PERFORMANCE CHARACTERISATION OF THE 64-CORE SG2042 RISC-V CPU FOR HPC

Nick Brown, EPCC

n.brown@epcc.ed.ac.uk

epcc

THE UNIVERSITY OF EDINBURGH

Sophon SG2042: A high-core count RISC-V CPU

- 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.

epcc | THE UNIVERSITY OF EDINBURGH

An HPC testbed for RISC-V

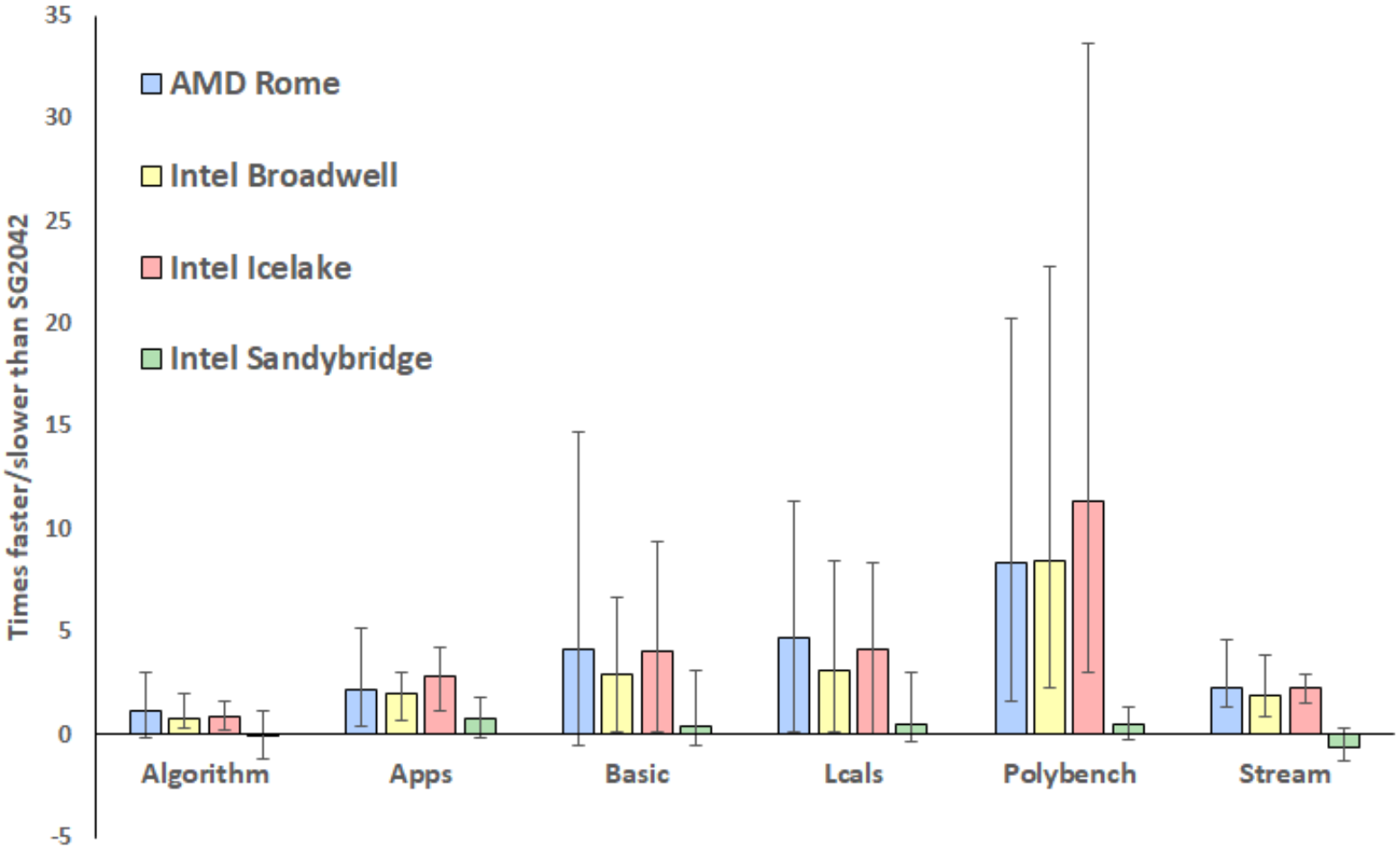
- Our SG2042s are in Milk-V Pioneer workstations, each with 128GB of DDR
- Set up with other nodes as an HPC-style system
  - With a login node, shared filesystem, all controlled under Slurm, compilers and libraries available via the module environment
  - Aim is to make this appear as HPC-like as possible to users

EXCALIBUR 10

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

epcc | THE UNIVERSITY OF EDINBURGH

Previous work exploring performance of SG2042

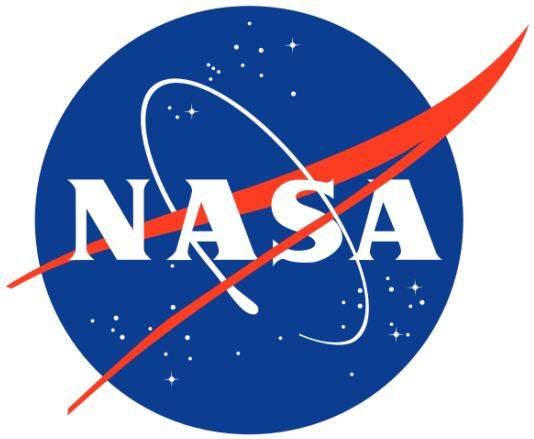


- Explored running the RAJAPerf benchmarking suite on the SG2042 and comparing against other architectures
  - Lots of individual kernels, some more representative than others of HPC workloads
  - Difficult to see individual patterns and draw conclusions

