A scenic view of the Clifton Suspension Bridge in Bristol, England, at sunset. The bridge is a suspension bridge with two large stone towers. The sky is a mix of orange, pink, and purple. Several hot air balloons are visible in the sky. The bridge is surrounded by lush green trees and a river below.

Simulation of Architectures for a Sustainable Computer Ecosystem

14th Aug
2024

Prof. Simon McIntosh-Smith
BriCS

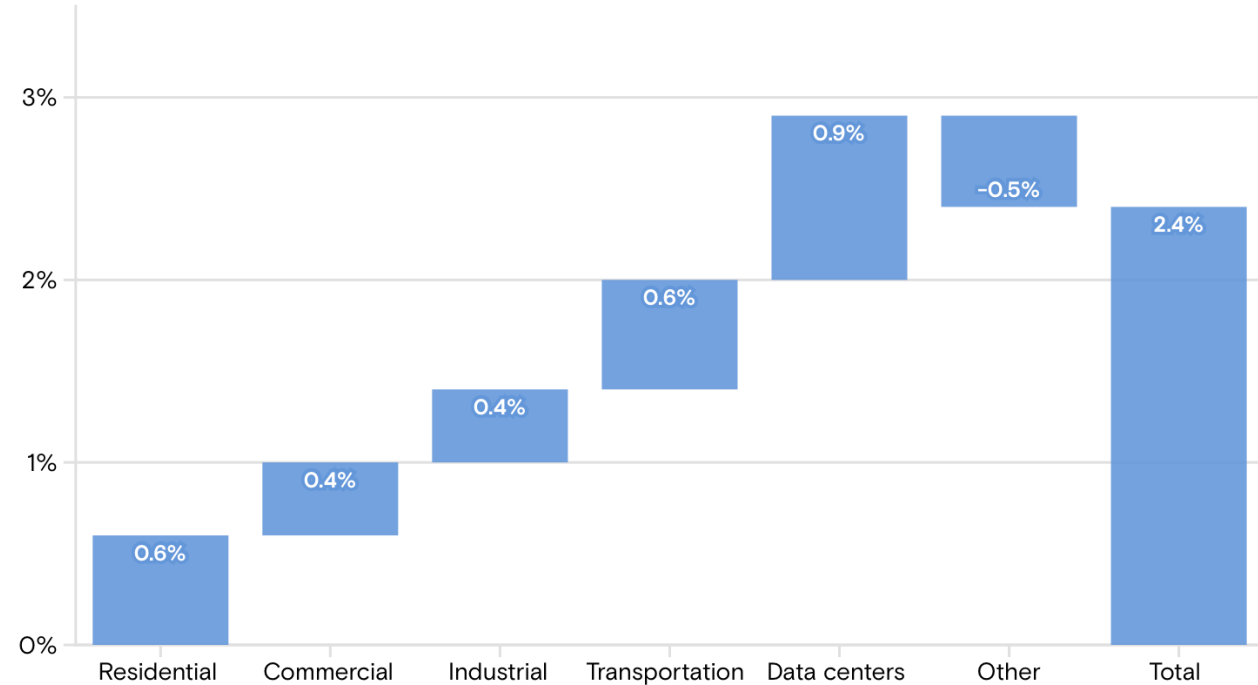


Director, Bristol Centre for Supercomputing (BriCS) and
University of Bristol High Performance Computing Group

