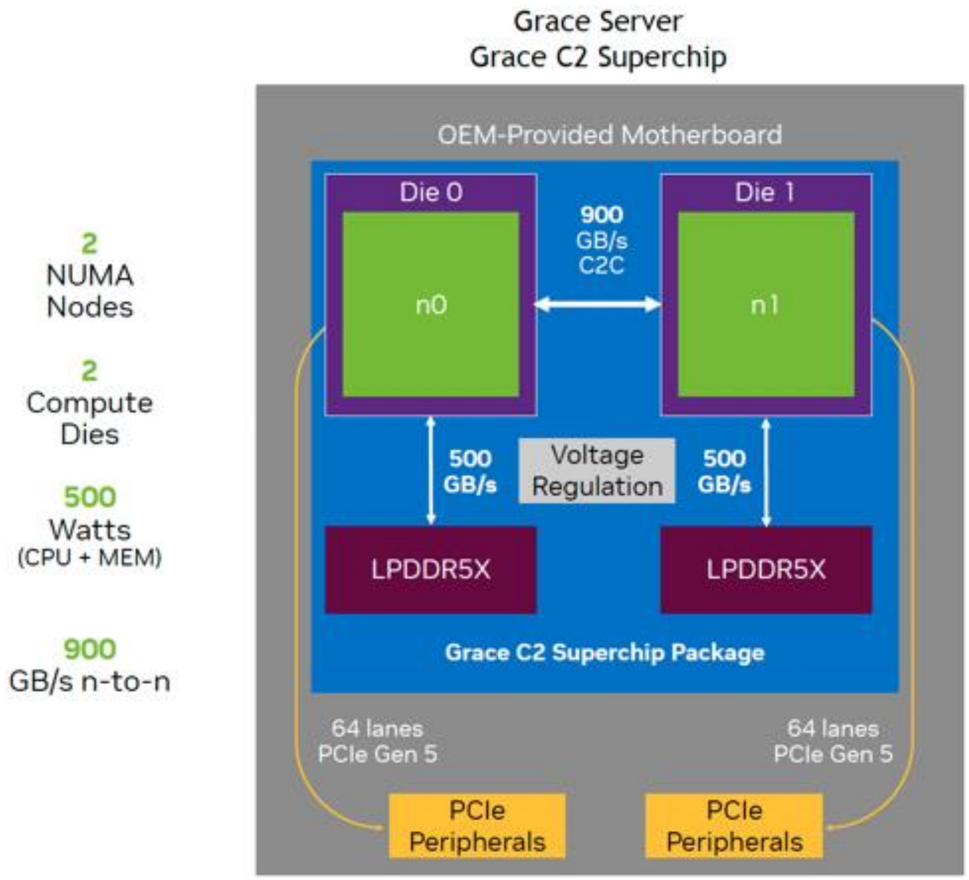

05-Dec-23 12:45


Isambard

Isambard 3 technical summary

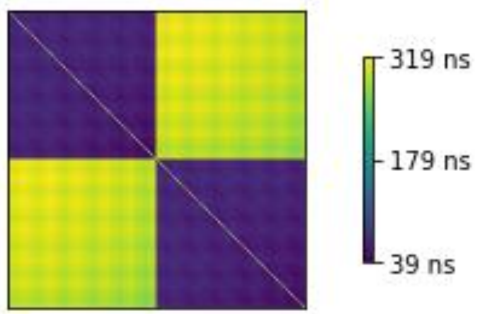
- **55,296** Armv9 cores, **384** nodes, 6 racks of NVIDIA Grace CPU Superchip, 2x72 core @ 3.1GHz (upto 3.3GHz)
  - ~2 PFLOP/s HPL, just outside the Top500 in Nov 2024
- **Slingshot 11** dragonfly network @ 200Gbps
- 2.0 PiByte of Cray/HPE ClusterStor E1000 **Lustre** storage
  - Mix of NVME/HDD, ~50GB/s
- Multi Architecture Comparison System (**MACS**)
  - AMD Milan, Genoa, Bergamo, Intel Sapphire Rapids + HBM, AMD/Nvidia GPUs…
  - Limited number of nodes
- In a new, dedicated **MDC** equivalent size of 15 racks total (also hosts Isambard-AI phase 1)

NVIDIA Grace Superchip



Source: NVIDIA

- Upgrade route to previous Arm-based Isambard 2 (**ThunderX2**)
- Strength in its memory bandwidth
- Use of LPDDR5X makes power of CPU+MEM efficient
- C2C low latency
- 4x 128b SVE2 per core



Available systems within Isambard 3

- Shared infrastructure
  - 2 PB ClusterStor (Lustre)
  - Slingshot **FMN 2.2**
  - Slingshot **SHS 11.1** (**11.0** on MACS)
  - SLES15sp5
  - Cray Programming Environment
  - HPCM 1.11

Processor	#	Mem [GB]	Cores	Base Clock Speed [GHz]	FP64 peak [TFLOP/s]	Default TDP [W]	Bandwidth [GB/s]
NVIDIA Grace CPU Superchip	384	240	2 x 72	3.1	7.1	1 x 500 (including memory)	1024.0
AMD EPYC 7713 (Milan)	12	256	2 x 64	2.0	4.0	2 x 225	409.6
AMD EPYC 9354 (Genoa)	2	384	2 x 32	3.25	3.3	2 x 280	921.6
AMD EPYC 9754 (Bergamo)	2	192	1 x 128	2.25	4.6	1 x 360	460.8
Intel Xeon Gold 6430 (Sapphire Rapids)	2	256	2 x 32	2.1	4.3	2 x 270	614.4
Intel Xeon CPU Max 9462 (Sapphire Rapids)	2	120	2 x 32	2.7	5.5	2 x 350	3276.8

https://docs.isambard.ac.uk/specs/#system-specifications-isambard-3-macs

Benchmarks*

Question from researchers

“How does X perform on Isambard compared to system Y?”

- Synthetic
  - **STREAM** [link]
  - **Arm-kernels** [link]
  - **CloverLeaf** [link]
  - **TeaLeaf** [link]
  - **SNAP** [link]
  - **Neutral** [link]
  - **OSU Micro-benchmarks** [link]
- Applications
  - **CASTEP** [link]
  - **CP2K** [link]
  - **GROMACS** [link]
  - **NAMD** [link]
  - **OpenFOAM** [link]

Based on previous studies on Isambard 2

* N.B. focus on effortless **science**!

University of BRISTOL

BriCS Bristol Centre for Supercomputing

Method to run benchmarks

Reframe

- Previous **scripting** method required changes to support Isambard 3
- Reframe provided method to run across **all clusters**
- Supported **Spack** as the self-service approach

Spack

- Supported **packages** being explored
- Provided mechanism to try **different** compilers
- Experienced issues with compilers except with GCC
  - Documented HPE approach results in CCE not in archspec
  - Mixing compilers has issues in 0.23.1

