INVESTIGATING A MULTI-SOCKET HIGH CORE-COUNT RISC-V SYSTEM FOR HPC WORKLOADS

Nick Brown, EPCC

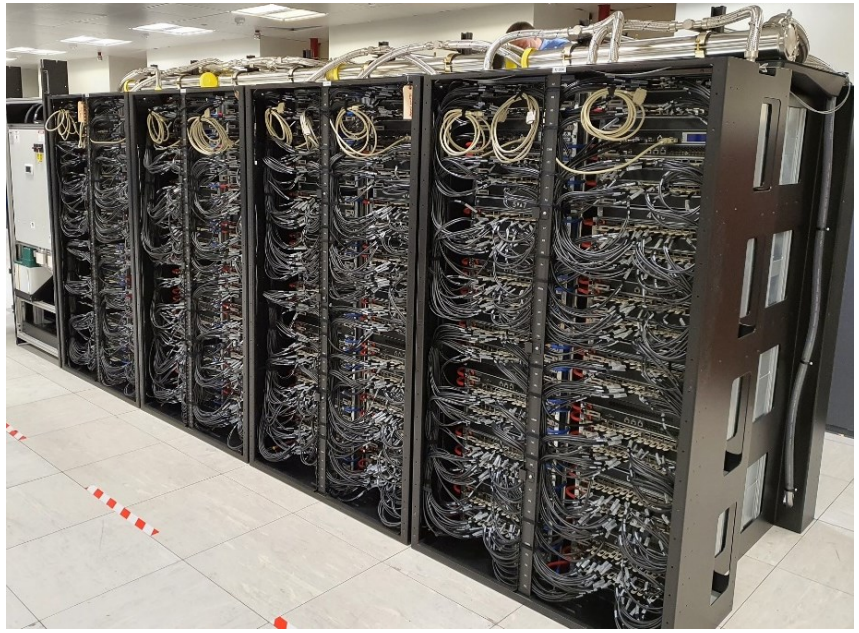
n.brown@epcc.ed.ac.uk

epcc

THE UNIVERSITY OF EDINBURGH


archer2


CE™ XA



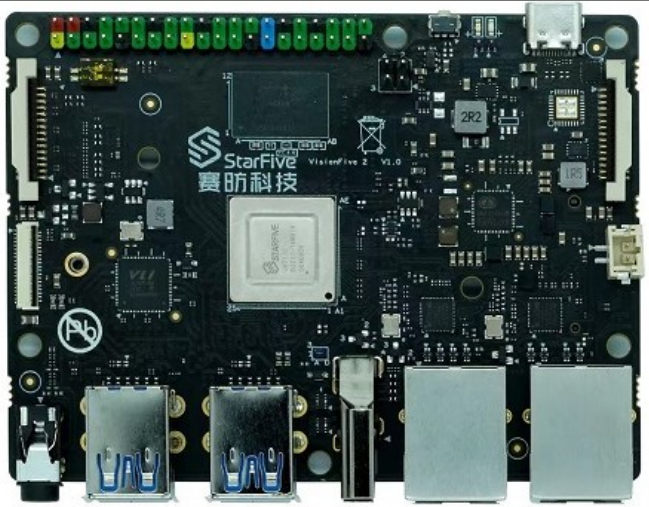
RISC-V testbed

- ExCALIBUR is a UK exascale programme
  - One key question is around what are the next generation hardware technologies that might be in our supercomputers in the next five to ten years
    - And how do we best support these for users?

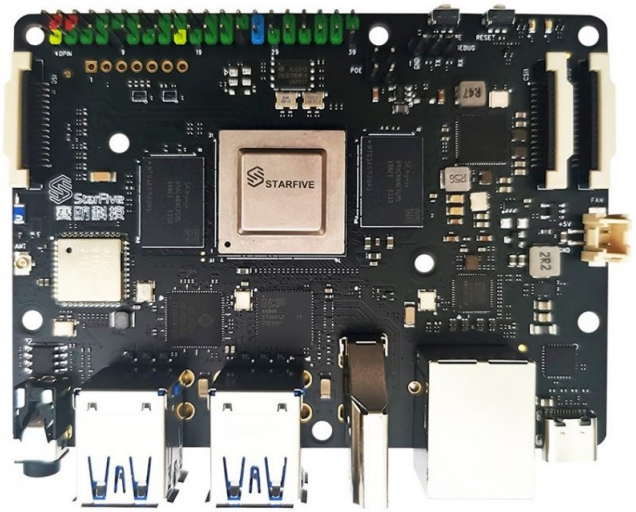
- RISC-V testbed aims to make this technology available to scientific software developers more widely to experiment with RISC-V
  - Idea is that it should feel like any other HPC style system with login node, compute managed by Slurm and common HPC libraries/compilers as modules, user accounts managed by SAFE
    - riscv.epcc.ed.ac.uk

