LOCA SERIES. Reading Club. 5th session

The general objective of LOCA Technical sessions is to value the knowledge, talent and experience of the center's workers and researchers, and support the creation of a multidisciplinary work team based on the strengths of each and every one of its members.

To contribute to a sustainable and organic growth of the hardware design area at BSC, we propose organizing Internal LOCA Tech sessions. A monthly event whose objectives are: 1) promotion of internal BSC talent (Tech-Talk sessions), and 2) continuous learning/development mechanism (Reading club sessions).
? Tech-Talk session: Technical presentations on specific topics with the aim of putting the spotlight on the most senior and post-doc community's background and know-how, and also making visible the research and/or interests areas of work of each of those members.
? Reading club session: Technical sessions in the form of a presentation with the objectives of: 1) actively involving members with less experience in research activities, and 2) offering a "light-weight" training/retraining mechanism for all members that allows them to keep up to date with the state of the art.

Speakers:

Marco Siracusa

- **Talk title**: A Tensor Marshaling Unit for Sparse Tensor Algebra on General-Purpose Processors
- **Talk short abstract**: During this talk, I'll introduce the Tensor Marshaling Unit (TMU), a near-core programmable dataflow engine for multicore architectures that accelerates tensor traversals and merging, the most critical operations of sparse tensor workloads running on today's computing infrastructures. The TMU leverages a novel multi-lane design that enables parallel tensor loading and merging, which naturally produces vector operands that are marshaled into the core for efficient SIMD computation. The TMU supports all the necessary primitives to be tensor-format and tensor-algebra complete. We evaluate the TMU on a simulated multicore system using a broad set of tensor algebra workloads, achieving 3.6x, 2.8x, and 4.9x speedups over memory-intensive, compute-intensive, and merge-intensive vectorized software implementations, respectively.
- **Short bio**: Marco received his Bachelor's Degree in Computer, Electronic and Telecommunication Engineering in 2016 from the Università degli studi di Parma, Italy. He received a Master's degree in Computer Science and Engineering in 2020 from Politecnico di Milano, Italy. He is currently a first-stage researcher at the Barcelona Supercomputing Center and second-year PhD student at Universitat Politècnica de Catalunya. His research interests include Compiler Infrastructures, Computer Architectures, and High-Performance Computing.

Max Dobla

- **Title**: GMX: Instruction Set Extensions for Fast, Scalable, and Efficient Genome Sequence Alignment
- **Abstract:** Sequence alignment remains a fundamental problem in computer science with practical

applications ranging from pattern recognition to computational biology. To address these challenges, we propose GMX, a set of ISA extensions that enable efficient sequence alignment computations based on dynamic programming (DP). Furthermore, we provide an efficient hardware implementation that integrates GMX extensions in a RISC-V-based edge System-on-Chip (SoC). Compared to widely-used software implementations, our hardware-software co-design leveraging GMX extensions obtains speed-ups from 25-265x, scaling to megabyte-long sequences. Compared to Domain-Specific Accelerators (DSA), we demonstrate that GMX-accelerated implementations demand significantly less memory bandwidth, requiring less area per Processing Element (PE). As a result, a single GMX-enabled core achieves a throughput per area between 0.35-0.52x that of state-of-the-art DSAs while being more flexible and reusing the core's resources.

- **Short bio**: Max Doblas received a B.Sc. in electrical engineering and computer science from Universitat Politècnica de Catalunya (UPC) in 2020. He received an M.Sc. in computer science from UPC in 2021. He is currently a second-year Ph.D. in computer architecture with the UPC. He also works as a Research Engineer in the project Designing RISC-V-based Accelerators for next-generation Computers (DRAC) at the Barcelona Supercomputing Center (BSC), in which he has designed a power-efficient processor with several extensions for domain-specific applications. His research interests include high-performance computing architectures and domine-specific accelerators with applications to genomics, computational biology, and sequence alignment.

Speakers

Presenters: Marco Siracusa, first-stage researcher at the Barcelona Supercomputing Center and second-year PhD student at Universitat Politècnica de Catalunya. **Max Doblas**, Research Engineer in the project Designing RISC-V-based Accelerators for next-generation Computers (DRAC) at the Barcelona Supercomputing Center (BSC) and second-year Ph.D. in computer architecture with the UPC
Chair: Teresa Cervero, Leading Research Engineer. Computer Sciences - European Exascale Accelerator at BSC
Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---