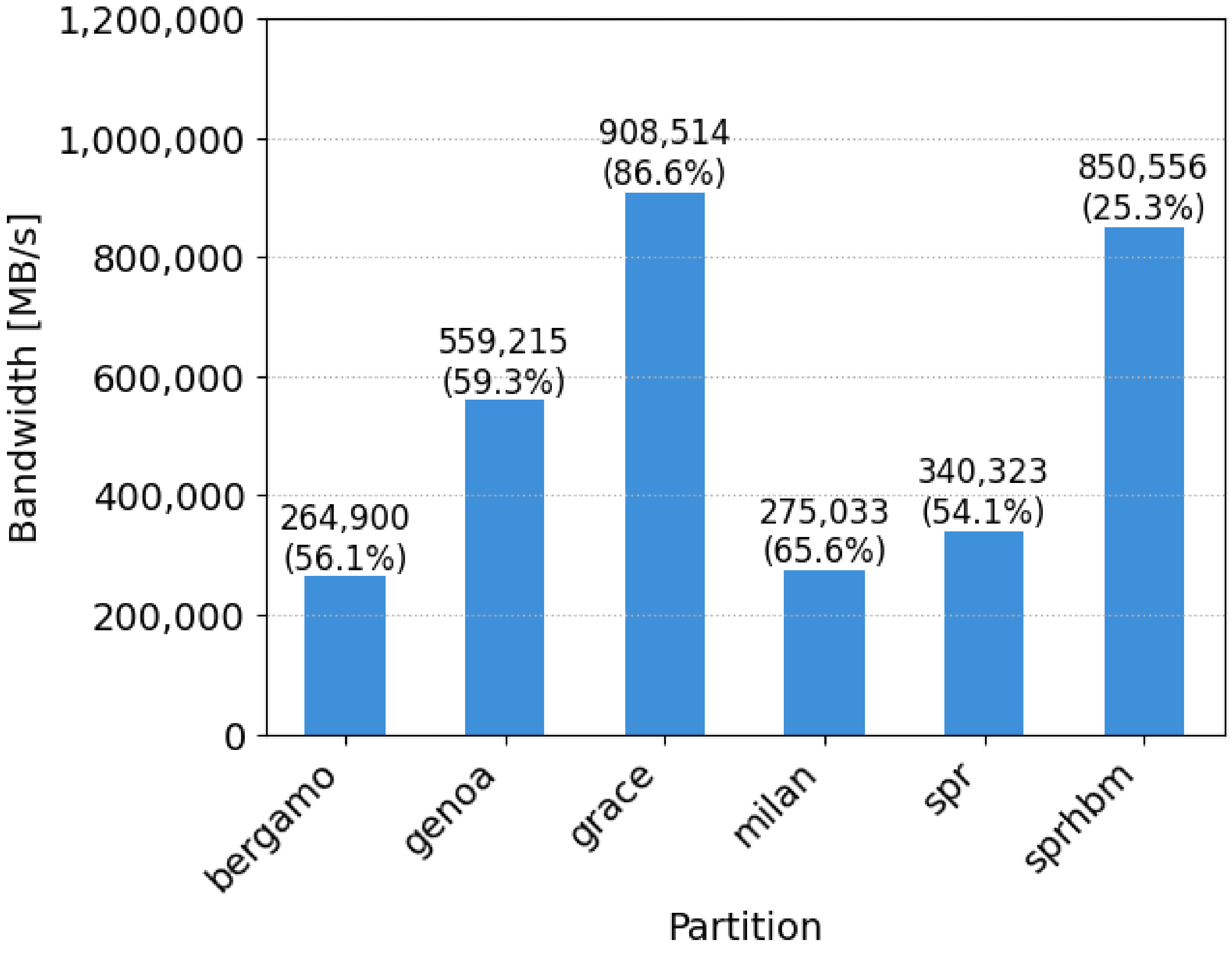
Configuration available at: https://github.com/isambard-sc/buildit

University of BRISTOL

BriCS Bristol Centre for Supercomputing

STREAM

- This synthetic benchmark provides a measure of bandwidth in MB/s
- Grace and Sapphire Rapids HBM are the expected top performers with their memory design
- Compiler choice influences results



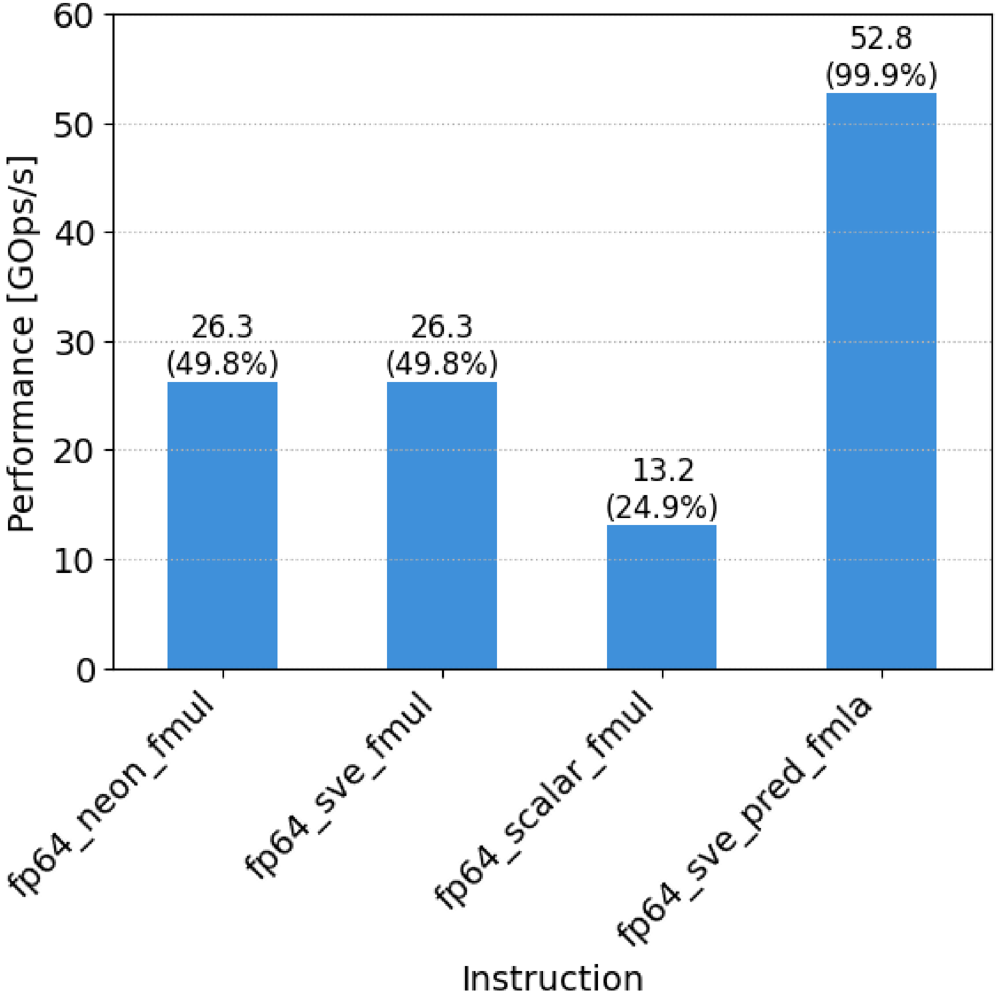
System: Isambard 3
Build: GCC 12.3, OpenMP,
Source: core Spack package, 5.10

University of BRISTOL

BriCS
Bristol Centre for Supercomputing

Arm-kernels

- Measures raw operations for each instruction
- Measure for each core (averaged across all cores)
- Only currently works with GCC, will be fixed in future Arm compiler (Clang based)
- Maximum at 3.3GHz with 2 FMA operations per cycle and 4x128bits SVE is 52.8 Gops/s at fp64