epcc | THE UNIVERSITY OF EDINBURGH

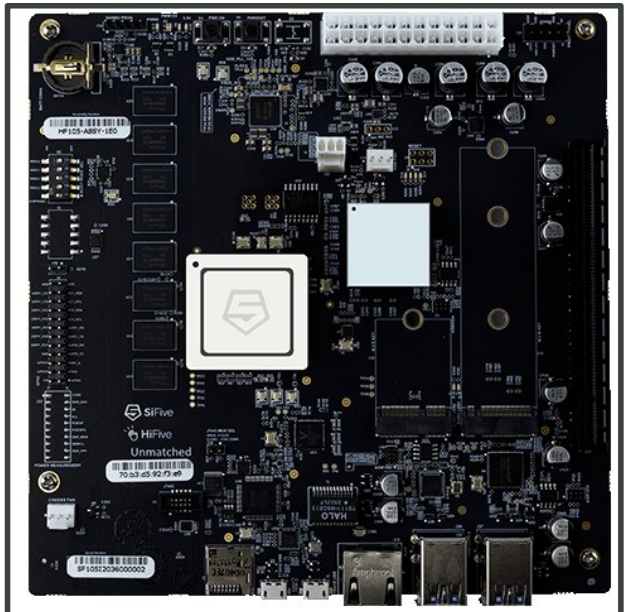
Initially SoCs only available



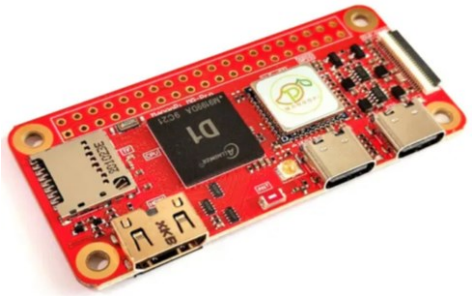
VisionFive V2: JH7200 SoC with U74 core @ 1.5 GHz. 8GB DDR



VisionFive V1: JH7100 SoC with U74 core @ 1.2 GHz. 8GB DDR



HiFive Unmatched: Freedome U740 SoC with U74 core @ 1.2 GHz. 16GB DDR



MangoPi: All Winner D1 SoC @ 1.0 GHz, with C906 core. 1GB DDR

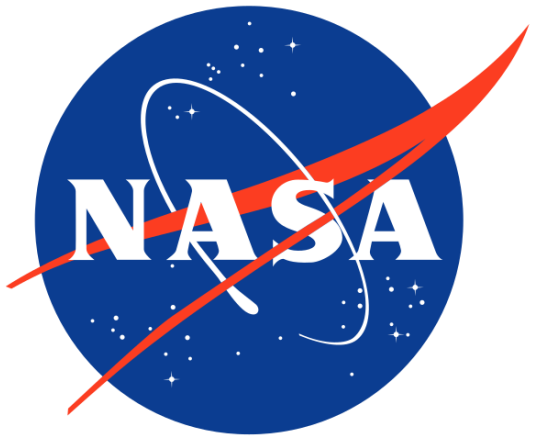
- U74: Dual Issue, in-order 8 stage.
- C906: In-order 5 stage. RVV v0.7.1 supported

The SG2042 was a game changer

- Comprises 64 XuanTie C920 cores
  - 12-stage out-of-order multiple issue superscalar pipeline design
  - RV64GCV instruction set, the C920 has three decode, four rename/dispatch, eight issue/execute and two load/store execution units
  - RVV v0.7.1 is supported with a vector width of 128 bits.
  - Each core contains 64KB of L1 instruction (I) and data (D) cache, 1MB of L2 cache which is shared between the cluster of four cores, and 64MB of L3 system cache which is shared by all cores in the package.
- The SG2042 also provides four DDR4-3200 memory controllers, and 32 lanes of PCI-E Gen4.

Free access if you are interested, see https://riscv.epcc.ed.ac.uk

NAS Parallel Benchmarks (NPBs)



- Suite developed by NASA's Advanced Supercomputing (NAS) division
  - Aim is to characterise HPC systems, especially for CFD applications
  - We focus on the five original kernels, and three pseudo-applications
  - Driven by a variety of problem size classes

Benchmark	Clock ticks cache stall	Clock ticks DDR stall	Time DDR bandwidth bound
Integer Sort (IS)	35%	0%	16%
Multi Grid (MG)	34%	20%	88%
Embarrassingly Parallel (EP)	11%	0%	0%
Conjugate Gradient (CG)	19%	18%	0%
Fast Fourier Transform (FT)	13%	9%	18%

