AI accelerators for HPC

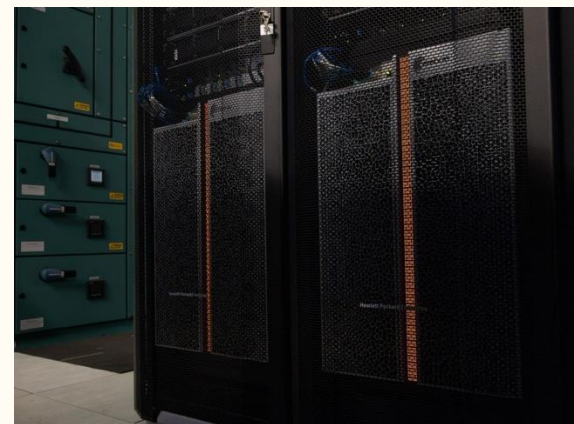Joseph Lee

EPCC

j.lee@epcc.ed.ac.uk

Who are we?

- Supercomputing centre at the University of Edinburgh, UK
- Providing world-class computing systems, data storage and support services
- Hosts the UK's national supercomputing service: ARCHER2

- ~150 staff
- Research, Education and Training: MSc and PhD programme

Who are we?

- Edinburgh International Data Facility (EIDF)
- High-powered data analytics and storage service that supports research and data-driven innovation
  - CPU & GPU
  - Cerebras CS-2 (x2)
  - Graphcore Bow Pod 64
- We have also been using these machines for HPC research

What are we looking at?

- ExCALIBUR (UK research programme for exascale algorithms and infrastructure)
- Enabling Hardware & Software Programme
    - Accelerators: RISC-V, FPGAs, Cerebras WSE

TensorFlow as a DSL for stencil-based computation on the Cerebras Wafer Scale Engine

Nick Brown[1], Brandon Echols[2], Justs Zarins[1], and Tobias Grosser[3]

- xDSL (Python x MLIR) for stencil computation
- PETSc on Cerebras
    - And more we want to do!
- Training: SDK tutorial at HPC Days – Durham UK

Experience

Positive 👍	Challenges 😩
Helpful support	Programming model/language (DSL)
Active community	Training & documentation
Tools (e.g. simulator)	Closed tools & reproducibility
Performance!	Big software changes
	Repeated/similar work (more collaboration?)

What are we looking for?

- Usable and accessible: Language, Library, Compiler
- Documentation and support
- AI Accelerator for HPC tutorials?
- Community!