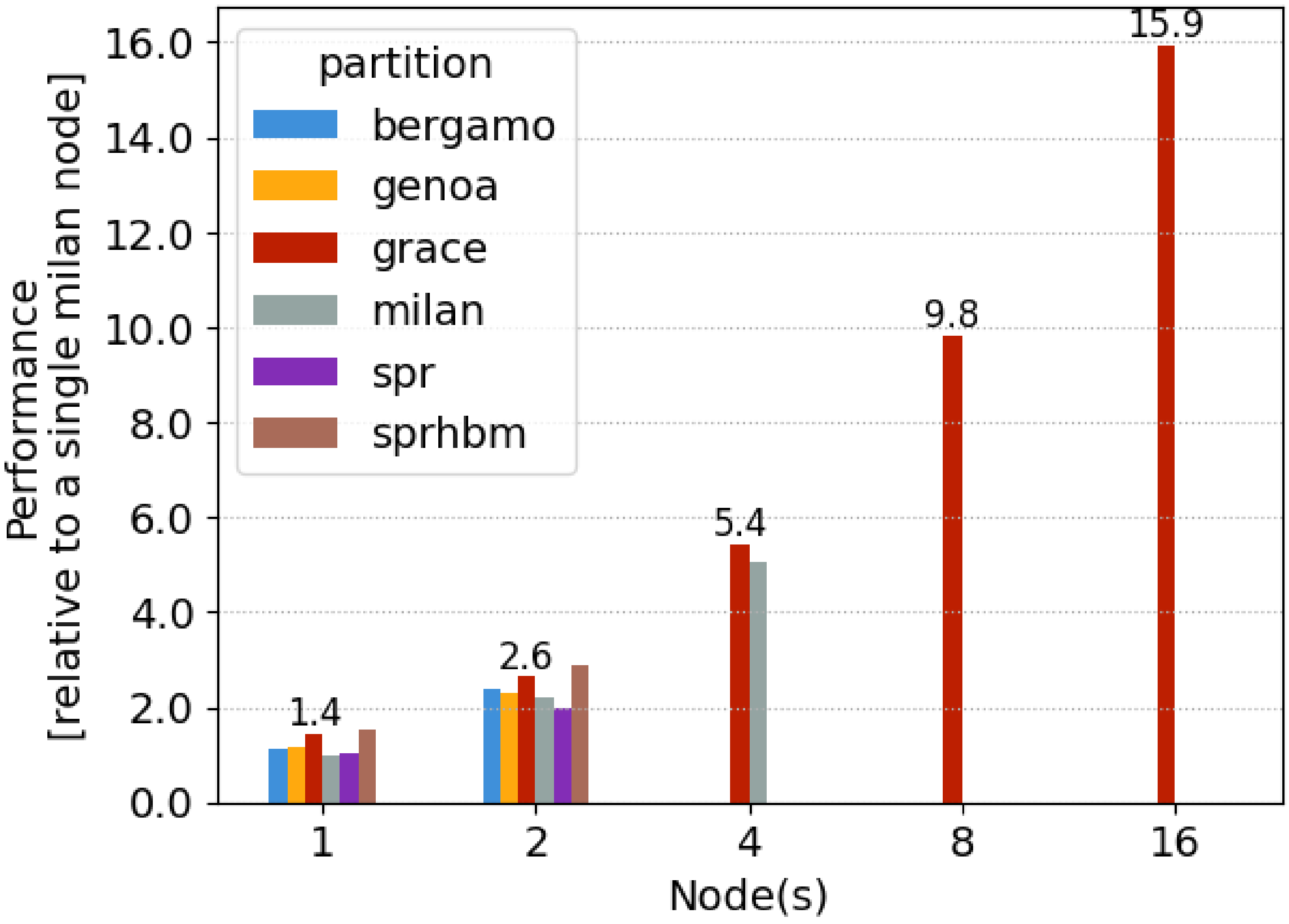
System: Isambard 3
Build: GCC 12.3
Source: core Spack package, d295e1f

University of BRISTOL

BriCS Bristol Centre for Supercomputing

CloverLeaf

- A mini-app that solves Euler's equations of compressible fluid dynamics
- **Memory-bandwidth** bound
- Also tests interconnect
- Both Grace and Sapphire Rapids HBM are the expected best performers



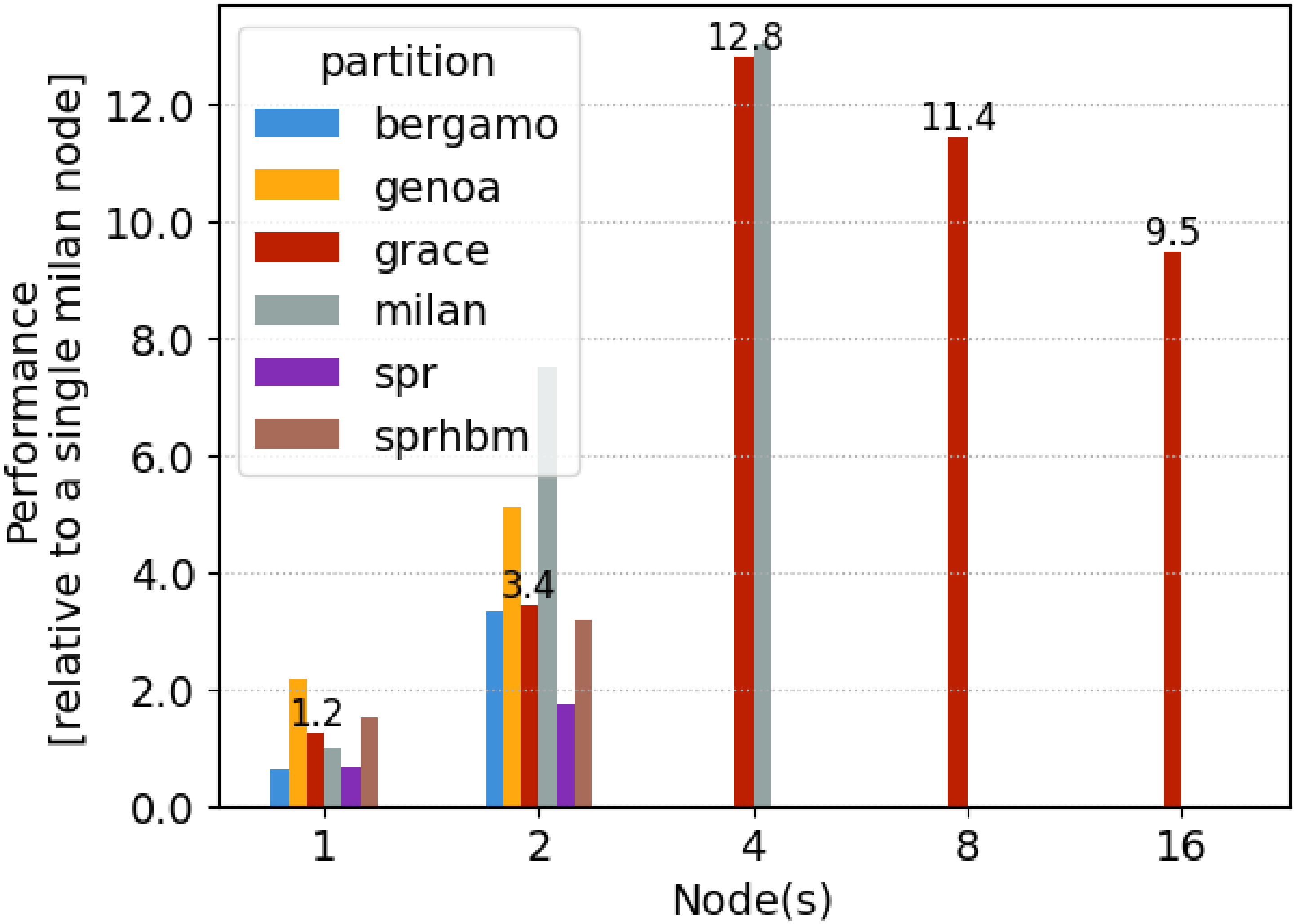
System: Isambard 3
Build: GCC 12.3
Source: created Spack package, 0fdb917

University of BRISTOL

BriCS Bristol Centre for Supercomputing

TeaLeaf

- A mini-app that solves linear heat conduction equation
- **Memory-bandwidth** bound
- Due to size of problem the super-scaling results are due to taking advantage of last-level caches
- Expected to be **communication bound** at higher node counts



