[Times in EDT Timezone]

Agenda Overview

Session 1: I/O, Persistency and Filesystem – Chair: Antonio J. Peña, Barcelona Supercomputing Center (BSC)

9:00	J-NVM: Off-heap Persistent Objects in Java	Anatole Lefort, Telecom SudParis
9:30	Principles of TSOPER and application in ArgoDSM	Per Ekemark, Uppsala University
10:00	DAOS - Accelerating HPC/AI Storage through Persistent Memory	Michael Hennecke, Intel
10:30	Break	

Session 2: Extending RAM Space – Chair: Joao Pedro Barreto, Universidade de Lisboa and INESC-ID

11:00	HeMem: Scalable Tiered Memory Management for Big Data Applications and Real NVM	Amanda Raybuck, The University of Texas at Austin
11:30	Getting objects allocations and management right in heterogenous memory systems	Maciej Maciejewski, Huawei
12:00	Adding machine intelligence to hybrid memory management	Thaleia Doudali, IMDEA
12:30	Break	

Session 3: Extending RAM Space (II) and Panel – Chair: Harald Servat, Intel

14:00	CSMV: A Highly Scalable Multi-Versioned Software Transactional Memory for GPUs	Diogo Nunes, IST/INESC-ID
14:20	Introducing the ecoHMEM Framework for Data Placement in Heterogeneous Memory Systems	Marc Jordà, Barcelona Supercomputing Center (BSC)
14:40	Panel	Moderator: Petar Radojkovic, Barcelona Supercomputing Center (BSC) Panelists: All presenters
15:30	End	

Full Agenda

SESSION 1: I/O, Persistency and Filesystem

Chair: Antonio J. Peña, Barcelona Supercomputing Center (BSC)

9:00 – 9:30

Talk: J-NVM: Off-heap Persistent Objects in Java

This talk presents J-NVM, a framework to access efficiently Non-Volatile Main Memory (NVMM) in Java. J-NVM offers a fully-fledged interface to persist plain Java objects using failure-atomic blocks. This interface relies internally on proxy objects that intermediate direct off-heap access to NVMM. The framework also provides a library of highly-optimized persistent data types that resist reboots and power failures. We evaluate J-NVM by implementing a persistent backend for the Infinispan data store. Our experimental results, obtained with a TPC-B like benchmark and YCSB, show that J-NVM is consistently faster than other approaches at accessing NVMM in Java.

Speaker: Anatole Lefort, Telecom SudParis

Anatole is a fourth-year and final year Ph.D. student at Télécom SudParis - Institut Polytechnique de Paris, advised by Pierre Sutra and Prof. Gaël Thomas from the computer science departement. His current research focuses on bringing persistent programming to cloud workloads and computations. Exploring language support for PMEM and revisiting programming abstractions for distributed computing runtimes and frameworks.

9:30 – 10:00

Talk: Principles of TSOPER and application in ArgoDSM

TSOPER is an approach for hardware-based strict TSO persistency, presented in HPCA 2021. In this talk, I present the principles behind TSOPER and the application of them in the realm of distributed shared memory, specifically ArgoDSM.
TSOPER provides a TSO persistency model where we allow coalescing of writes in caches but hide the coalescing by forming atomic groups of cachelines that are persisted all at once. To achieve atomicity, we stage all writes from each group in a power-backed buffer, akin to Intel's WPQ/ADR. To order atomic groups between cores, the caches collectively track coherence transactions and enforce persistence in the same order – but decoupled form – consistency.
When moving to the distributed domain, it turns out that the principles of TSOPER are still applicable with some variation. For example, in ArgoDSM we need to deal with coalescing, and thus grouping, of not only writes but also entire data-race-free regions. Additionally, to maintain order between different nodes, we develop a protocol that allows nodes to track coherence, carried by the use of locks, and signal when groups are persisted.

Speaker: Per Ekemark, Uppsala University
Per Ekemark is a PhD student in Computer Architecture at Uppsala University, Sweden. His research focus is on efficient and productive use of persistent memory in both single and distributed systems.

He has previously also explored compiler optimisations involving software prefetching and data-race-freedom in conjunction with his Master's thesis work.