Profiling with Vtune on a Xeon 8170 via OpenMP over all 26 cores

- IS: Indirect, random memory access & integer performance
- MG: Memory bound
- EP: Tests floating point computer performance
- CG: Irregular memory access and nearest neighbour interactions
- FT: All to All neighbour interactions

epcc | THE UNIVERSITY OF EDINBURGH

Comparing against other RISC-V cores

Benchmark	SG2042	VisionFive V2	VisionFive V1	SiFive U740	All Winner D1
IS	60.6	17.84 *(29%)*	6.36 *(10%)*	9.09 *(15%)*	5.41 *(9%)*
MG	1210.05	288.65 *(24%)*	72.31 *(6%)*	90.28 *(7%)*	163.19 *(13%)*
EP	31.35	12.01 *(38%)*	7.55 *(24%)*	9.08 *(29%)*	9.23 *(29%)*
CG	205.25	43.61 *(21%)*	21.96 *(11%)*	20.09 *(10%)*	12.99 *(6%)*
FT	857.64	245.99 *(29%)*	88.35 *(10%)*	116.59 *(14%)*	DNR

- Reported is Mops/s **(higher is better)**
- In red is the percentage performance of the C920 in the SG2042 provided by this specific CPU core

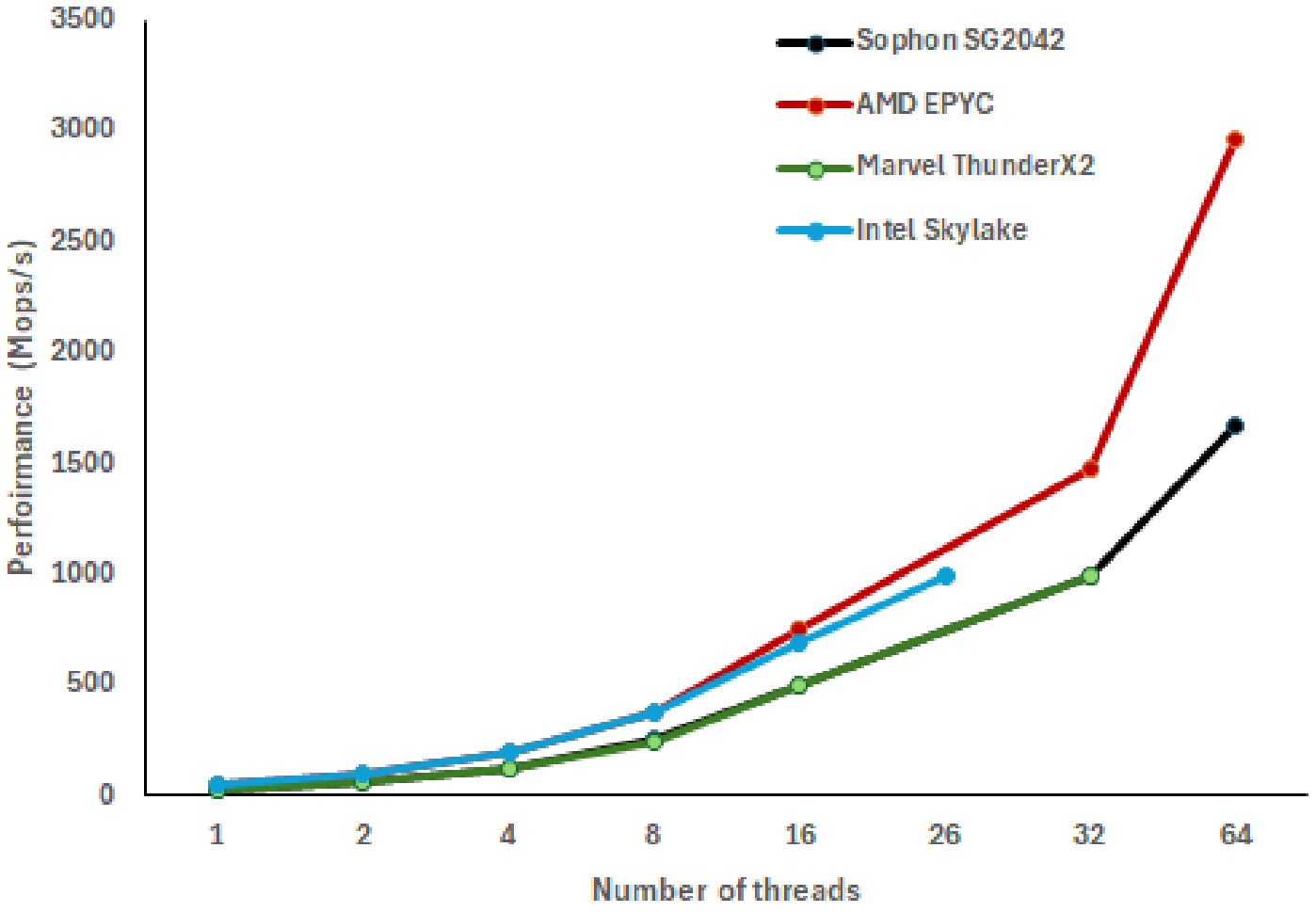
- Running class B of the suite
- Using GCC 8.4
- Using O3

epcc | THE UNIVERSITY OF EDINBURGH

Comparing against other architectures

CPU	ISA	Part	Base clock	Number of cores	Vector
AMD EPYC	x86-64	EPYC 7742	2.25GHz	64	AVX2
Intel Skylake	x86-64	Xeon Platinum 8170	2.1 GHz	26	AVX512
Marvell ThunderX2	ARMv8.1	CN9980	2 GHz	32	NEON
Sophon SG2042	RV64GCV	SG2042	2 GHz	64	RVV v0.7.1