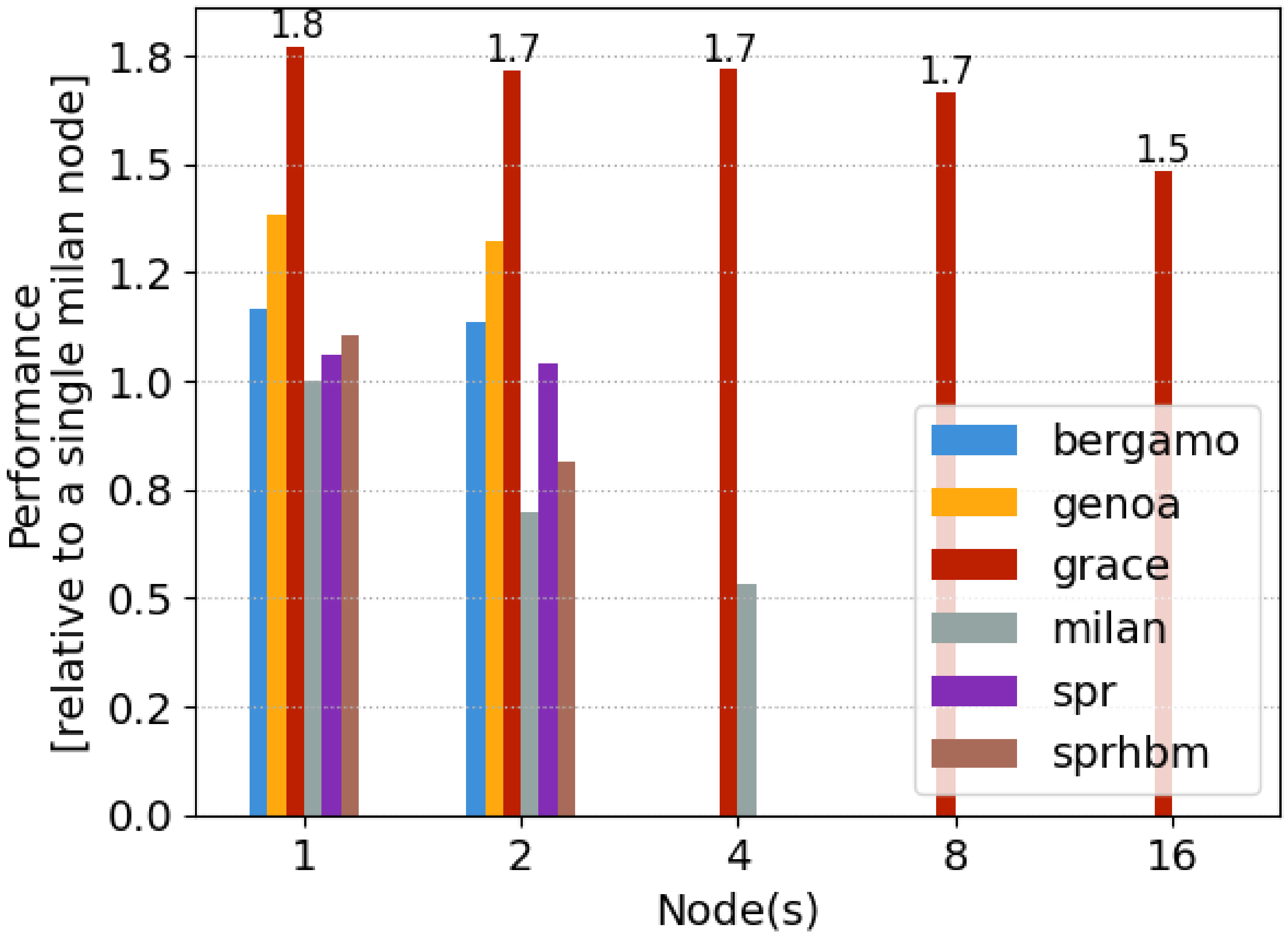
System: Isambard 3
Build: GCC 12.3
Source: core Spack package, 019aa2f

University of BRISTOL

BriCS Bristol Centre for Supercomputing

SNAP – uob-hpc

- The SNAP mini-app solves the linear radiation pseudo-transport problem on a structured mesh
- Influenced by **cache behaviour**
- The weak scaling behaviour provides confidence Slingshot performing well for Grace
- Some show reduction in scaling even at modest core counts that is worth exploring further (e.g. milan nodes)



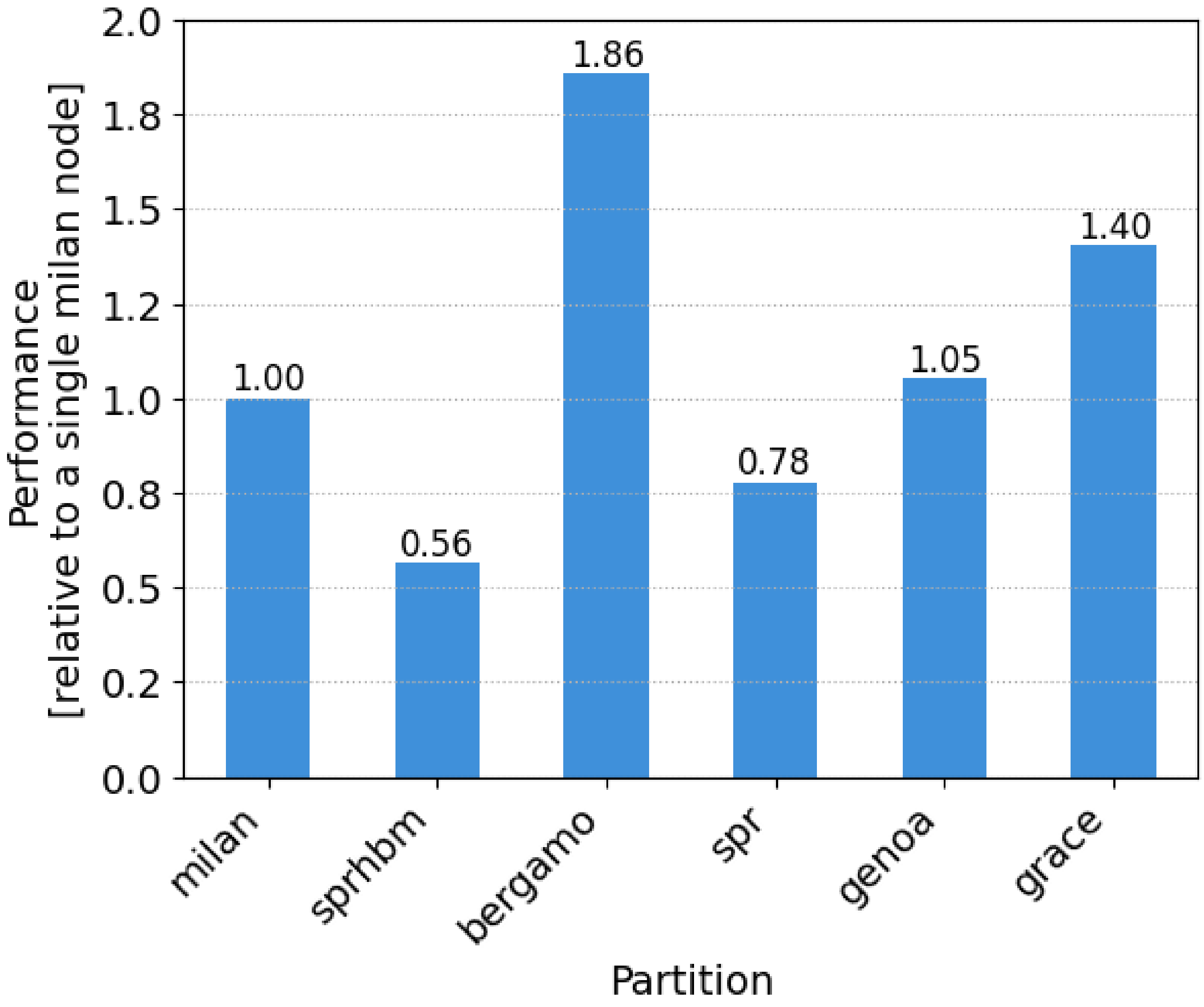
System: Isambard 3
Build: GCC 12.3, cray-mpich 8.1.30, OpenMP
Source: created Spack package, e7ab43d