epcc

THE UNIVERSITY OF EDINBURGH

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

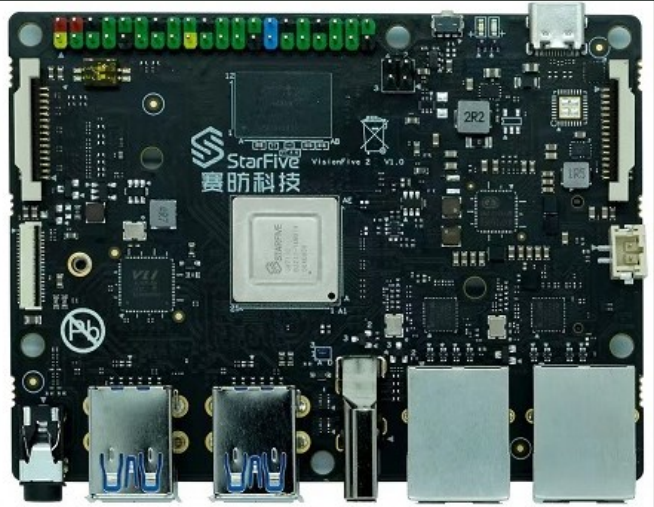
Benchmark	Clock ticks cache stall	Clock ticks DDR stall	Time DDR bandwidth bound
Integer Sort (IS)	35%	0%	16%
Multi Grid (MG)	34%	20%	88%
Embarrassingly Parallel (EP)	11%	0%	0%
Conjugate Gradient (CG)	19%	18%	0%
Fast Fourier Transform (FT)	13%	9%	18%
Block Tridiagonal (BT)	8%	9%	0%
LU Gauss Seidel (LU)	12%	11%	0%
Scalar Pentadiagonal (SP)	20%	21%	0%

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

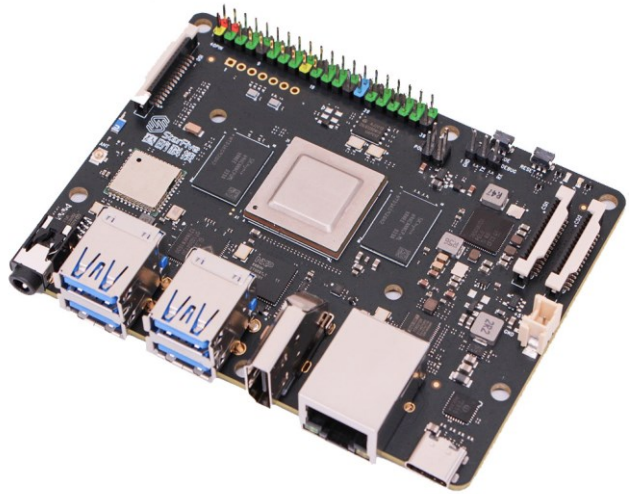
- IS: Indirect, random memory access & integer performance
- MG: Memory bound
- EP: Tests floating point computer performance
- CG: Irregular memory access and nearest neighbour interactions
- FT: All to All neighbour interactions

epcc | THE UNIVERSITY OF EDINBURGH

Comparing against other RISC-V cores

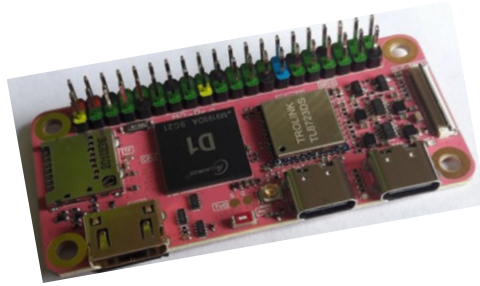


VisionFive V2: JH7200 SoC with U74 core @ 1.5 GHz. 8GB DDR



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

HiFive Unmatched: Freedom 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

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

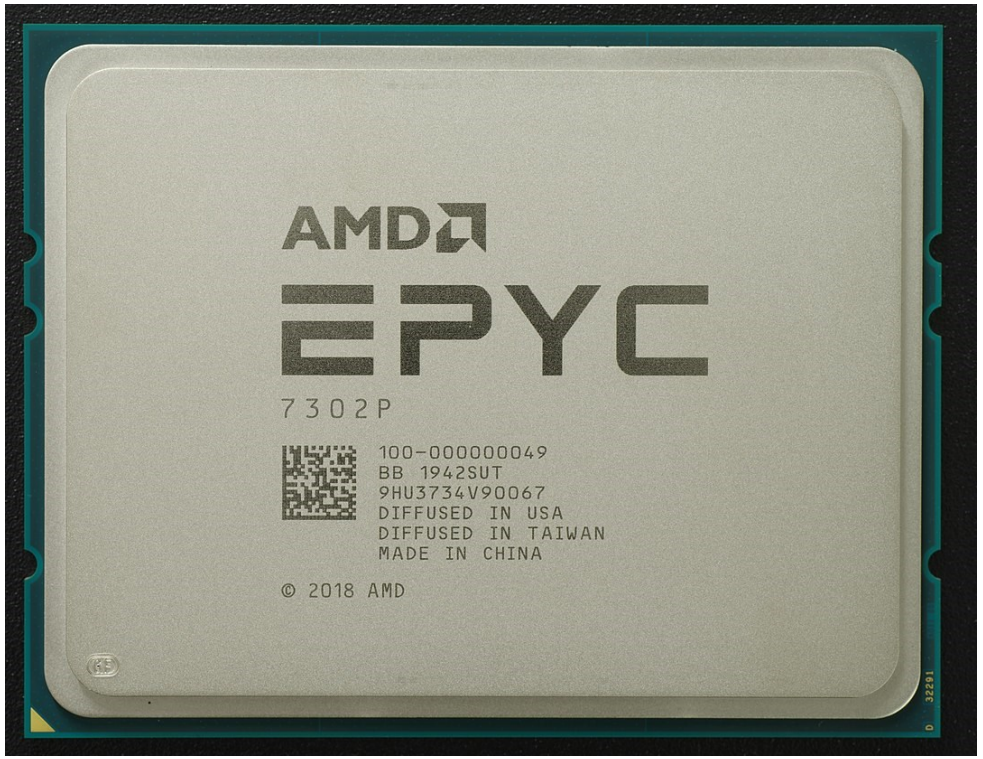
- Reported is Mops/s **(higher is better)**
- In red is the percentage performance of the C920 in the SG2042 provided by this specific CPU core