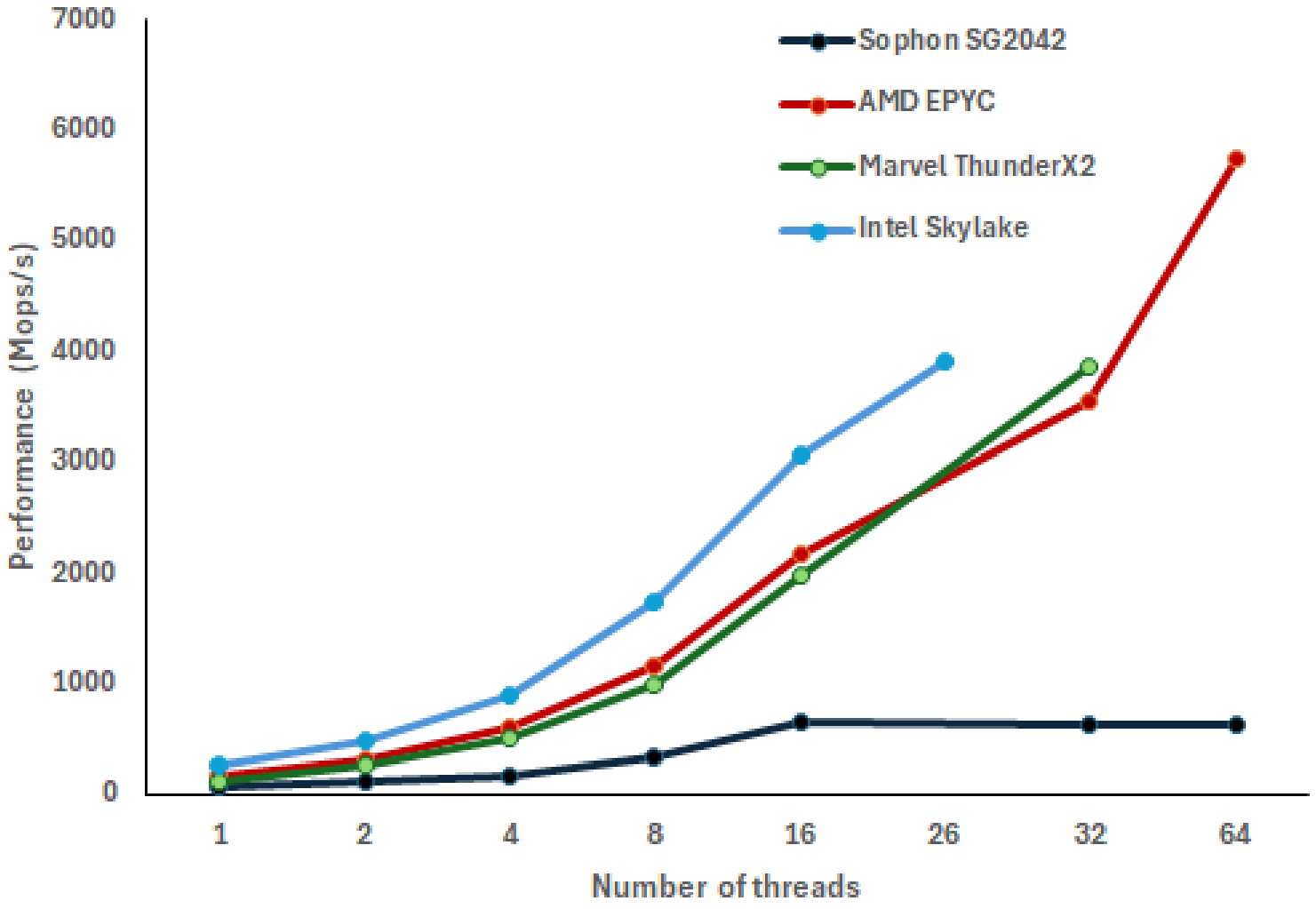
Embarrassingly Parallel (EP) benchmark



- Focusses on compute performance
  - Two groups of performance here
- SG2042 and ThunderX2 are both 128-bit wide vector (although ThunderX2 has two FPUs per core).
- Skylake and EPYC provide wider vectorisation