University of BRISTOL

BriCS Bristol Centre for Supercomputing

Neutral

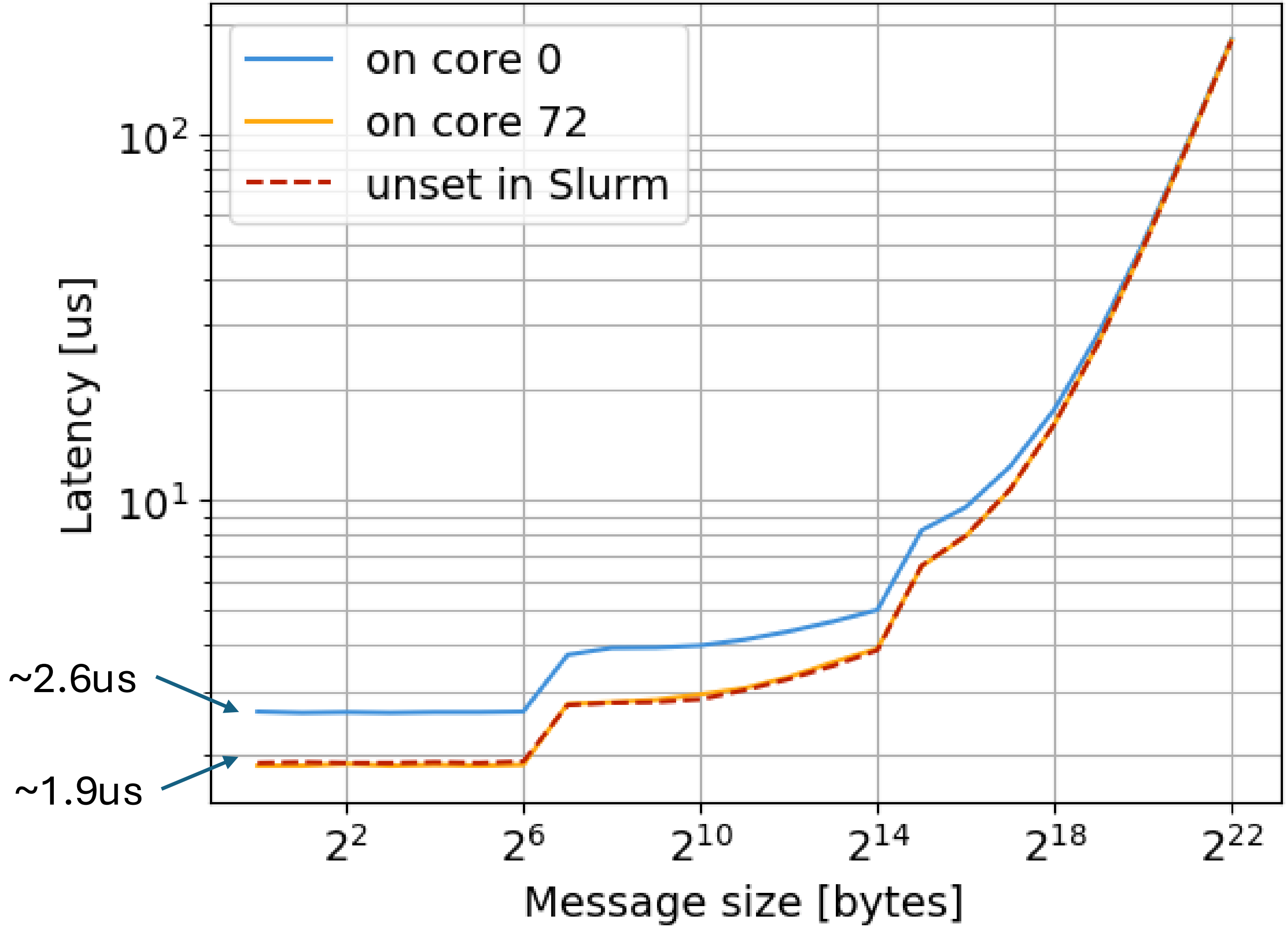
- Neutral is a Monte Carlo neutron transport mini-app (single node)
- Influenced by **cache behaviour**
- Random memory access benefits the Bergamo configuration
- Grace performs well within the range of other configurations



System: Isambard 3
Build: GCC 12.3, OpenMP
Source: created Spack package, d983598

OSU Microbenchmarks

- **Latency** within Grace clearly depends on MPI task placement
- Slingshot card connected to 2nd "socket"
- Default behaviour performs sensibly, with MPI being placed near Slingshot



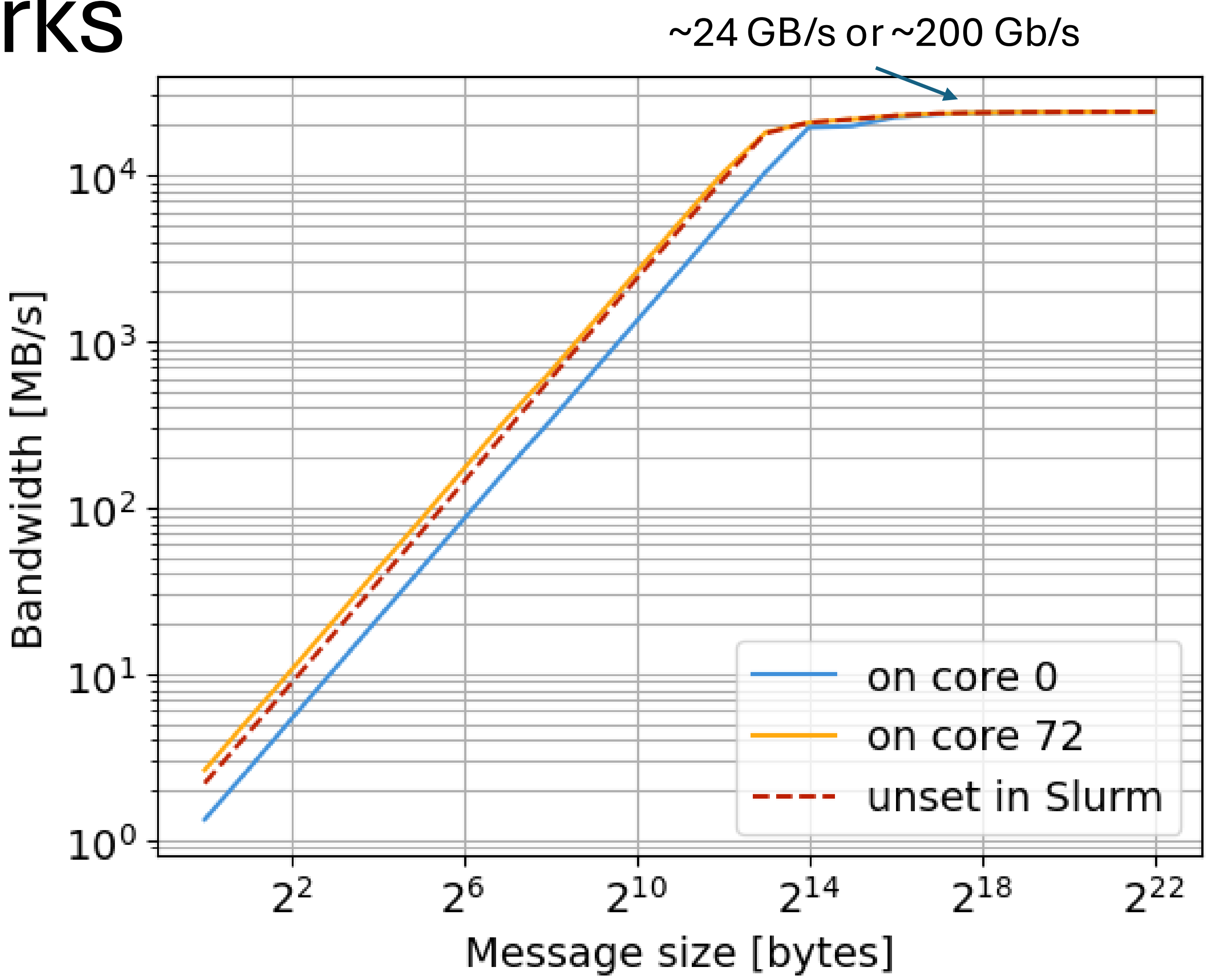
System: Isambard 3 (2 nodes)
Build: GCC 12.3 , cray-mpich 8.1.30
Source: created Spack package, 7.5