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

epcc | THE UNIVERSITY OF EDINBURGH

Comparing against other architectures

AMD EPYC



- EPYC 7742
- 64 cores across 4 NUMA regions
- 8 memory channels and controllers.
- Each core contains
  - 32KB L1 I and D cache
  - 512KB L2 cache
  - 16MB L3 shared between four cores
- Provides AVX2 (256 bits wide)
  - Can process two AVX2 instructions per cycle
- Part of ARCHER2, a Cray-EX
- 256GB memory per node
- Using GCC 11.2

Intel Skylake

- Xeon Platinum 8170
- 26 cores
- 6 memory channels and 2 controllers.
- Each core contains
  - 32KB L1 I and D cache
  - 1MB L2 cache
  - 35.75MB L3 shared between all cores
- Provides AVX512 (512 bits wide)
  - Can process two AVX512 instructions per cycle
- 192GB memory per node
- Using GCC 8.4

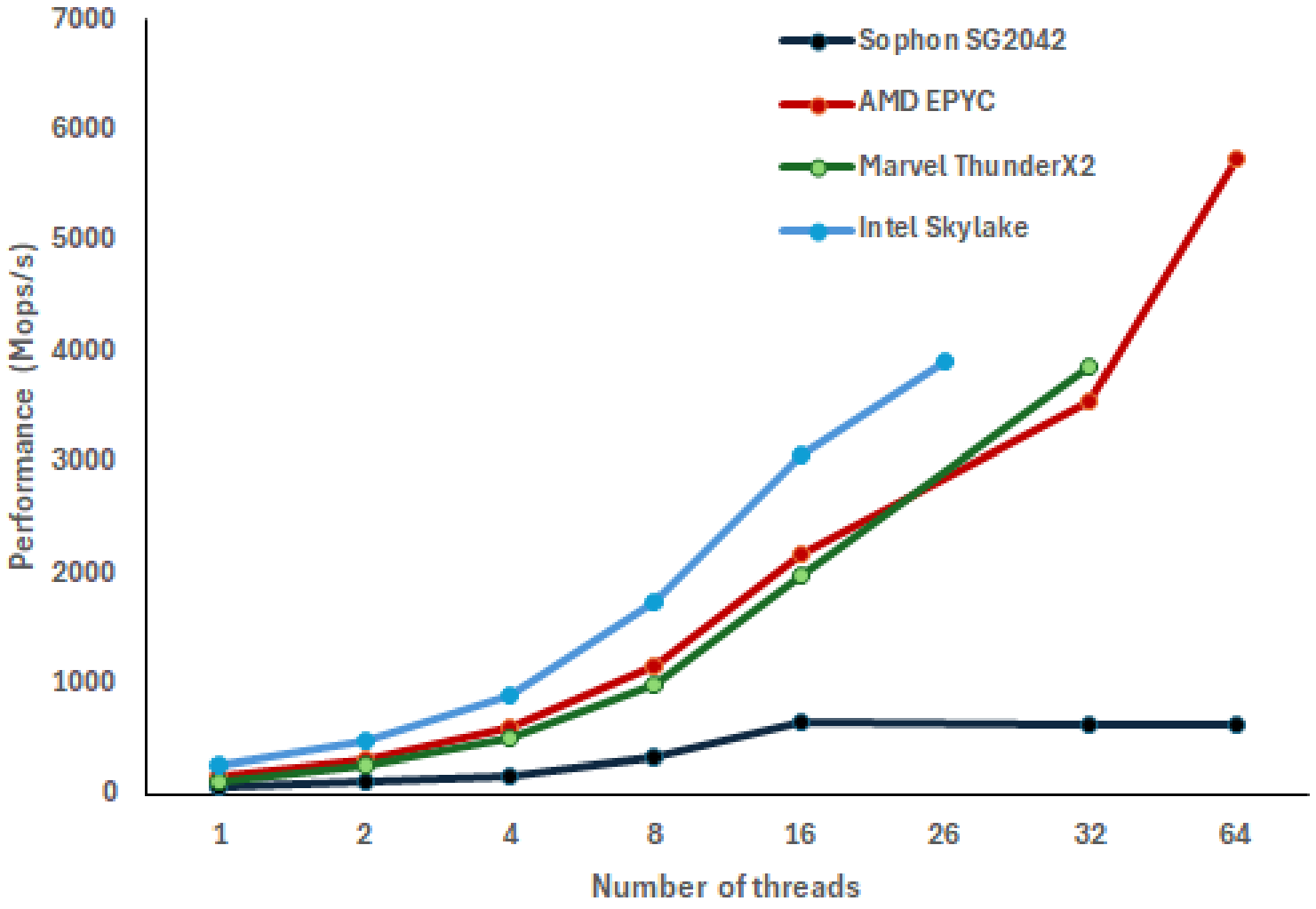
Marvell ThunderX2

- CN9980
- 32 cores implementing ARMv8.1
- 8 memory channels and 2 controllers.
- Each core contains
  - 32KB L1 I and D cache
  - 256MB L2 cache
  - 32MB L3 shared between all cores
- Provides NEON (128 bits wide)
  - Contains two FPUs
- 128GB memory per node
- Using GCC 9.2