Why do we care about sustainable computer architectures? – Our environment

The sectoral growth in US power demand

The demand for electricity is forecast to rise at 2.4% CAGR between 2022–2030



Source: Goldman Sachs Research, EIA

**Goldman
Sachs**

Why do we care about sustainable computer architectures? – Our environment



News & Events ▾

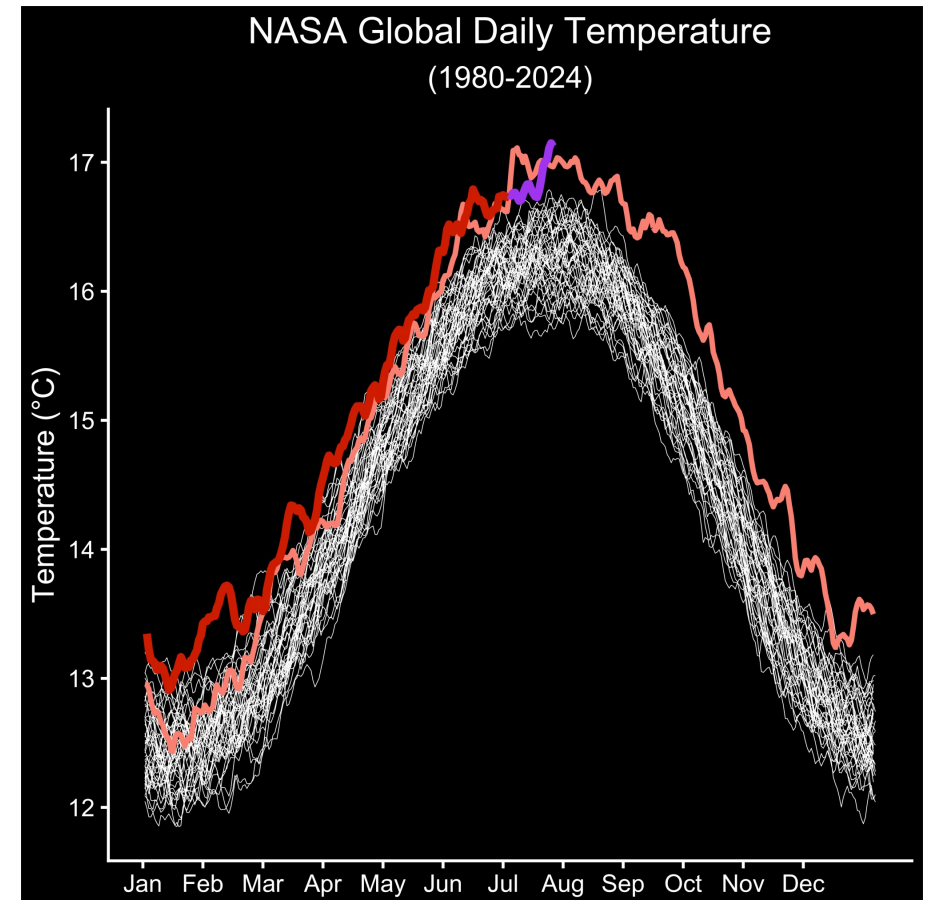
Multimedia ▾

NASA+ **LIVE**

2 MIN READ

NASA Data Shows July 22 Was Earth's Hottest Day on Record

Source: <https://www.nasa.gov/earth/nasa-data-shows-july-22-was-earths-hottest-day-on-record/#:~:text=July%2022%2C%202024%2C%20was%20the,record%2C%20set%20in%20July%202023.>



Why do we care about sustainable computer architectures? – Our environment



Working towards net zero carbon

Our pledge to become a net zero carbon campus means making significant and rapid changes to reduce carbon emissions to the lowest amount – and offsetting as a last resort. We've set a target to reach net zero scope 1 and 2 carbon emissions from our buildings by 2030, and are committed to getting scope 3 emissions to net zero as soon as possible.

Why do we care about sustainable computer architectures? – Performance



BLOG | JULY 12, 2021

Performance Per Watt is the New Moore's Law

The need to decarbonize compute for the sake of our planet means the technology roadmap can no longer prioritize processing power, says Rob Aitken