University of BRISTOL

BriCS Bristol Centre for Supercomputing

OSU Microbenchmarks

- **Bandwidth** penalty when not near the Slingshot card
- Slingshot card connected to 2nd "socket"
- Default behaviour performs sensibly, with MPI being placed near Slingshot



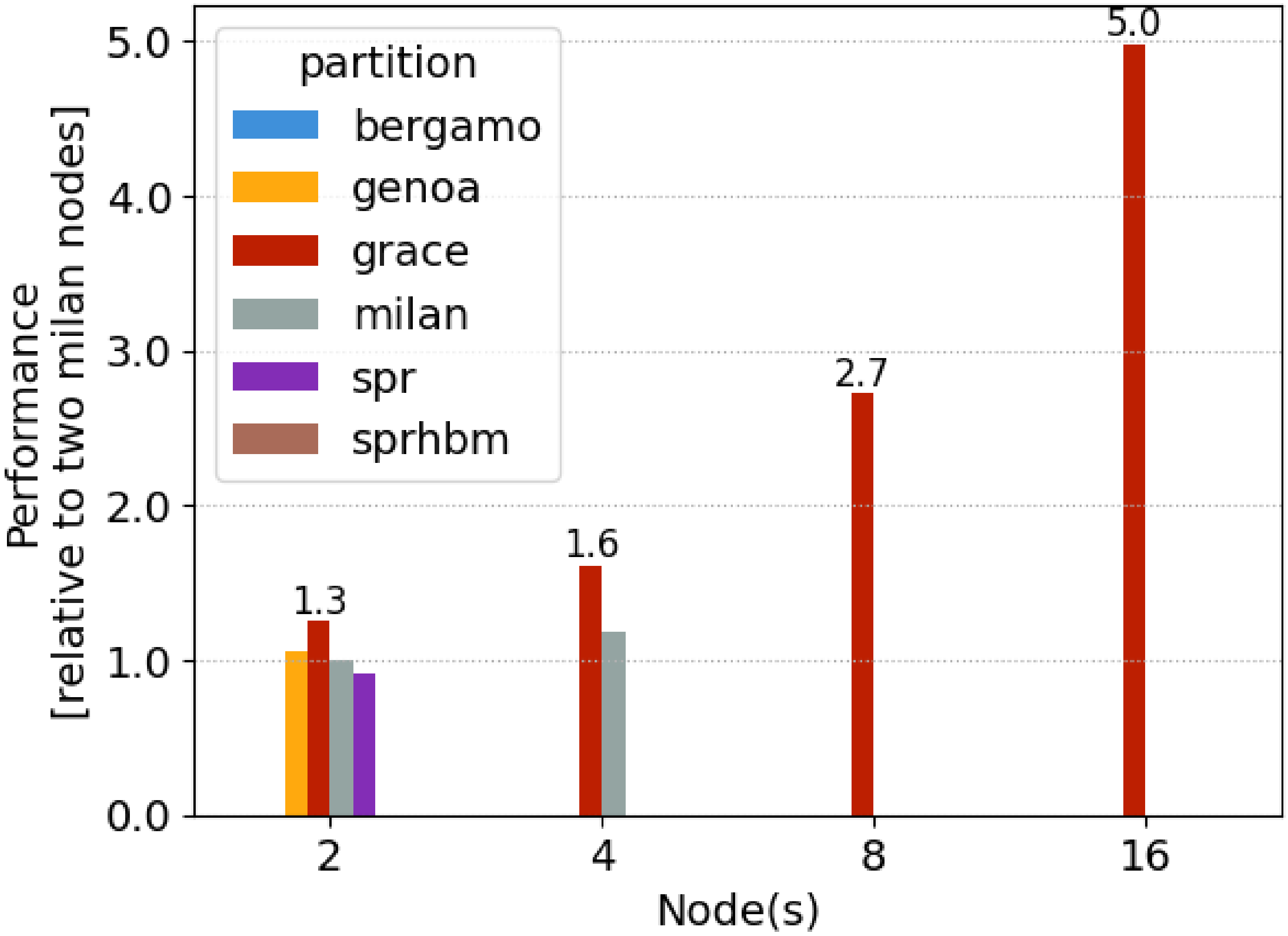
System: Isambard 3 Grace (2 nodes)
Build: GCC 12.3, cray-mpich 8.1.30
Source: core Spack package, 7.5

University of BRISTOL

BriCS
Bristol Centre for Supercomputing

CASTEP - crambin

- Popular code used to calculate material properties from first principles
- Mainly **memory bandwidth bound**
- Memory requirements resulted in minimum of 2 nodes
- For Bergamo and Sapphire Rapids HBM there was not enough memory available
- Strong scaling looks good



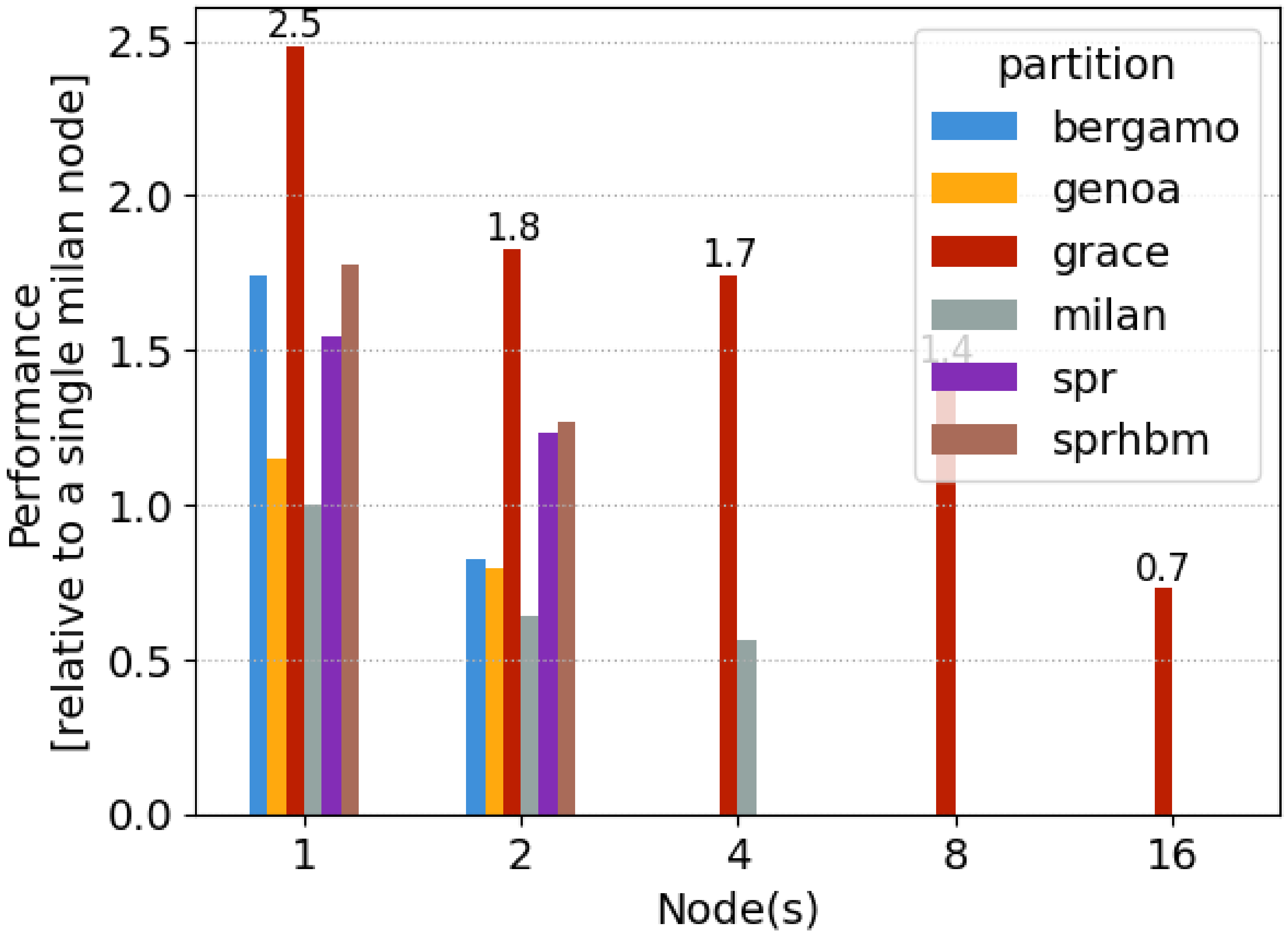
System: Isambard 3
Build: GCC 12.3, cray-mpich 8.1.30
Source: created Spack package, 25.1.1