Other architectures

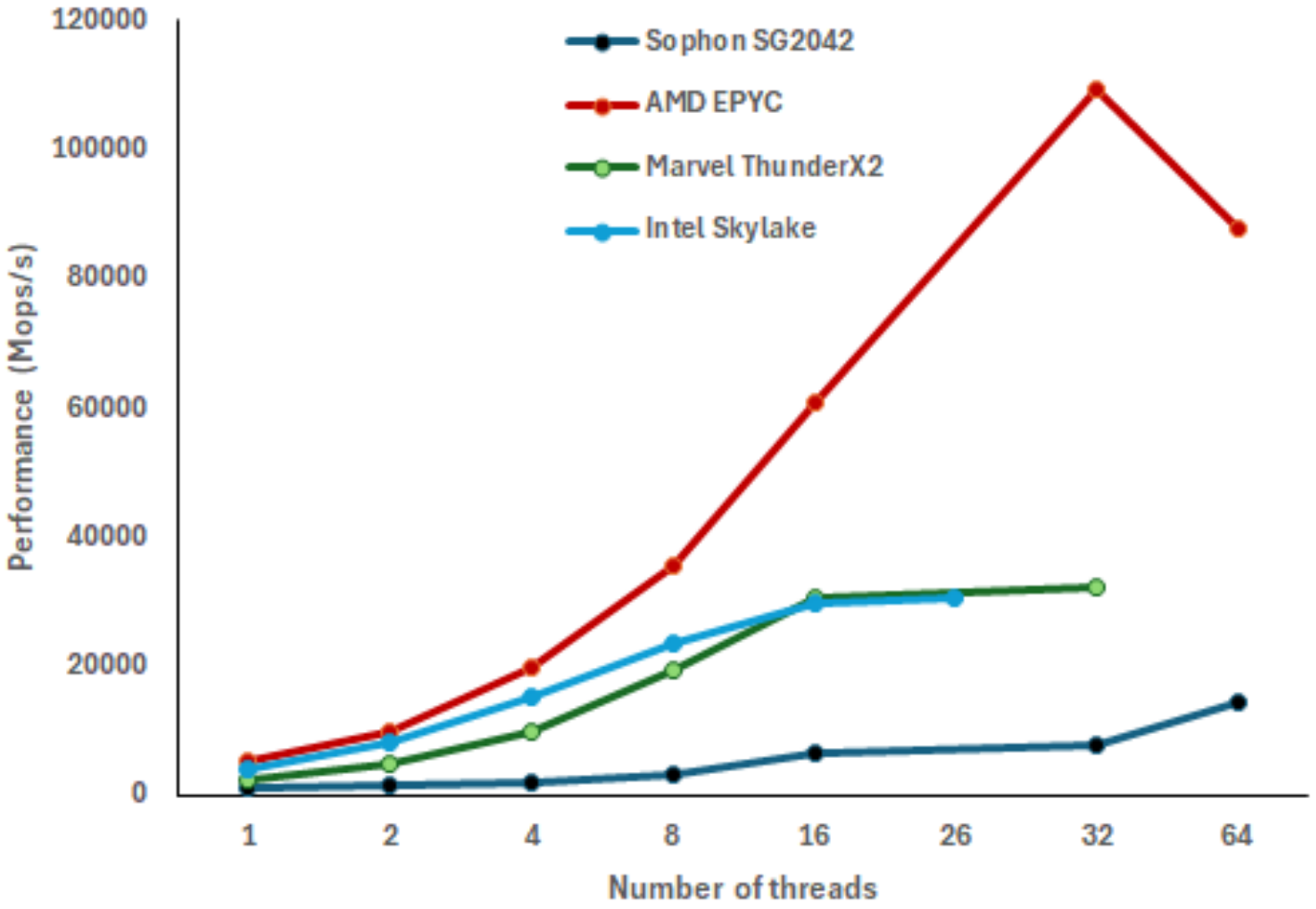
CPU	ISA	Part	Base clock	Number of cores	Vector
AMD EPYC	x86-64	EPYC 7742	2.25GHz	64	AVX2
Intel Skylake	x86-64	Xeon Platinum 8170	2.1 GHz	26	AVX512
Marvell ThunderX2	ARMv8.1	CN9980	2 GHz	32	NEON
Sophon SG2042	RV64GCV	SG2042	2 GHz	64	RVV v0.7.1

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.