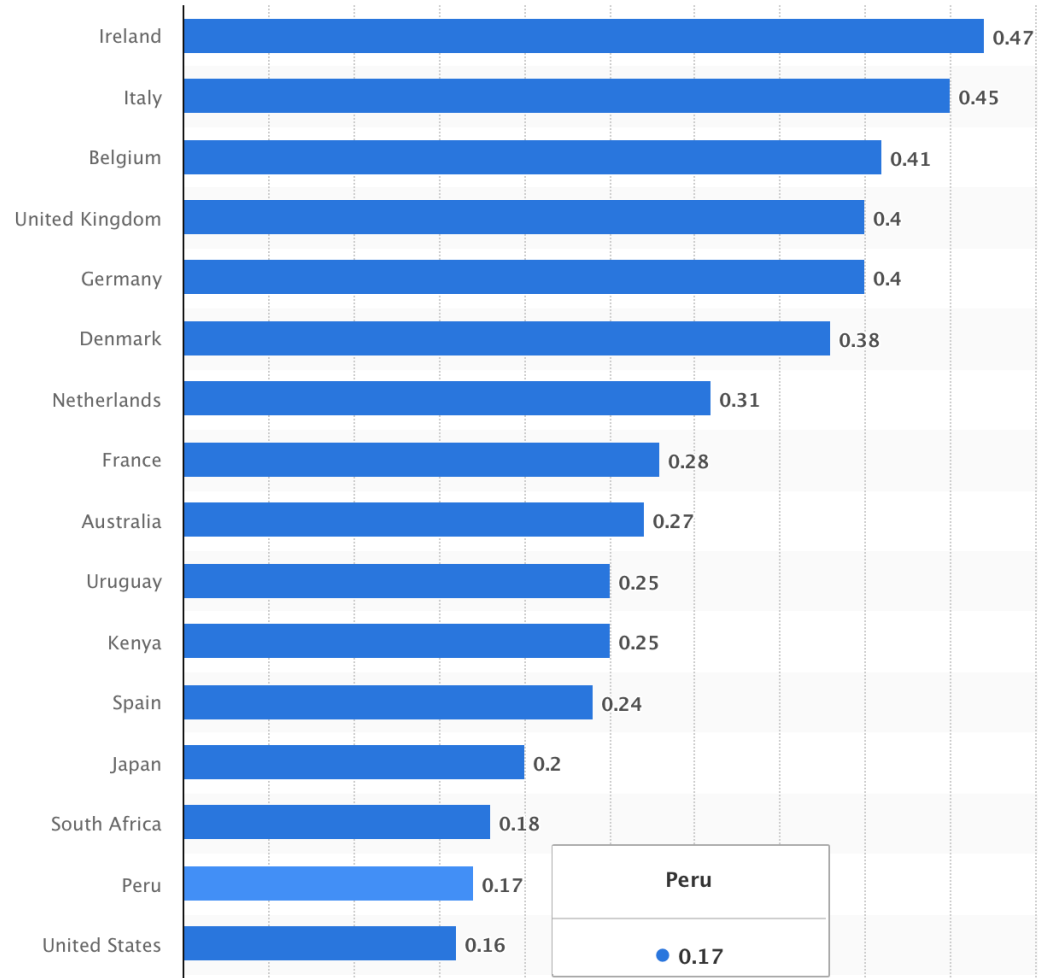
By [Rob Aitken](#), Fellow & Director of Technology, Arm

Share     

Reading time 7 min

Source: <https://newsroom.arm.com/blog/performance-per-watt>

Household electricity prices worldwide in December 2023, (in U.S. dollars per kilowatt-hour)

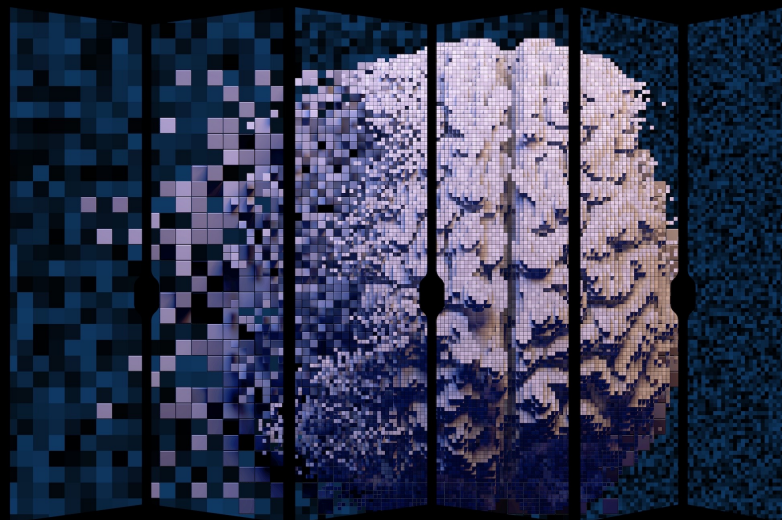


Why do we care
about
sustainable
computer
architectures? –
Cost

Source:

<https://www.statista.com/statistics/263492/electricity-prices-in-selected-countries/>