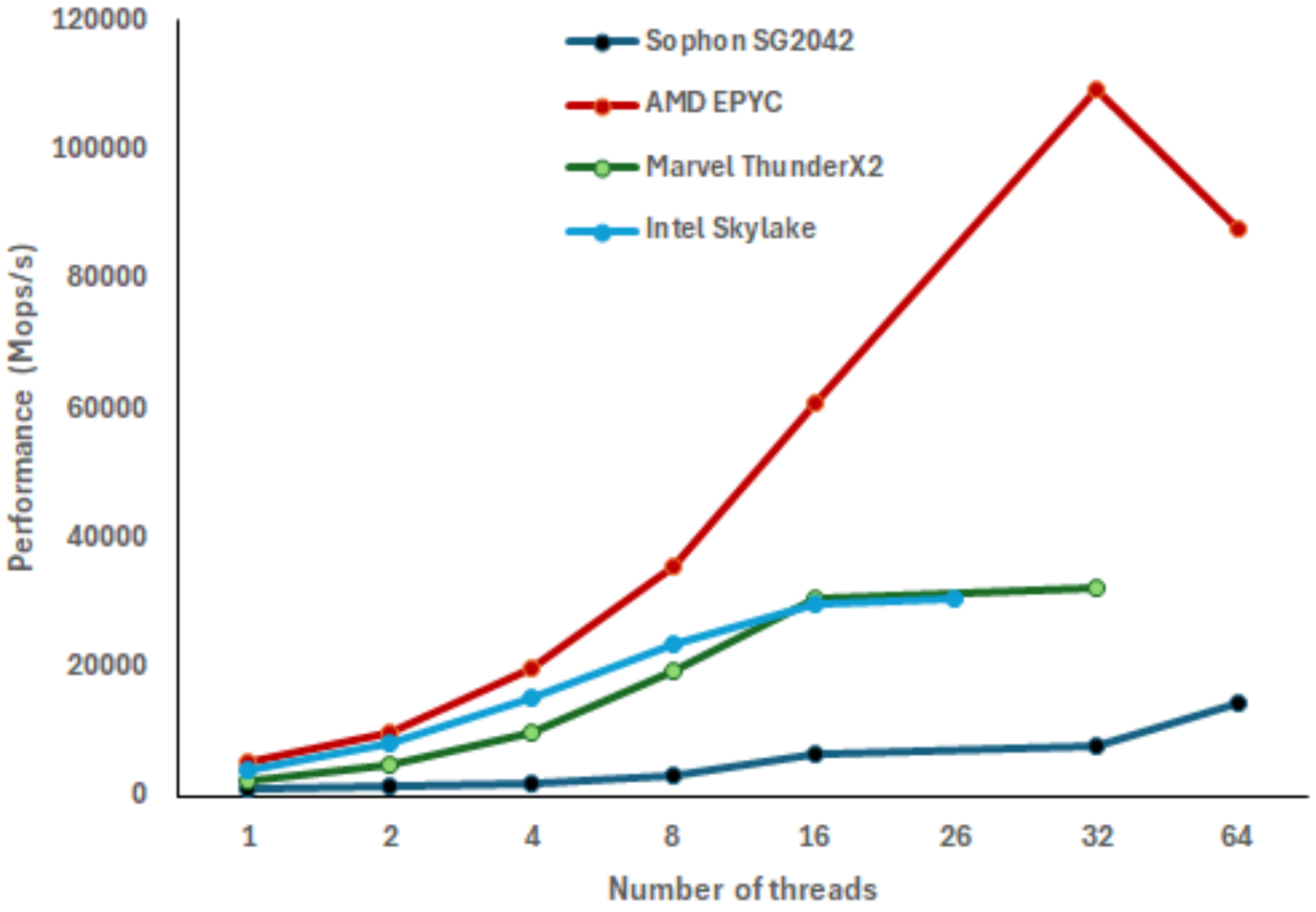
epcc | THE UNIVERSITY OF EDINBURGH

IS benchmark



- Integer comparison and indirect, random, memory accesses
- SG2042 plateaus at 16 cores and performs worse than all the others
- Possibly to do with cache hierarchy
  - Skylake has largest L2 cache (1MB per core) compared to 256KB per core for SG2042 & ThunderX2. EPYC has 512KB per core.

