Research Statement - Thaleia Dimitra Doudali - thdoudali@gatech.edu

Problem Statement.

The exponential data growth in datacenter, high performance and exascale computing environments is creating a huge need to accelerate data access, storage and manipulation, at scale. Industry is responding via new classes of memory and storage technologies such as non-volatile memories (NVMs) or smart storage arrays, specialized accelerators such as TPUs and FPGAs, programmable network elements, and new standards [14] to interconnect these components into extremely scalable systems. The continued increase in demand for data and data analytics, coupled with the slowdown of Moore's Law, will continue to fuel this industry trend. As a result, future computing will encompass unprecedented levels of heterogeneity across the technology substrate, manifesting itself (simultaneously) across multiple dimensions such as cost, speed, capacity, reliability, persistence and consistency.

The extreme heterogeneity promises, though, to deliver factors or even orders of magnitude application speedup for emerging applications. However, leveraging its potential to realize desired performance or system efficiency levels, requires innovation across the stack to build solutions that transparently map compute and data to the appropriate underlying technologies, combining information regarding application level behavior, system level resource availability and hardware level properties. My dissertation research contributes data management methodologies and system level abstractions that boost application performance and system cost efficiency when the main memory subsystem consists of hybrid hardware technologies, such as DRAM and Non Volatile Memory (NVM). My long-term vision is to extend my research across the hardware layers (compute, memory, storage, network), especially with respect to identifying cross-layer abstractions needed to open up support for new types of interactions and information flows.

Contributions.

The focus on my dissertation research is on resource management for hybrid memory systems. In this area, existing solutions [7-10] are predominantly focusing on properly identifying the data that needs to be placed and migrated between DRAM and NVM, so as to speed up most of the memory accesses. Without such solutions the application performance can be reduced by several factors, depending on the underlying memory access patterns.

My most recent publication explores the role of Machine Intelligence in management of memory allocation and migration, the impact it can have on performance enhancements, and the feasibility of realizing solutions which can be deployed in practice. This is a very timely research question, in an era where machine learning approaches are extremely popular and show promise to solve similar system problems such as cloud resource scheduling [5] or data prefetching [6]. To this extent I built Kleio: a hybrid memory page scheduler with machine intelligence [1], a **best paper award finalist** at HPDC '19. Kleio is able to bridge by 80% the performance gap that exists between current state-of-the-art and oracular solutions. Also, Kleio is designed in a way that lays the grounds for its practical integration into future systems, because it cleverly identifies a small subset of pages whose timely allocation in DRAM via machine intelligent management increases application performance. This, combined with

future generation of accelerators, will open up new opportunities for use of machine intelligence in online, dynamic resource management tasks across the software stack.

My earlier work proposes new questions we need to address in the research area of hybrid memory systems. While other work is focused on methodologies for boosting performance of *single* applications in systems with *fixed capacities* of DRAM and NVM, my research introduces two new dimensions of the problem space: capacity sizing and efficient resource sharing across applications. First, we need to have a way to understand how much capacity of DRAM and NVM we need, given a cost budget and desired performance levels. Our findings illustrate opportunities for significant reduction of the memory system cost with a managed, and in some cases only trivial, impact on performance. Using key-value stores as an application driver, due to their popularity and high data capacity demand, I built Mnemo [2], a tool that automatically discovers insights into the impact of hybrid memory capacity sizing on the workloads' cost-performance tradeoffs, which can be of tremendous practical value as cloud systems start introducing different types of memory technologies in their infrastructure offerings [11,12]. Second, we need to have an efficient way to distribute the memory resources across collocated applications or workload components. I built CoMerge [3], a methodology that dynamically adjusts the memory resources and maximizes the resource utilization as well as aggregate application performance by several factors, compared to static solutions. This functionality is valuable both in multi-tenant datacenters, and in supercomputing systems where compute nodes are shared among complex simulation and analytics workflow components.

The next steps in my dissertation research are tackling questions surrounding coordinated management of memory and interconnect resources in hybrid memory systems or rack-scale systems with disaggregated, "far" memories [4]. In future systems, with more profound differences in the memory capabilities of different components , and the data paths that interconnect them, current solutions designed for small scale NUMA topologies can have a detrimental impact on both application performance and resource efficiency. My goal is with my ongoing research to contribute new abstractions, mechanisms and algorithms that address the missing system-level decision support for coordinated memory fabric management.