Isambard AI



Hewlett Packard
Enterprise

NVIDIA

arm

NATIONAL
COMPOSITES
CENTRE

UK Research
and Innovation

Department for
Science, Innovation
& Technology

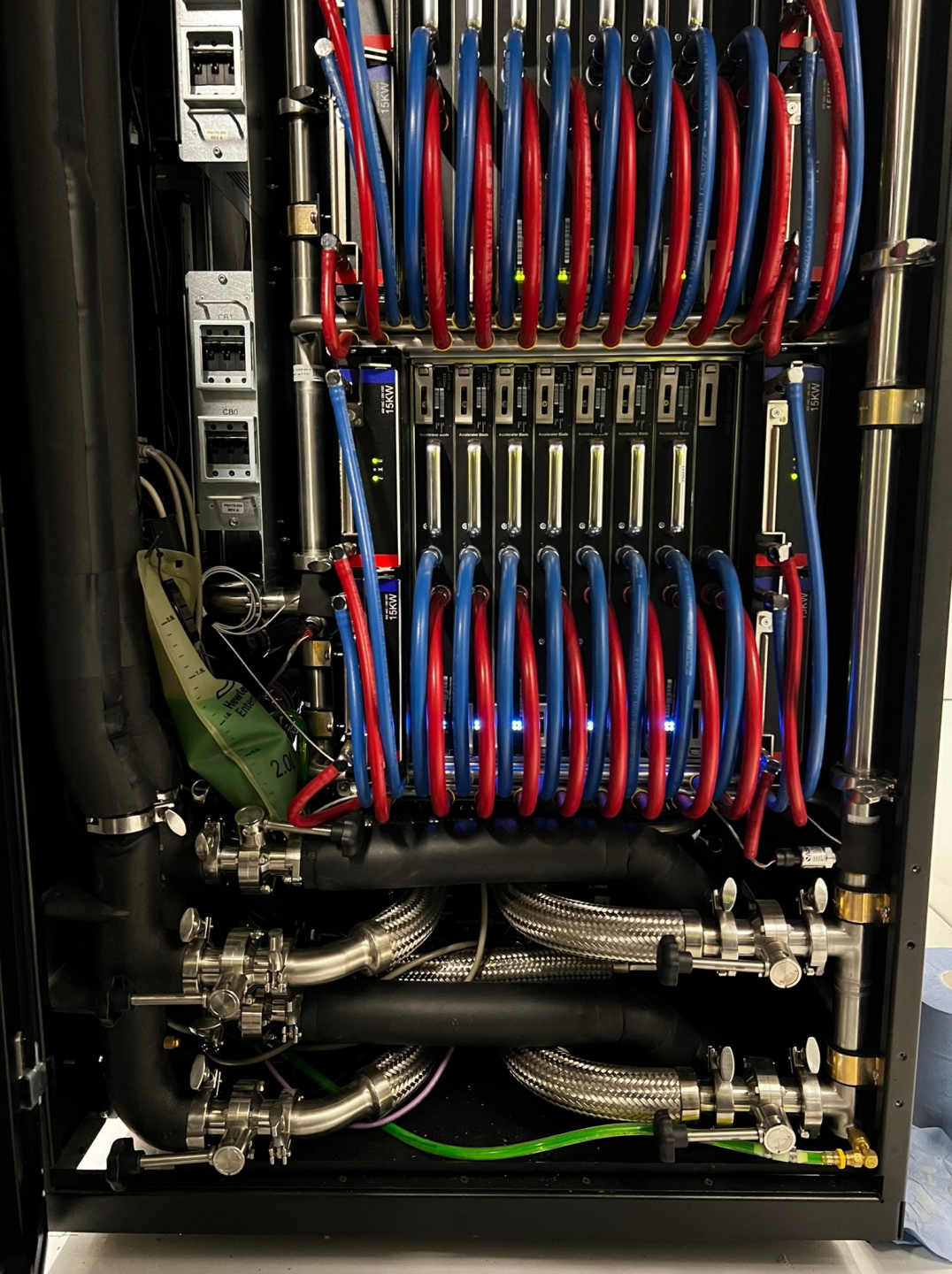
Isambard-AI



University of
BRISTOL

- **>£300M** investment by UK Government in AI capability
- Funding **~5,500 NVIDIA Grace-Hopper GPUs** in a new, **5MW** HPE modular data centre (MDC) facility in Bristol, UK
 - ~21 ExaFLOP/s of 8-bit for AI, ~250 PFLOP/s 64-bit
 - In Top 10 in the world
- **Extremely rapid deployment** a key requirement:
 - **First conversation** with UK Government on **Aug 18th 2023**
 - **>£200M procurement** written in **1 week**, run in just **2 weeks**
 - Contract signed and **ground broken** in November 2023
 - Site chosen for power availability and rapid planning permission procedures (8 weeks)





top500.org

June 2024 | TOP500

R_{max} and R_{peak} values are in PFlop/s. For more details about other fields, check the TOP500 description.