University of BRISTOL

BriCS
Bristol Centre for Supercomputing

CP2K – H2O-64

- The CP2K code simulates the Ab-initio electronic structure and molecular dynamics of different systems
- No single factor seen to influence behaviour
- Due to size of problem it provides limited scaling results
- Single node result clearly shows Grace performing well



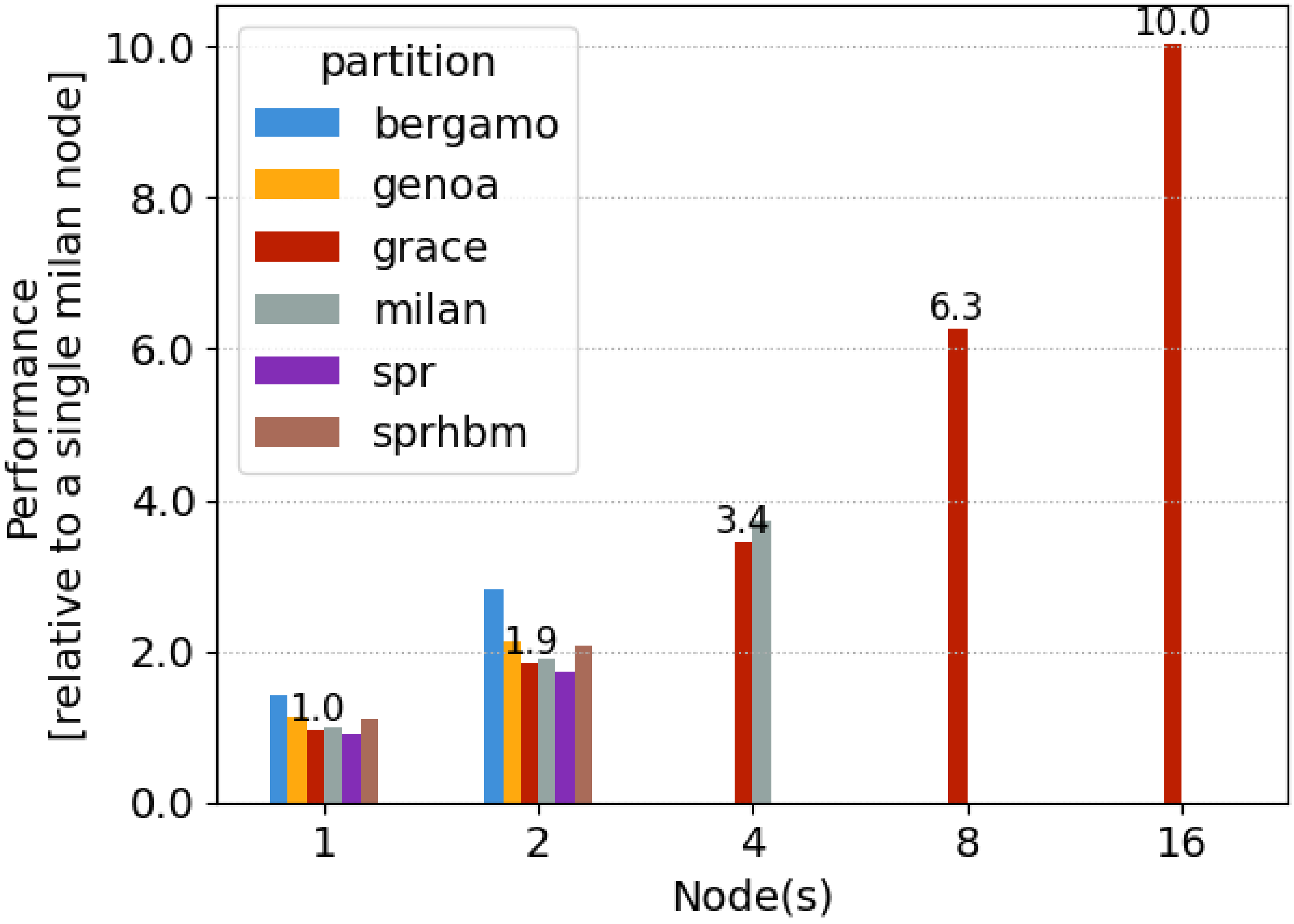
System: Isambard 3
Build: GCC 12.3, cray-mpich 8.1.30
Source: core Spack package, 2024.3

University of BRISTOL

BriCS
Bristol Centre for Supercomputing

Gromacs - TestCaseB

- A molecular dynamics package that solves Newton's equations of motion
- Considered a **compute bound** problem
- Grace performs reasonably relative to the other platforms



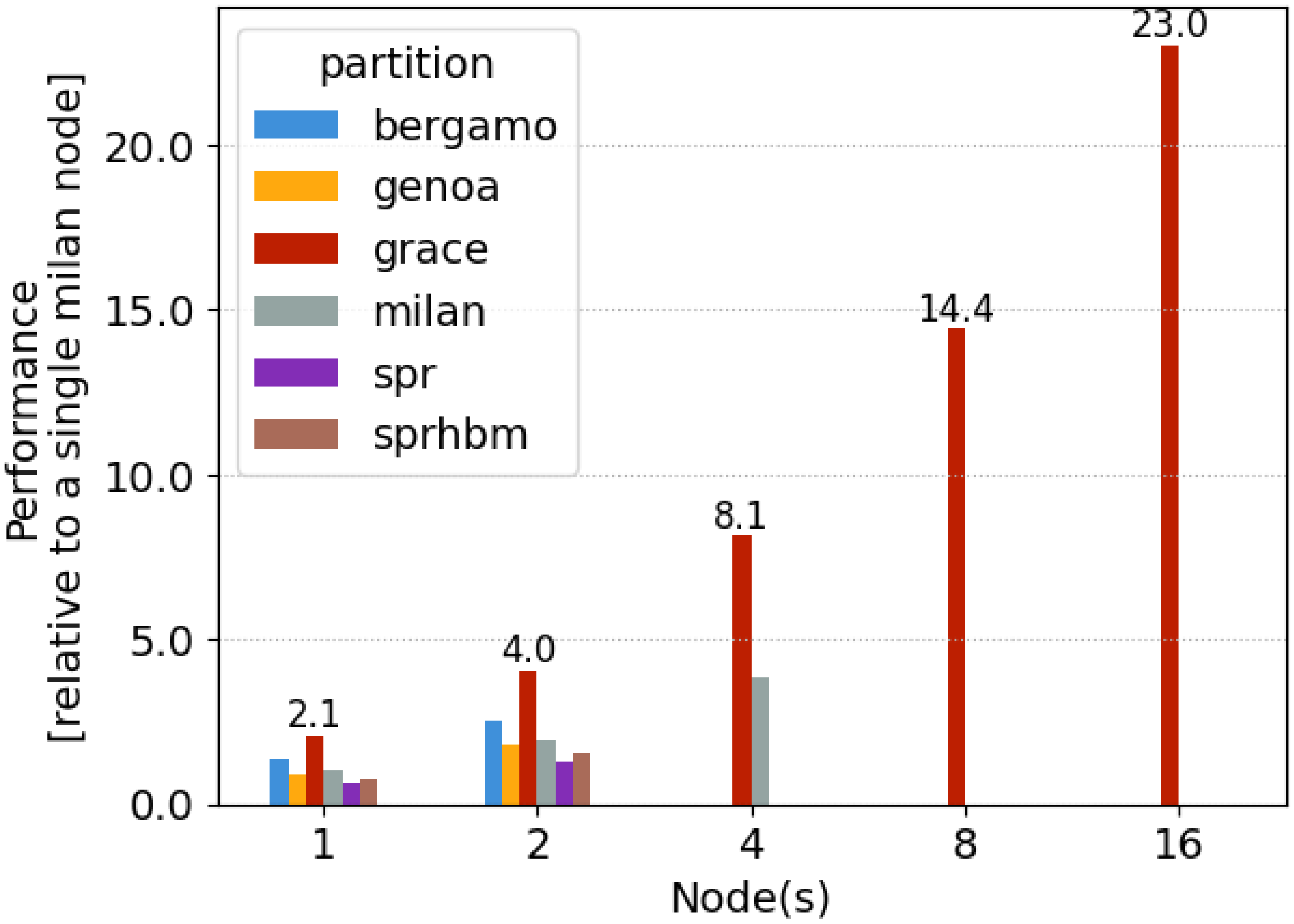
System: Isambard 3
Build: GCC 12.3, , cray-mpich 8.1.30
Source: core Spack package, 2024.3