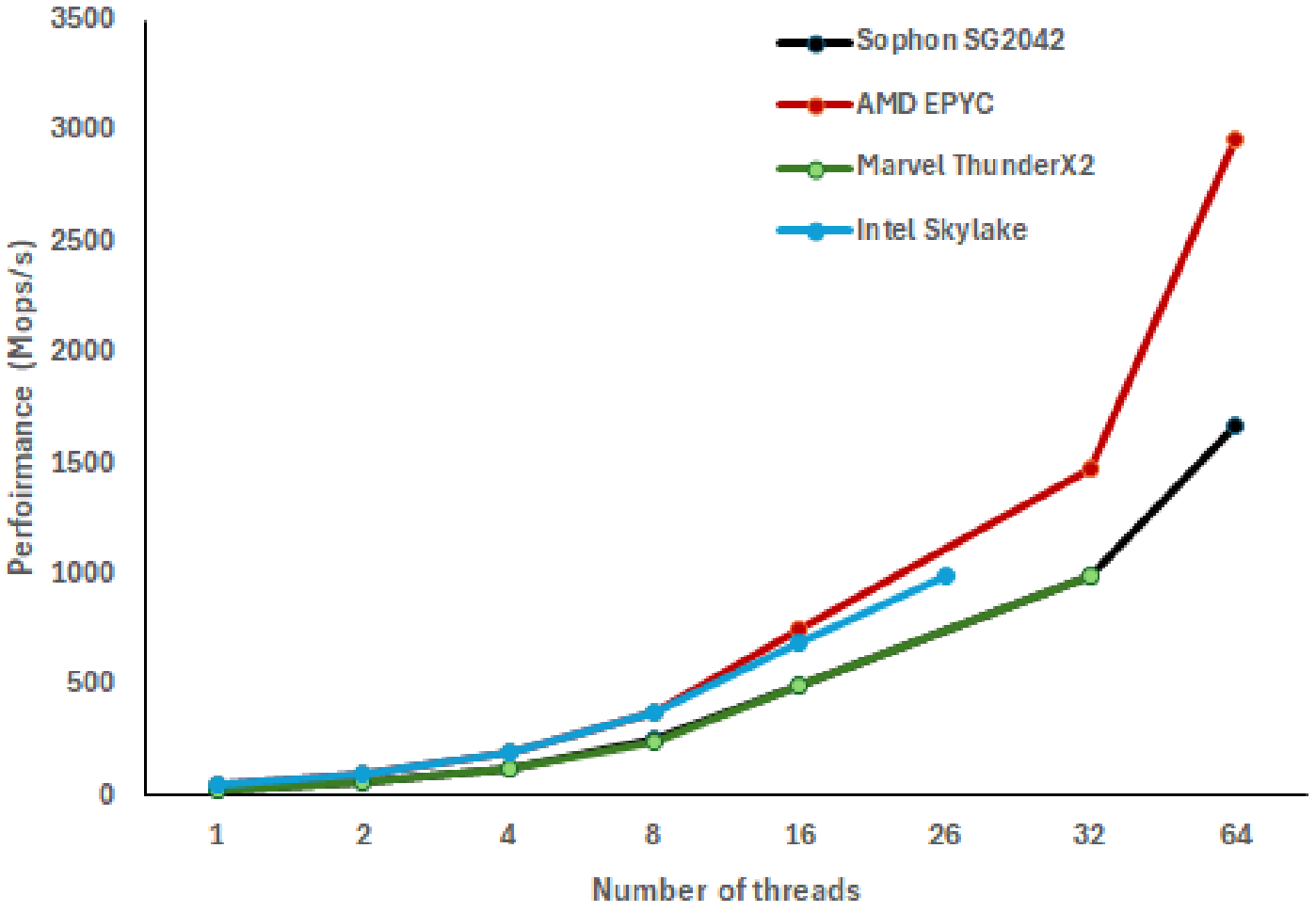
MultiGrid (MG) benchmark



- 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

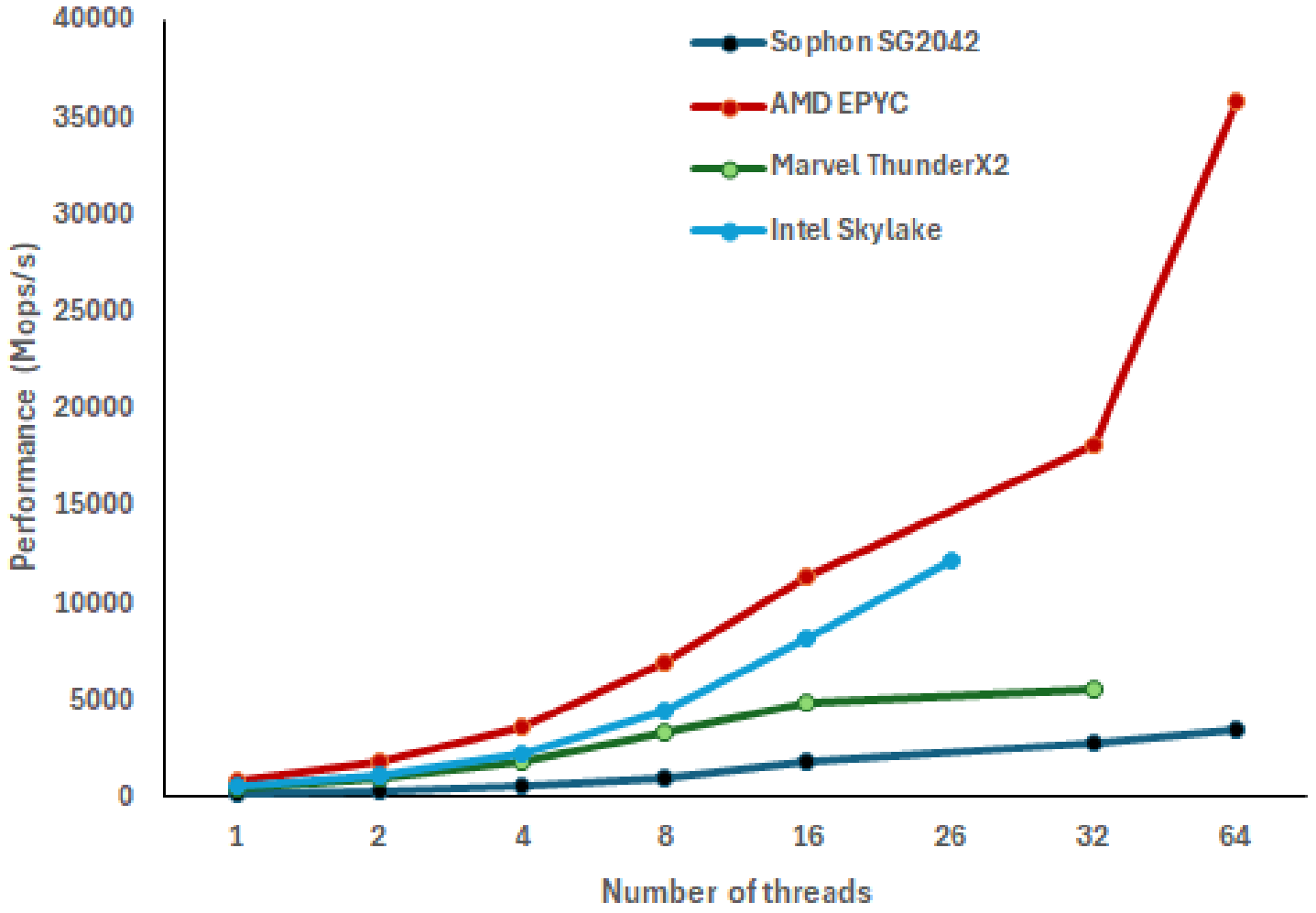
Embarrassingly Parallel (EP) benchmark



- 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

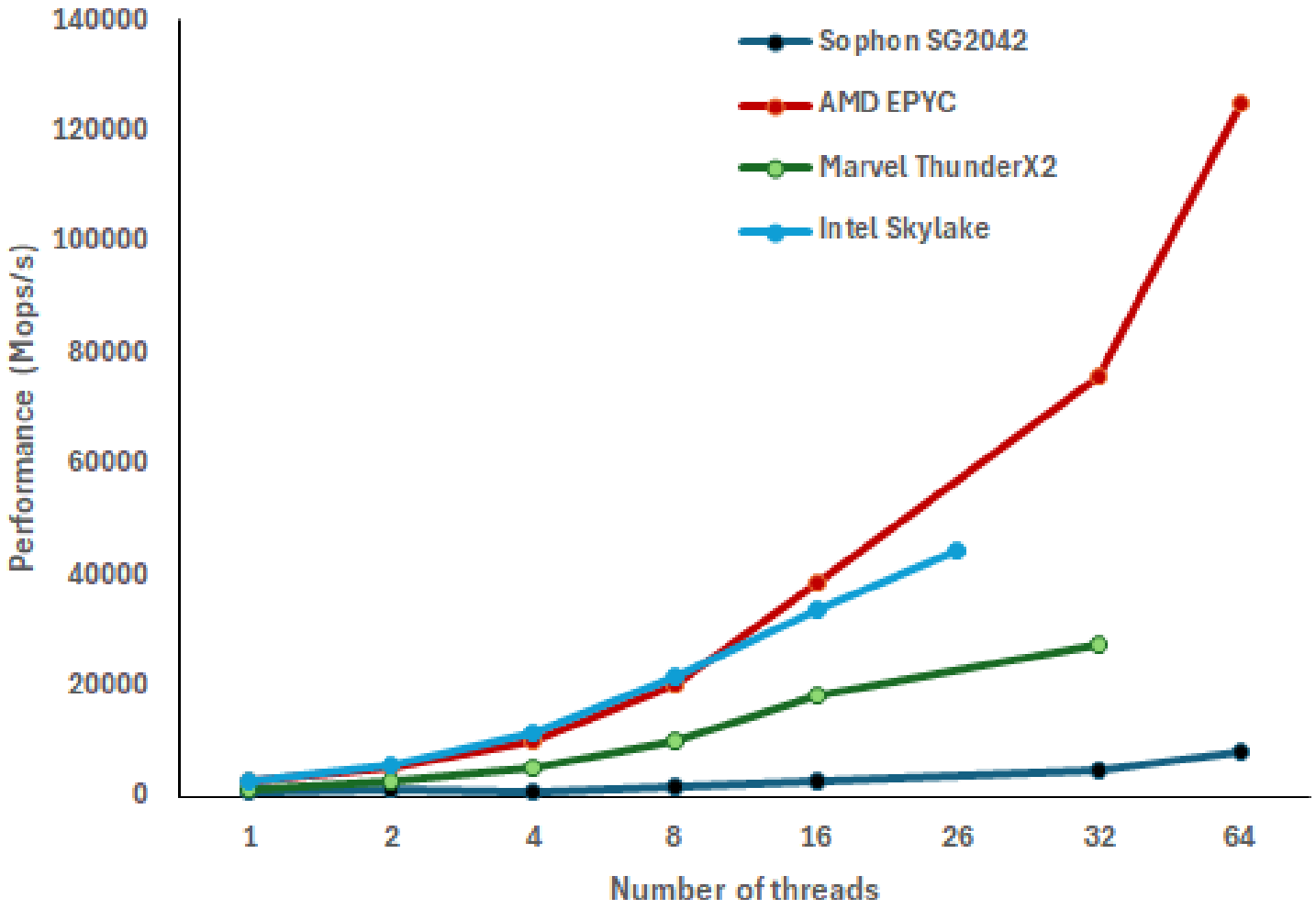
epcc | THE UNIVERSITY OF EDINBURGH

Conjugate Gradient (CG) benchmark



- CG spends considerable time stalled on cache and DDR memory access
  - Irregular memory accesses and nearest neighbour communications
- SG2042 is closer to the ThunderX2 than we had anticipated (around 50% of the performance)
  - Both have the same L2 (256KB) and 1MB L3 cache per core
  - EPYC has 4MB per core, Skylake 1.3MB per core (but large L2 cache too)