R_{peak} values are calculated using the advertised clock rate of the CPU. For the efficiency of the systems you should take into account the Turbo CPU clock rate where it applies.

Green500 Data

Rank	TOP500 Rank	System	Cores	Rmax (PFlop/s)	Power (kW)	Energy Efficiency (GFlops/watts)
1	189	JEDI - BullSequana XH3000, Grace Hopper Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail NVIDIA InfiniBand NDR200, ParTec/EVIDEN EuroHPC/FZJ Germany	19,584	4.50	67	72.733
2	128	Isambard-AI phase 1 - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE University of Bristol United Kingdom	34,272	7.42	117	68.835
3	55	Helios GPU - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Cyfronet Poland	89,760	19.14	317	66.948

TOP

Isambard-AI phase 2 system now completed testing in HPE's factory in the Czech Republic



Factory visit on June 11th 2024. 5,280 GPUs now all built, in Top10 in the world

5th Dec 2023 to 13th Aug 2024



Site Wednesday Aug 14th 2024



Simulating energy efficient architectures

- Wanted to enable research into processor microarchitecture
- Needed a fast, accurate, easy to use simulator framework
- Have developed the “Simulation Engine”, or **SimEng**
- First commit Dec 2018 (prototype work for 2 years before this)
- Initially focusing on the Arm ISA, now also RISC-V
- Being used in projects with partners including Arm, RIKEN on FugakuNEXT and SiPearl on European processors

SimEng enables **exploration of the design space** for future processors:

- **Core**
 - OoO and processor width, configuration of vector and matrix units, cache prefetch capability, branch predictors...
- **Co-processor**
 - Accelerators for vector–matrix maths, FFTs, ...
- **Memory hierarchy**
 - Smart prefetchers, sparse access support, ...
- **Network**
 - Network-on-Chip, inter-chip etc.

SimEng Design Goals



Fast – millions of Out-of-Order (OoO) instructions per second on a single core



Accurate – cycle counts typically within ~5-10% of real hardware



Easy to modify – just a few days to produce a radically different processor model



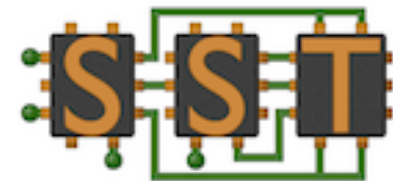
Use existing frameworks where possible

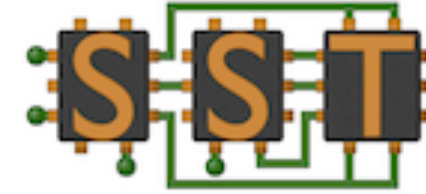
CAPSTONE for instruction decode, SST for memory hierarchy/multicore

Gem5-compatible tracing, checkpointing, ...

Currently Supported Features

- **AArch64** Armv9.2-a with SVE, SVE2 and SME extension support
 - Other extensions are supportable, but not yet targeted
 - ~1000 instruction variants supported with tests (~16% coverage)
 - Imminent SME2 support which extends the ISA from accelerating just GEMM to GEMV, CONV, etc.
- **RISC-V** rv64imafdc (base, mul/div, atomic, SP-FP, DP-FP, compressed)
 - 98%+ coverage
- Executes statically linked Linux ELF binaries
 - Support for the most common system calls
- Single-Thread OpenMP support
 - Multi-thread support running in prototype form

The ARM logo, consisting of the word "arm" in a blue, lowercase, sans-serif font.The RISC-V logo, featuring a stylized "R" icon in blue and yellow, followed by the text "RISC-V" in a blue, sans-serif font with a registered trademark symbol.The OpenMP logo, with the word "Open" in a teal, sans-serif font and "MP" in a larger, bold, teal, sans-serif font, underlined.