10:00 – 10:30

Talk: DAOS - Accelerating HPC/AI Storage through Persistent Memory

The Distributed Asynchronous Object Storage (DAOS) is an open source scale-out storage system that is designed from the ground up to support Storage Class Memory (SCM) and NVMe storage in user space. This talk introduces the DAOS storage architecture, presents theavailable APIs and DAOS-enabled middleware, and explains how DAOS uses transactions in Persistent Memory (based on PMDK's libpmemobj) to overcome the limitations of traditional parallel filesystems.

Speaker: Michael Hennecke, Intel

Michael is a Principal Engineer for HPC Storage in the Intel DAOS engineering team. He has over 28 years of experience in High Performance Computing and HPC Storage. Michael holds a master degree in physics from Ruhr-Universität Bochum (Germany), and a "Distinguished IT Specialist" certification from The Open Group.

10:30 – 11:00

Break

SESSION 2: Extending RAM Space

Chair: Joao Pedro Barreto, Universidade de Lisboa and INESC-ID

11:00 – 11:30

Talk: HeMem: Scalable Tiered Memory Management for Big Data Applications and Real NVM

HeMem is a tiered main memory management system designed from scratch for commercially available NVM and the big data applications that use it. HeMem manages tiered memory asynchronously, batching and amortizing memory access tracking, migration, and associated TLB synchronization overheads. HeMem monitors application memory use by sampling memory access via CPU events, rather than page tables. This allows HeMem to scale to terabytes of memory, keeping small and ephemeral data structures in fast memory, and allocating scarce, asymmetric NVM bandwidth according to access patterns.

Finally, HeMem is flexible by placing per-application memory management policy at user-level. On a system with Intel Optane DC NVM, HeMem outperforms hardware, OS, and PL-based tiered memory management, providing up to 50% runtime reduction for the GAP graph processing benchmark, 16% lower tail-latency under performance isolation for a key-value store, and up to 10x less NVM wear than the next best solution, without application modification.

Speaker: Amanda Raybuck, The University of Texas at Austin

Amanda Raybuck is a PhD student at the University of Texas at Austin. Her research interests are in modern memory management, particularly in the face of new hardware such as non-volatile memory.

11:30 – 12:00
Talk: Getting objects allocations and management right in heterogenous memory systems

Heterogenous memory systems are constantly evolving. From distant socket memory, through HBM to Intel Optane PMEM nowadays. With the enablement of cache coherency with CXL protocol, we can expect increasing popularity and impact of systems equipped with multiple different memory types. In this talk we shall analyze methods for deciding the data placement within memory subsystem. Along with questioning their strengths and weaknesses, we will try to find the optimal combination providing memory access performance to applications.

Speaker: Maciej Maciejewski, Huawei

Maciej Maciejewski is a software professional working in a high-tech industry since 16 years. He received MSc. in optoelectronics from the Gdansk University of Technology in 2002, where he undertook PhD in an Optical Coherence Tomography area. For eight years, he has worked at ADVA Optical Networking being an architect of distributed and stateless applications within network management systems area. For 6 years, he worked at Intel on adoption of Persistent Memory (PMEM) by software applications. Working at Intel, he lead multiple adaptations of databases, HPC applications, and cloud workloads to optimally utilize PMEM. Since 2021 he leads Memory Acceleration Technology Center at Huawei.

12:00 – 12:30

Talk: Adding machine intelligence to hybrid memory management

Nowadays, computing platforms use a mix of different hardware technologies, to scale application performance, resource capacities and achieve cost effectiveness. However, this heterogeneity, along with the greater irregularity in the behavior of emerging workloads, render existing resource management approaches ineffective. In the first part of this talk, I will describe how we can use machine learning methods at the operating system-level, in order to make more intelligent resource management decisions and speed up application performance. In the second part of the talk, I will present how we can accelerate certain components of such systems using smart and visual insights, as well as computer vision methods. Finally, I will conclude with remaining challenges in the practical use of machine learning at the system-level hybrid memory management.