Fast Fourier Transform (FT) benchmark



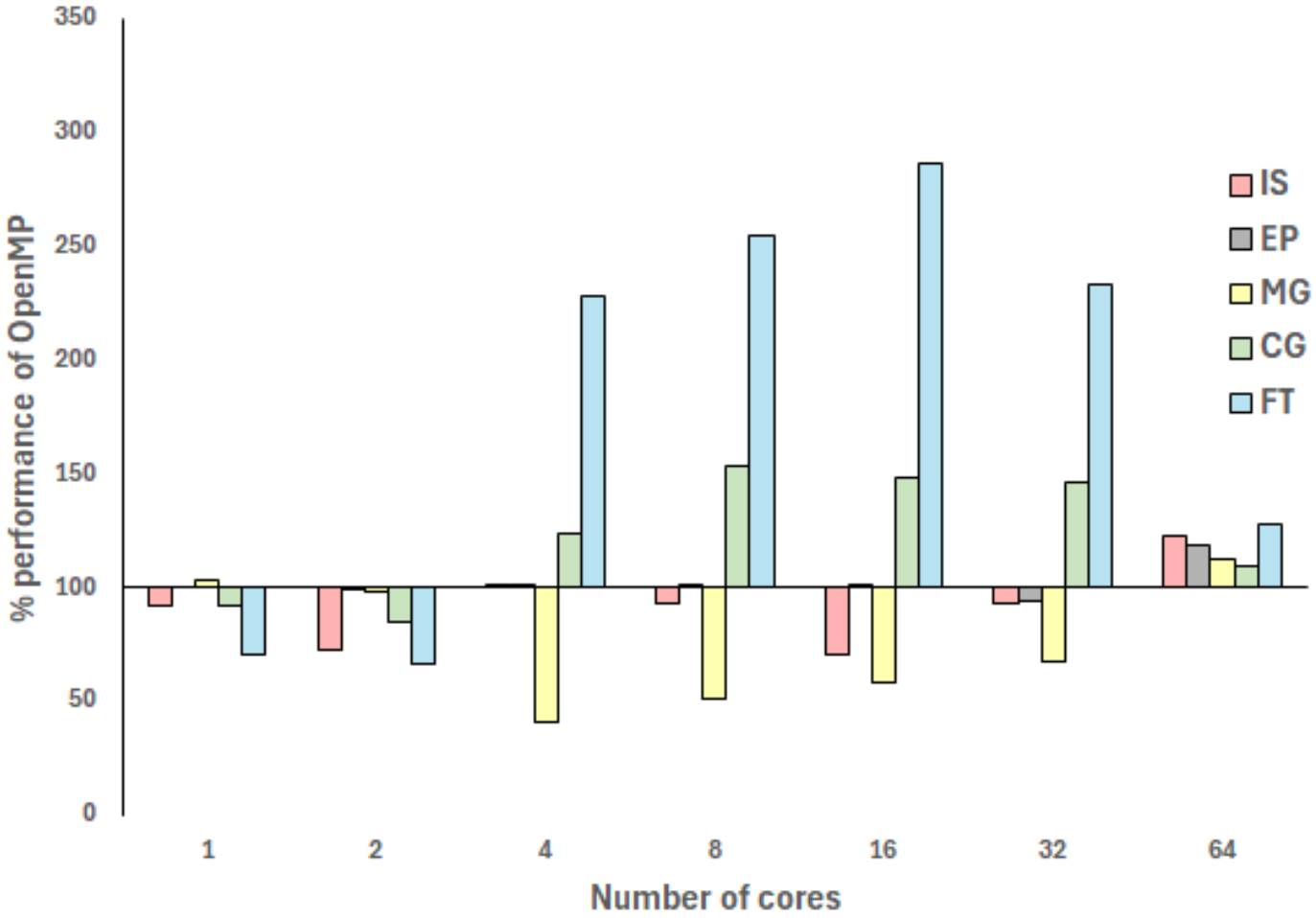
- All to all communications, with some time stalled due to cache and DDR access but also 18% of time spent under high DDR utilisation

Pseudo applications

Number cores	BT benchmark			LU benchmark			SP benchmark		
	EPYC	Skylake	ThunderX2	EPYC	Skylake	ThunderX2	EPYC	Skylake	ThunderX2
16	3.23	3.28	2.43	3.65	4.15	2.86	5.01	3.91	3.65
26	3.57	2.97	2.69	3.20	3.16	2.62	6.25	3.48	3.57
32	3.68	-	2.64	3.40	-	2.94	5.26	-	3.22
64	4.19	-	-	2.95	-	-	4.22	-	-

- Performance reported is how many times faster it is than the SG2042
- Based on the initial profiling of the suite and what we have seen around challenges with memory performance, we assumed that the SG2042 would perform best for the BT benchmark, and worst for SP.
  - Generally, this pattern holds

Should we use OpenMP or MPI in a node?