University of BRISTOL

BriCS
Bristol Centre for Supercomputing

NAMD - STMV

- A molecular dynamics simulation program designed to scale up to millions of atoms
- **No single bound** and Grace performs well for this workload
- Strong scales well



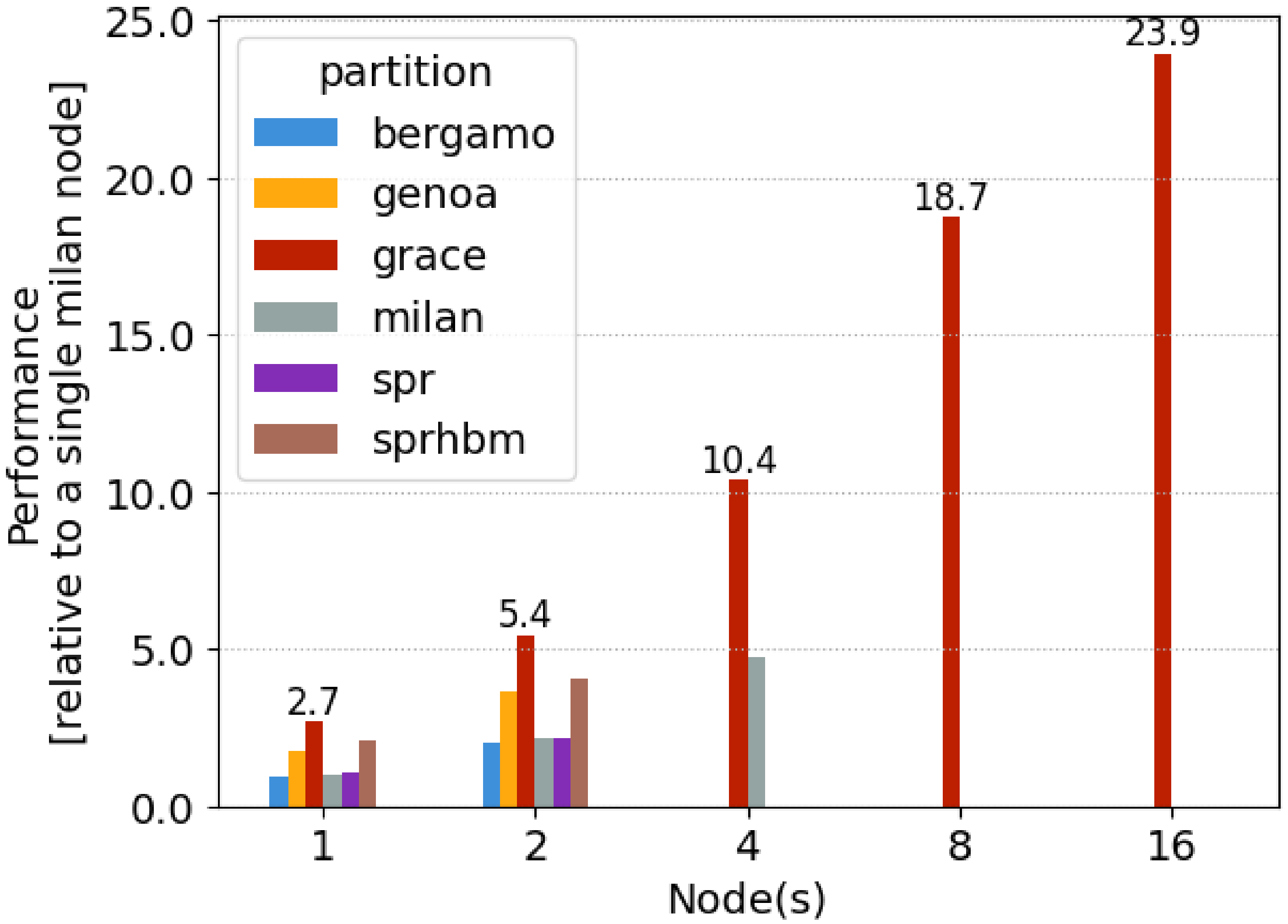
System: Isambard 3
Build: GCC 12.3, , cray-mpich 8.1.30
Source: created Spack package, 3.0

University of BRISTOL

BriCS
Bristol Centre for Supercomputing

OpenFOAM – HPC Motorbike

- OpenFOAM is a modular C++ framework aiming to simplify writing custom computational fluid dynamics (CFD) solvers
- A **memory bandwidth bound** code, should perform well on Grace and Sapphire Rapids HBM
- OpenFoam performs well in strong scaling as expected



System: Isambard 3
Build: GCC 12.3, , cray-mpich 8.1.30
Source: core Spack package, 2312

University of BRISTOL

BriCS
Bristol Centre for Supercomputing

Early user feedback

"On the software side, I was generally pleased with the Spack setup."

"[Firedrake] its up and running, its fast and considering this is just some initial testing we get decent parallelization."

"Our best-case performance result was a 4-node run on Isambard 3 which was 28% faster than the equivalent run on 4 nodes of [system based on 2 x AMD 7H12 processors]."

A supercomputer for scientific applications!

- NVIDIA Grace Superchip
  - Performs well across wide **range of software**
  - **Competitive** against a range of other CPUs
  - **Memory bandwidth** clear advantage (e.g. OpenFOAM)
- System design
  - Well suited but MPI placement requires careful consideration due to Grace C2C
- Suitability of **self-service approach** with Spack
  - Worked well with GCC
  - Other compilers need further work with packages
  - Cray Compiler requires consideration of compiler flags from archspec used in Spack
- Early user feedback
  - Positive experience
  - Fix signed vs unsigned char

Future plans

- Welcome **further projects** onto Isambard 3
- Move to the soon to be released **Spack 1.0** where compiler configuration has major changes
- Further investigate **Grace socket MPI** behaviour
- Continue to improve the **build configuration** for applications

Acknowledgements

The authors acknowledge the use of resources provided by the Isambard 3 Tier-2 HPC Facility.

Isambard 3 is hosted by the University of Bristol and operated by the GW4 Alliance (https://gw4.ac.uk) and is funded by UK Research and Innovation; and the Engineering and Physical Sciences Research Council [EP/X039137/1].

brics-enquiries@bristol.ac.uk

https://www.bristol.ac.uk/supercomputing/

University of BRISTOL

BriCS
Bristol Centre for Supercomputing