MultiGrid (MG) benchmark



- Memory bandwidth bound
- EPYC provides best performance, ThunderX2 and Skylake are similar and plateau at 16 cores
- SG2042 is always lowest performance here
- EPYC: 8 memory channels, 8 controllers connected to DDR4-3200
- Skylake & ThunderX2: 2 memory controllers (6 channels for Skylake, 8 for ThunderX2). Connected to DDR4-2666
- SG2042: 4 memory controller and 4 channels, connected to DDR4-3200

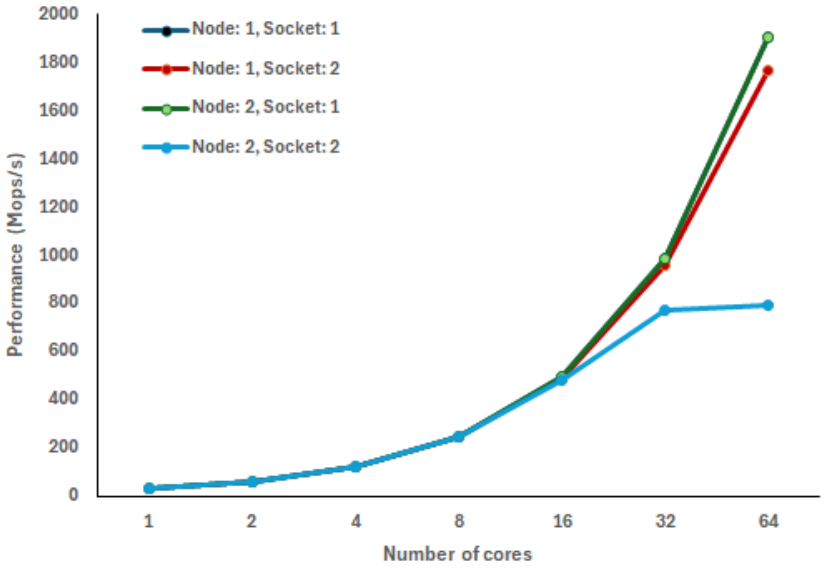
epcc | THE UNIVERSITY OF EDINBURGH

Going multi-socket

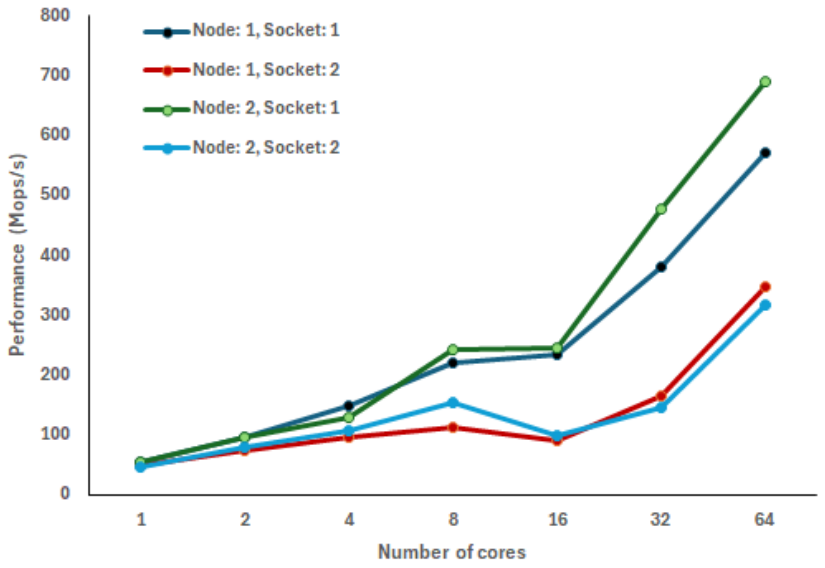
- It seemed like the SG2042 was memory bound
  - But is this due to limits in the CPU or the Milk-V Pioneer workstation?

- E4 Computer Engineering developed a prototype dual-socket SG2042 system that we then re-ran these benchmarks on
  - This is more in the form factor we would deploy in an HPC centre
  - Dual-socket is also a very common configuration in HPC machines

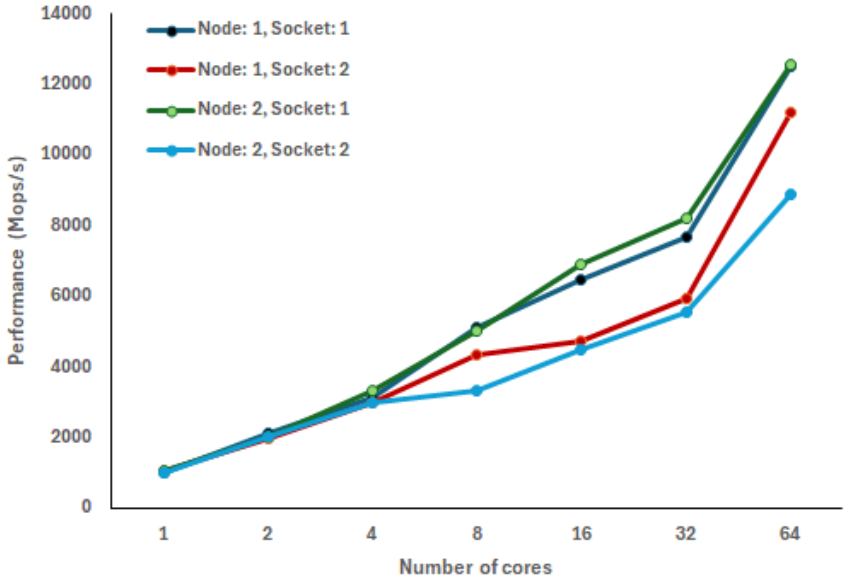
Inter-socket performance



Embarrassingly Parallel (EP)