Future Directions

I look forward to a career in an academic research environment, where I can collaborate with experts in programming languages, compilers, computer architecture and databases, and build effective cross-stack solutions. Many emerging hardware technologies are only now beginning to enter the commercial market [11] and the design specifications of datacenter [12] and exascale [13] systems, restructuring traditional architectures into memory centric topologies [14]. As the heterogeneity and scale of these systems increases, new online, lightweight, practical and intelligent solutions will become ever more critical for closing the performance and efficiency gaps left by existing approaches. My technical background and preparation, the breath of my current research, and my collaborative approach to research, make me well-positioned to succeed in making significant future contributions in this space.

References

[1] Thaleia Dimitra Doudali, Sergey Blagodurov, Abhinav Vishnu, Sudhanva Gurumurthi, and Ada Gavrilovska. 2019. Kleio: A Hybrid Memory Page Scheduler with Machine Intelligence. In The 28th International Symposium on High-Performance Parallel and Distributed Computing (HPDC '19), June 22–29, 2019, Phoenix, AZ, USA. ACM, New York, NY, USA, 12 pages.

[2] Thaleia Dimitra Doudali and Ada Gavrilovska. Mnemo: Boosting Memory Cost Efficiency in Hybrid Memory Systems. In proceedings of the 5th IEEE International Workshop on High-Performance Big Data, Deep Learning, and Cloud Computing (HPBDC 2019). In conjunction with the 33rd IEEE International Parallel and Distributed Processing Symposium (IPDPS 2019). Rio de Janeiro, Brazil, May 2019.

[3] Thaleia Dimitra Doudali and Ada Gavrilovska. 2017. CoMerge: toward efficient data placement in shared heterogeneous memory systems. In Proceedings of the International Symposium on Memory Systems (MEMSYS '17). ACM, New York, NY, USA, 251-261.

[4] Thaleia Dimitra Doudali and Ada Gavrilovska. Network-aware Data Management for Disaggregated Memory Systems. Poster in the PhD Forum of the 33rd IEEE International Parallel and Distributed Processing Symposium (IPDPS 2019).

[5] Ana Klimovic, Heiner Litz, and Christos Kozyrakis. 2018. Selecta: Heterogeneous Cloud Storage Configuration for Data Analytics. In Proceedings of the 2018 USENIX Conference on Usenix Annual Technical Conference (USENIX ATC '18). USENIX Association, Berkeley, CA, USA, 759–773.

[6] Hashemi, M., Swersky, K., Smith, J., Ayers, G., Litz, H., Chang, J., Kozyrakis, C. & Ranganathan, P.. (2018). Learning Memory Access Patterns. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:1919-1928

[7] Subramanya R. Dulloor, Amitabha Roy, Zheguang Zhao, Narayanan Sundaram, Nadathur Satish, Rajesh Sankaran, Jeff Jackson, and Karsten Schwan. 2016. Data tiering in heterogeneous memory systems. In Proceedings of the Eleventh European Conference on Computer Systems(EuroSys '16). ACM, New York, NY, USA, Article 15, 16 pages.

[8] Du Shen, Xu Liu, and Felix Xiaozhu Lin. 2016. Characterizing emerging heterogeneous memory. In Proceedings of the 2016 ACM SIGPLAN International Symposium on Memory Management (ISMM 2016). ACM, New York, NY, USA, 13-23.

[9] Kai Wu, Yingchao Huang, and Dong Li. 2017. Unimem: runtime data managementon non-volatile memory-based heterogeneous main memory. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC '17). ACM, New York, NY, USA, Article 58, 14 pages

[10] Kai Wu, Jie Ren, and Dong Li. 2018. Runtime data management on non-volatile memory-based heterogeneous memory for task-parallel programs. In Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC '18). IEEE Press, Piscataway, NJ, USA, Article 31, 13 pages.

[11] https://www.intel.com/content/www/us/en/architecture-and-technology/intel-optane-technology.html

[12]https://www.intel.com/content/www/us/en/products/docs/memory-storage/optane-persistent-memory/google-partner-video.html

[13] https://www.olcf.ornl.gov/summit/

[14] http://genzconsortium.org/