Speaker: Thaleia Doudali, IMDEA

Thaleia Dimitra Doudali is an Assistant Research Professor at the IMDEA Software Institute in Madrid, Spain. She received her PhD from the Georgia Institute of Technology (Georgia Tech) in the United States. Prior to that she earned an undergraduate diploma in Electrical and Computer Engineering at the National Technical University of Athens in Greece. Thaleia's research lies at the intersection of Systems and Machine Learning, where she explores novel methodologies, such as machine learning and computer vision, to improve system-level resource management of emerging hardware technologies. In 2020, Thaleia was selected to attend the prestigious Rising Stars in EECS academic workshop. Aside from research, Thaleia actively strives to improve the mental health awareness in academia and foster diversity and inclusion.

12:30 – 14:00

Break

Session 3: Extending RAM Space (II) and Panel

Chair: Harald Servat, Intel

14:00 – 14:20

Short Talk: CSMV: A Highly Scalable Multi-Versioned Software Transactional Memory for GPUs

GPUs have traditionally focused on streaming applications with regular parallelism. Over the last years, though, GPUs have also been successfully used to accelerate irregular applications in a number of application domains by using fine grained synchronization schemes.
Unfortunately, fine-grained synchronization strategies are notoriously complex and error-prone. This has motivated the search for alternative paradigms aimed to simplify concurrent programming and, among these, Transactional Memory (TM) is probably one of the most prominent proposals.

This paper introduces CSMV (Client Server Multi-versioned), a multi-versioned Software TM (STM) for GPUs that adopts an innovative client-server design. By decoupling the execution of transactions from their commit process, CSMV provides two main benefits: (i) it enables the use of fast on chip memory to access the global metadata used to synchronize transaction (ii) it allows for implementing highly efficient collaborative commit procedures, tailored to take full advantage of the architectural characteristics of GPUs.

Via an extensive experimental study, we show that CSMV achieves up to 3 orders of magnitude speed-ups with respect to state of the art STMs for GPUs and that it can accelerate by up to 20× irregular applications running on state of the art STMs for CPUs.

Speaker: Diogo Nunes, IST/INESC-ID

Diogo Nunes obtained a MsC degree at IST in Lisbon, Portugal, in 2021. He is currently working as a junior researcher at IST/INESC-ID. His research interests include high performance computing, GPU programming and Transactional Memory.

14:20 – 14:40

Short Talk: Introducing the ecoHMEM Framework for Data Placement in Heterogeneous Memory Systems

Hybrid memory systems are an emerging trend to provide larger RAM sizes at reasonable cost and energy consumption. Recent byte-addressable persistent memory (PMEM) technology offers capacities comparable to storage devices and access times much closer to DRAMs than other non-volatile memory technology. To palliate the large gap with DRAM performance, DRAM and PMEM are usually combined. Users have the choice to either manage allocations to different memory

spaces manually or leverage the DRAM as a cache for the virtual address space of the PMEM. In this talk, we present the ecoHMEM framework for automatic object-level placement, which addresses the performance shortcomings of previous solutions, yielding a framework which is competitive in performance with respect to the state of the art, while enabling a much simpler workflow. Our experiments leveraging Intel Optane Persistent Memory show from matching to greatly improved performance with respect to state–of–the–art software and hardware solutions, attaining over 2x runtime improvement in miniapplications and over 6% in OpenFOAM, a complex production application.

Speaker: Marc Jordà, Barcelona Supercomputing Center (BSC)

Marc Jordà received his M.S. in Computer Architecture in 2012 from the Universitat Politècnica de Catalunya, Barcelona. Since then, he has been a research engineer at the Barcelona Supercomputing Center working in several topics related to high-performance computing, including application acceleration using GPUs, GPU hardware simulation, performance analysis, and automatic data placement for heterogeneous memory systems.

14:40 – 15:30

HMEM ’22 Panel

Moderator: Petar Radojkovic, Barcelona Supercomputing Center (BSC)

Petar Radojkovic is the Memory systems team leader at the Barcelona Supercomputing Center.
He is the PI of BSC collaboration with Micron Technologies US and Huawei China.
He was also leading a multi-project collaboration between the BSC and Samsung Electronics, Korea.

Panelists: All presenters

Registration

Please click here for registration page by ICS2022.