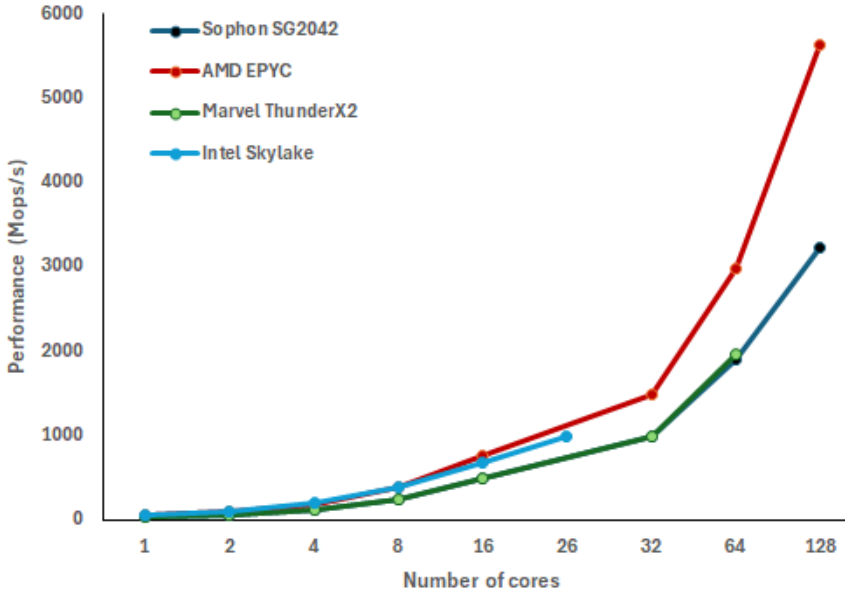
Integer Sort (IS)



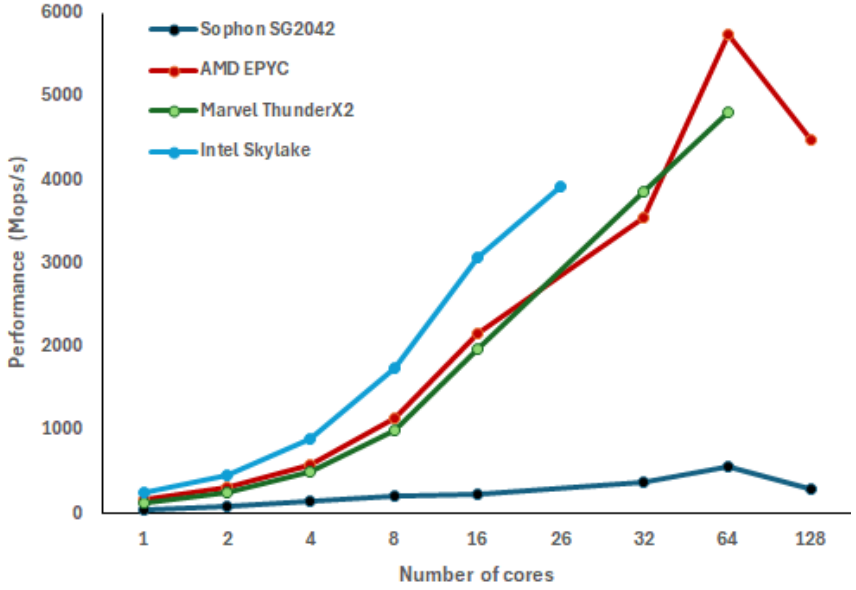
Multi Grid (MG)