- Over one or two cores, it's better to use OpenMP regardless
- At 64 cores, all benchmarks run faster using MPI
- MPI always slower than OpenMP on EPYC
- Same for Skylake apart from CG, where MPI about twice as fast
  - When profiling, clock ticks stalled due to cache reduced to 5.5% with MPI (was 19% with OpenMP). MPI has no clock ticks stalled for DDR access (18% in OpenMP version)

General comments

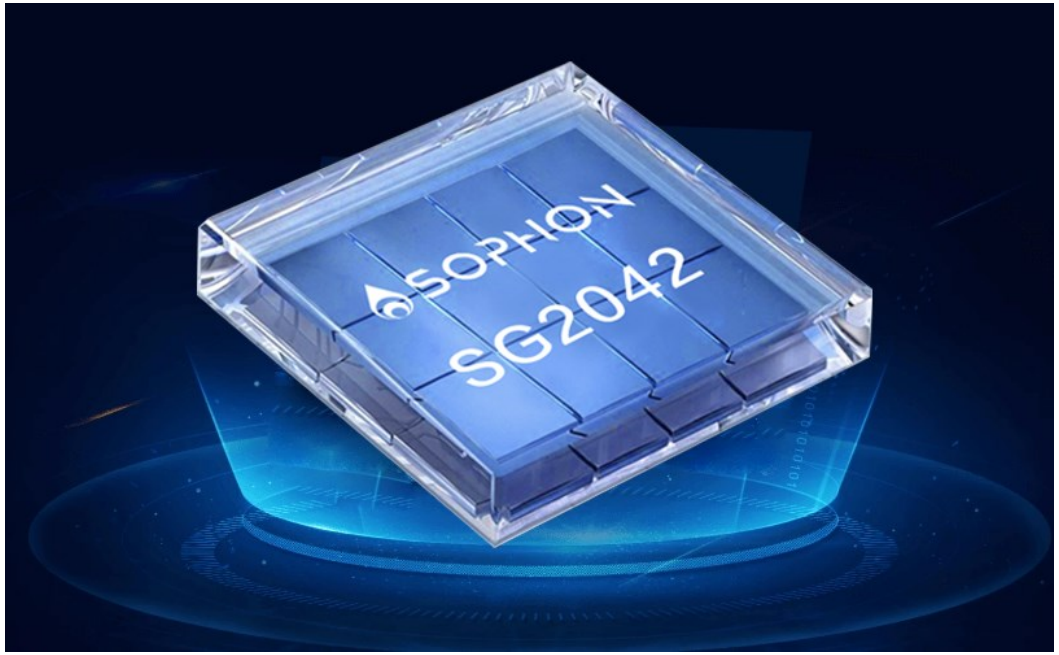
- Would be nice to use the *simd* directive of OpenMP
  - Supported in LLVM but that required RVV v1.0
- Profiling is immature, and this would be nice to develop
  - There are some performance counters on the SG2042, but often difficult to see the link between the numbers reported and the performance observed
  - Profiling tooling would be nice (I think extrae from BSC is ported to RISC-V)

```
branch-instructions OR branches                    [Hardware event]
branch-misses                                      [Hardware event]
bus-cycles                                         [Hardware event]
cache-misses                                       [Hardware event]
cache-references                                   [Hardware event]
cpu-cycles OR cycles                               [Hardware event]
instructions                                       [Hardware event]
ref-cycles                                         [Hardware event]
stalled-cycles-backend OR idle-cycles-backend      [Hardware event]
stalled-cycles-frontend OR idle-cycles-frontend    [Hardware event]
L1-dcache-load-misses                              [Hardware cache event]
L1-dcache-loads                                    [Hardware cache event]
L1-dcache-prefetch-misses                          [Hardware cache event]
L1-dcache-prefetches                               [Hardware cache event]
L1-dcache-store-misses                             [Hardware cache event]
L1-dcache-stores                                   [Hardware cache event]
L1-icache-load-misses                              [Hardware cache event]
L1-icache-loads                                    [Hardware cache event]
L1-icache-prefetch-misses                          [Hardware cache event]
L1-icache-prefetches                               [Hardware cache event]
LLC-load-misses                                    [Hardware cache event]
LLC-loads                                          [Hardware cache event]
LLC-prefetch-misses                                [Hardware cache event]
LLC-prefetches                                     [Hardware cache event]
LLC-store-misses                                   [Hardware cache event]
LLC-stores                                         [Hardware cache event]
branch-load-misses                                 [Hardware cache event]
branch-loads                                       [Hardware cache event]
dTLB-load-misses                                   [Hardware cache event]
dTLB-loads                                         [Hardware cache event]
dTLB-prefetch-misses                               [Hardware cache event]
dTLB-prefetches                                    [Hardware cache event]
dTLB-store-misses                                  [Hardware cache event]
dTLB-stores                                        [Hardware cache event]
iTLB-load-misses                                   [Hardware cache event]
iTLB-loads                                         [Hardware cache event]
node-load-misses                                   [Hardware cache event]
node-loads                                         [Hardware cache event]
node-prefetch-misses                               [Hardware cache event]
node-prefetches                                    [Hardware cache event]
node-store-misses                                  [Hardware cache event]
node-stores                                        [Hardware cache event]
```

epcc

THE UNIVERSITY OF EDINBURGH

Conclusions



- SG2042 is a very interesting technology, but there is still work to do against existing architectures popular in HPC
- Outperformed by:
  - AMD EPYC between 1.77 and 15.06 times
  - Intel Skylake between 0.59 and 5.98 times
  - Marvel ThunderX2 between 0.59 and 5.91 times
    - SG2042 only outperformed Skylake and ThunderX2 for EP benchmark
- Best suited to computational workloads, struggles with algorithmic patterns that are memory bandwidth or latency bound
- Interesting, the SG2044 is supposed to have 3 times the memory performance of the SG2042 (and RVV 0.7.1)

epcc | THE UNIVERSITY OF EDINBURGH