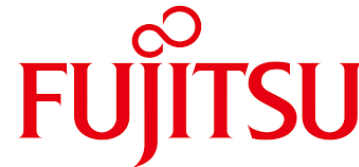


SimEng-SST Integration

- Infinite L1D Cache, supported by base SimEng but integration with the Structural Simulation Toolkit (SST) allows us to increase simulation accuracy and scope
- SST developed by Sandia National Laboratories
- Allows for much more complex memory hierarchy simulations
 - i.e. L1->L2->L3->DRAM (or HBM)
 - Configurable bandwidths + link delays
- SST drives SimEng simulation + handles cache coherency
- Multi-core model support in prototype form (ran 128 cores in parallel)
- Investigating current prefetcher models to improve accuracy in SST (Next-n-block, stride, stream)

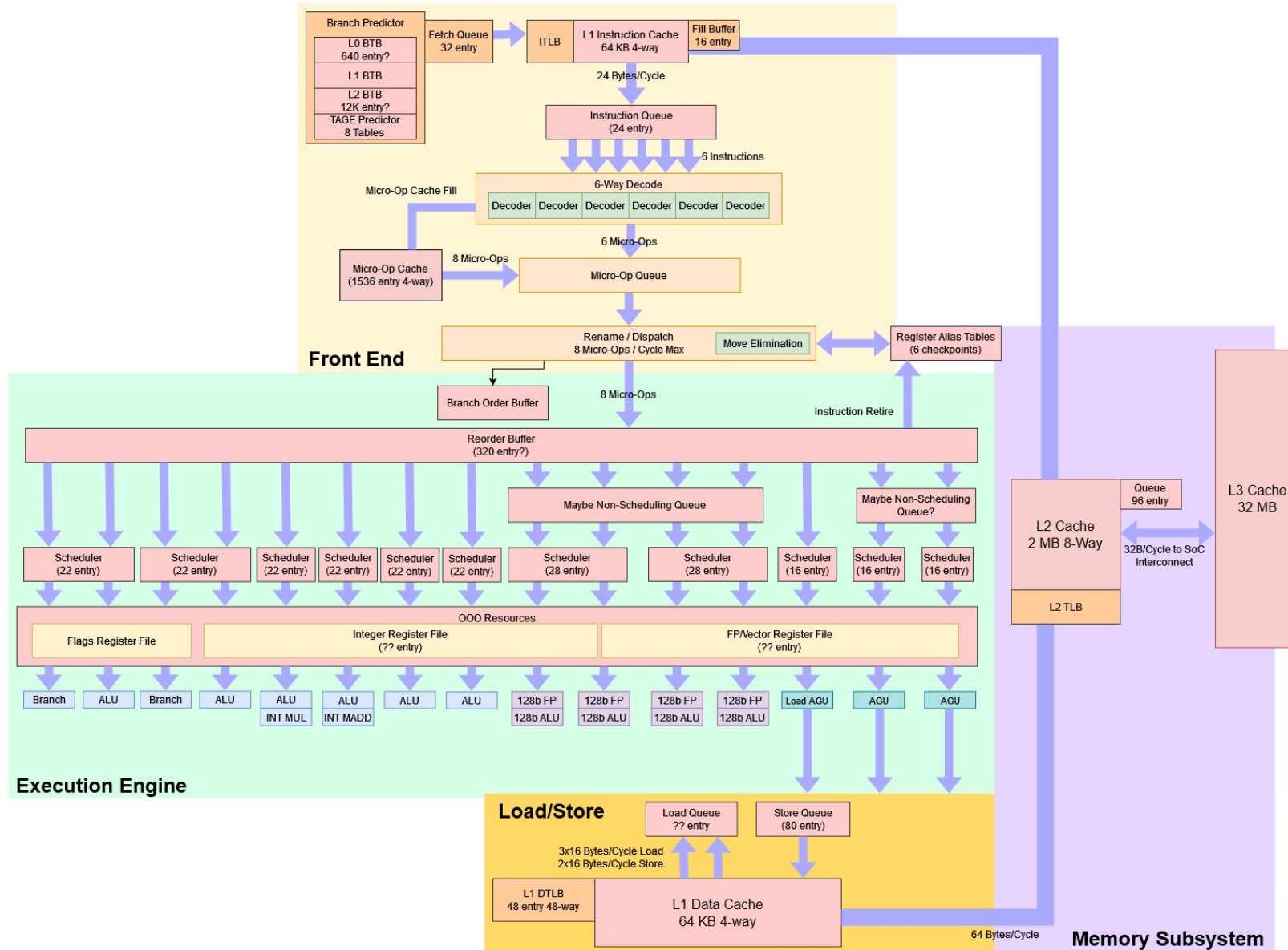
Core models already implemented in SimEng