More details at https://arxiv.org/pdf/2502.10320

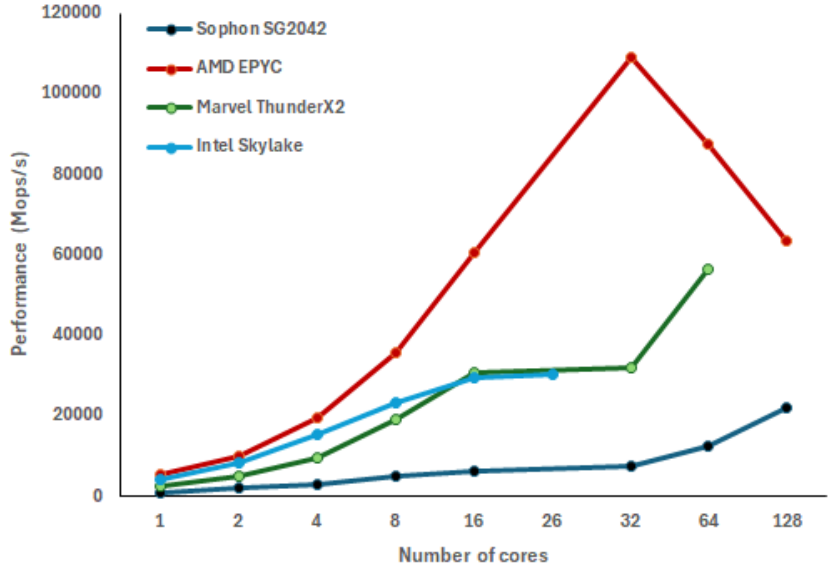
Multi-socket performance comparison



Embarrassingly Parallel (EP)



Integer Sort (IS)



Multi Grid (MG)

More details at https://arxiv.org/pdf/2502.10320

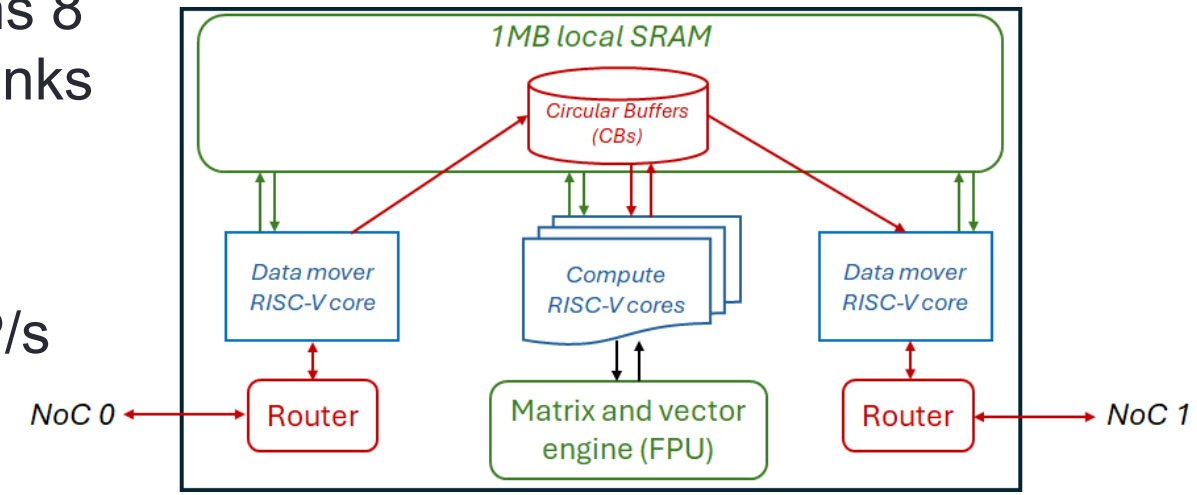
epcc

THE UNIVERSITY OF EDINBURGH

Tenstorrent Grayskull PCIe accelerator

- PCIe accelerator card, initially designed for AI workloads but can program directly using TT-Metalium API.
  - Hence potential for scientific computing workloads

- In this work we use the e150 which contains 8 GiB of DRAM which is split across eight banks
- Running at 1.2 GHz
- 120 Tensix cores
- Quoted as providing up to 332 FP8 TFLOP/s
- Available for purchase and very affordable



Optimised performance and energy efficiency

- A stencil based code which are very common in HPC
- 163 times better performance than our initial unoptimized version
- Across the entire e150 we slightly outperform the Xeon Platinum but at around 5 times less energy

Type	Total cores	Cores in Y	Cores in X	Performance (GPt/s)	Energy (Joules)
CPU	1	-	-	1.41	1657
CPU	24	-	-	21.61	588
e150	1	1	1	1.06	2094
e150	2	1	2	2.48	893
e150	4	1	4	2.92	744
e150	8	4	4	7.99	276
e150	32	8	4	9.20	240
e150	64	8	8	12.96	170
e150	72	8	9	17.26	128
e150	108	12	9	22.06	110
e150 x 2	216	24	9	44.12	102
e150 x 4	432	48	9	86.75	108

More details at https://arxiv.org/pdf/2409.18835

epcc

THE UNIVERSITY OF EDINBURGH

Conclusions

- RISC-V is advancing for HPC and there are some very interesting technologies
- SG2042 is a serious contender, but there is still work to do against existing architectures that are popular in HPC

- PCIe accelerators are potentially where we are going to see RISC-V adopted in HPC initially because it is easier to add to an existing ecosystem
  - Vendors such as Tenstorrent have produced some impressive hardware both in terms of performance but also energy usage
  - Still there are challenges around porting codes to these architectures, we are learning a great deal about optimising HPC codes for them

epcc | THE UNIVERSITY OF EDINBURGH