Research Statement - Thaleia Dimitra Doudali - thdoudali [at] gatech.edu

Problem Statement

The explosion of data in high performance computing environments and datacenters creates the need for acceleration of compute and data retrieval times. The industry is responding via new classes of memory and storage technologies, such as non-volatile memories (NVMs) [13], which have already become part of the design specifications of datacenter [14] and exascale [15] systems. These new memory systems are coupled with specialized accelerators such as TPUs and FPGAs, programmable network elements, and new standards [16] to interconnect these components into extremely scalable systems, that promise 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 NVM.

Contributions

The focus of my dissertation research is on intelligent and efficient resource management for hybrid memory systems. In this area, existing solutions [9-12] are predominantly focusing on properly identifying the data that needs to be placed and moved across heterogeneous memory components (such as 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 recent publication explores the role of machine intelligence in the management of memory allocation and migration, the impact it can have on performance enhancements, and the feasibility of realizing such solutions 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 [7] and data prefetching [8]. To this extent, I built Kleio: a hybrid memory page scheduler with machine intelligence [4], a **best paper award finalist** at HPDC ‘19. Kleio is able to bridge up to 80% of the performance gap that exists among 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 the use of machine intelligence in online, dynamic resource management tasks across the software stack.

My earlier work identifies new questions we need to address in the research area of hybrid memory systems. While other work is focused on methodologies for boosting the performance of *individual* 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 [5], 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 [13,14]. Second, we need to have an efficient way to distribute the memory resources across collocated applications or workload components. I built CoMerge [6], 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. Third, we validated some of the simulation-based observations made in the CoMerge work on systems with real persistent memories, once Intel's Optane PMEM modules became available [13]. Interestingly, we also observed inconsistencies in some of the experimental trends on the real system; this led to new insights on the impact of existing persistent memories on application performance and on the design and configuration of mainstream operating systems support for these new memory components [3].

For the last part of my thesis I have been developing for now a tool [1], later hopefully an online system-level solution, that automates the configuration of tunable parameters of the operating system-level memory management stack. Particularly with respect to data movement, parameters such as the periodicity of the access scanning, frequency and thresholds of page migration operations, etc., are currently all empirically configured. Even considering recent work on heterogeneous or disaggregated memory management, the question of parameter configuration is glossed over. The current Linux policies seem to work reasonably well for common enterprise workloads and small scale NUMA systems, but we have shown that there is a significant loss in attainable performance and platform efficiency when considering data-intensive workloads, and when considering memory heterogeneity or disaggregation [2].

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. 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 breadth 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, Daniel Zahka, Ada Gavrilovska. Cori: Dancing to the Right Beat of Periodic Data Movements over Hybrid Memory Systems. In preparation. September 2020.

[2] Thaleia Dimitra Doudali, Daniel Zahka, Ada Gavrilovska. The Case for Optimizing the Frequency of Periodic Data Movements over Hybrid Memory Systems. To appear in MEMSYS 2020.

[3] Unexpected Performance of Intel Optane DC Persistent Memory. Tony Mason, Thaleia Dimitra Doudali, Margo Seltzer, Ada Gavrilovska. In IEEE Computer Architecture Letters, vol.19, no.1, pp.55-58, 1 Jan.-June 2020.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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

[9] 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.

[10] 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.

[11] Kai Wu, Yingchao Huang, and Dong Li. 2017. Unimem: runtime data management on 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

[12] 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.

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

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

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

[16] http://genzconsortium.org/