- Marvell ThunderX2 – ARMv8.1
- Fujitsu A64fx – ARMv8.3
- Apple M1 Firestorm – ARMv8.4
- Arm Neoverse V2 (NVIDIA Grace, Graviton 4) – ARMv9.0
- RISC-V (generic OoO 4-way superscalar) – rv64imafdc



Neoverse V2

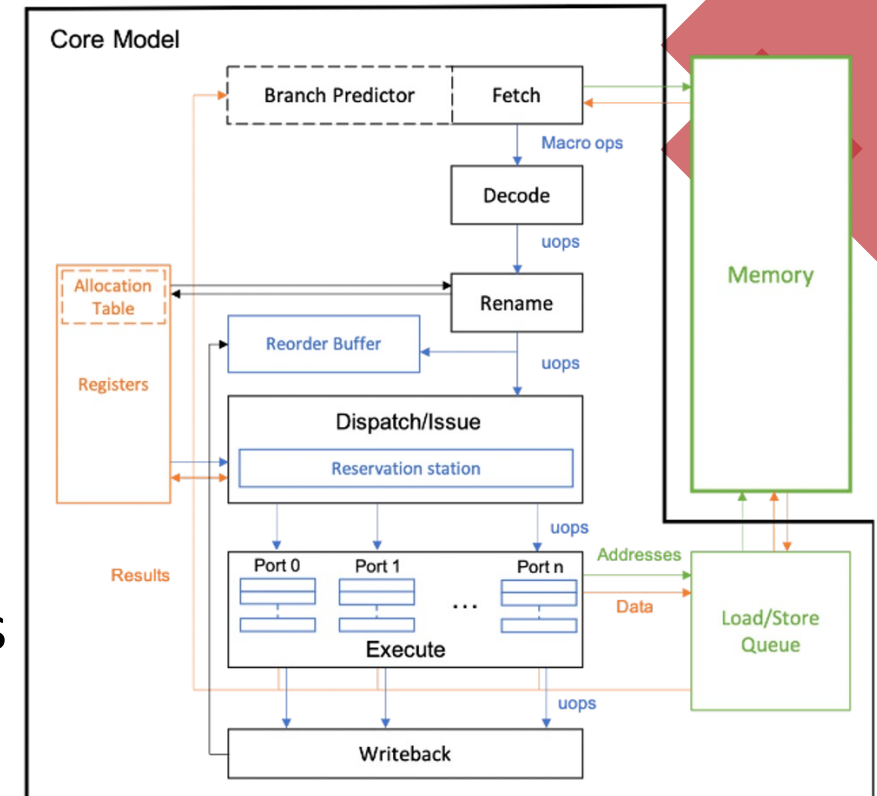
Diagram By Clamchowder



Some of the experiments enabled by SimEng

- Micro-architecture design studies
 - Execute pipeline length / width
 - Instruction splitting / merging
 - Reservation station configuration
 - OoO resource allocation (ROB, Phys. Registers, LSQ)
 - Number of vector units
 - Matrix engine composition
- Different branch prediction algorithms
- ISA comparisons across otherwise identical cores
- Cache prefetcher algorithms*
- TLB walks*

* In conjunction with SST





Prof. Simon McIntosh-Smith
PI



Mr Jack Jones
Lead Developer



Mr Finn Wilkinson



Mr Dan Weaver



Mr Alex Cockrean

<https://uob-hpc.github.io/SimEng>



Mr Joseph Moore

Useful Links + Contacts

- Simon McIntosh-Smith : s.mcintosh-smith@bristol.ac.uk
- Jack Jones : jj16791@bristol.ac.uk
- SimEng Repository : <https://github.com/UoB-HPC/SimEng>
- SimEng Documentation : <https://uob-hpc.github.io/SimEng/>
- PMBS-22 SME Evaluation Paper : <https://ieeexplore.ieee.org/document/10024029>
- ModSim-23 SME Evaluation Poster : https://uob-hpc.github.io/SimEng/_downloads/modsim23_poster.pdf
- Second International Workshop on RISC-V for HPC RISC-V vs AArch64 Comparison: <https://dl.acm.org/doi/abs/10.1145/3624062.3624233>