Han Dong
Research Statement

Boston University
Fall 2024

INTRODUCTION

Enhancing the efficiency of computer systems has far-reaching impacts on many aspects of our lives. For instance, accelerating high-performance computing (HPC) applications can drive advances in fields like drug discovery and scientific simulations, while more efficient data analytics can lead to insights that improve our health and safety. Efficiency in computing is growing even more important due to climate change, as the accelerated growth and use of Cloud computing can have significant global impacts. These impacts include increases in power consumption, carbon emissions, hazardous material used in manufacturing and e-waste. My research focuses the construction of novel, efficient, high-performance computer systems that can address the ever-growing needs of modern Cloud applications. My research approach revolves around conducting rigorous quantitative analysis of applications and systems in order to understand their limitations and opportunities for growth. Then, I design solutions targeting those opportunities through systems optimization or leveraging advances in machine learning (ML). Throughout my work, I have successfully collaborated with scientists across diverse fields, including physics and radiology, as well as with industry partners.

ENHANCING EFFICIENCY IN PARALLEL COMPUTE

My research career began during my undergraduate studies when I volunteered for a research class project focused on a novel parallel programming language. I successfully demonstrated its efficacy in generating high-performance code for distributed applications [1]. This experience fueled my desire to continue working in computational science. As an intern at Los Alamos National Laboratory, I worked alongside physicists and mathematicians to parallelize a physics simulator [2]. Later, during my master's, I combined software and hardware optimizations to accelerate a NASA climate physics program [3]. For both these projects, I collaborated with experts from diverse fields, learning to translate their algorithms into more efficient, high-performance versions.

ENHANCING EFFICIENCY IN DATA ANALYTICS

During my master's, I also became interested in how system efficiency can enhance data analytics to deliver valuable insights. Towards this, I developed systems to process social media data, aiming to provide accurate, real-time "boots on the ground" information to support evacuation efforts and emergency services during major weather disasters [4][5]. Later, as a researcher at the Oak Ridge Institute for Science and Education at the US Food and Drug Administration, I collaborated with radiologists and led efforts to create a sensor-based system that tracks the real-time location and posture of technicians and nurses in medical imaging rooms in order to reduce their exposure to radiation [6]. The International Commission on Radiological Protection recognized this work as a model for future research to minimize radiation exposure within medical communities [13].

ENHANCING EFFICIENCY IN OPERATING SYSTEMS

Reflecting on my previous experiences, my motivation for pursuing a PhD was to delve deeper into the systems stack, with a quantitative focus at the boundary between the Operating System (OS) software and hardware. In EbbRT [7], I developed benchmarks to help explain its performance improvements and lead efforts to optimize EbbRT to run directly and efficiently on hardware. This involved developing a network device driver from scratch and aggressively optimizing custom OS-network paths for various applications. These contributions enabled EbbRT to achieve significant performance and energy efficiency gains over Linux by 2X. In SEUSS [8], I built the virtual machine infrastructure for deploying thousands of Linux containers, enabling comparisons with a modern serverless system and ensuring experimental reproducibility. Both OSes have also been taught as part of upper-level computer systems courses at institutions such as Univ. of Waterloo, IIT Delhi, Virginia Tech, Harvard, etc. As part of the tradition in my lab, we often extracted self-contained components of these larger systems in order to have undergraduate and graduate involvement in research and I have worked closely with these students as part of their class projects. At Hamilton, I intend to replicate this approach to promote student engagement in research.

Building on my interest in using empirical data to identify and optimize system efficiency, as part of my dissertation, I conducted a data-driven study on how different hardware settings can be tuned to improve application and OS energy efficiency by 2X [9].Subsequently, I collaborated with researchers at Red Hat to develop an open-source tool that leverages ML to automatically tune these parameters [10]. I have also been invited to present this work at industry conferences like NetDev and DevConf. In my postdoctoral research, I am serving as co-PI on three Red Hat-funded projects aimed at quantifying and improving performance and energy efficiency of Cloud-based systems and large-scale inferencing systems. Recently, in [11], I demonstrated that optimizing OSes on older hardware can actually outperform modern OSes on the latest hardware for Cloud applications, suggesting new approaches to address sustainability in computing.

RESEARCH AGENDA

Anthropogenic climate change poses one of the greatest threats to both us and our planet's health. Currently, the computing industry accounts for around 3.9% of global carbon emissions and this is expected to grow [14]. Recent advancements in generative AI have intensified this trend, as their resource demands are expected to accelerate emissions growth, potentially by orders of magnitude. The research agenda below represents my plan to reimagine computer systems with sustainability at their core.

Reduce: Motivated by my doctoral research, which revealed that a data-driven analysis of an OS can uncover fundamental settings in hardware and software that significantly enhance application performance while reducing energy consumption, I believe the next logical steps over the next 1-2 years is to investigate the feasibility of applying these insights to other systems.Though I leveraged ML to automatically tune these settings, I am also interested in exploring whether ML can be employed to discover new settings across existing and emerging systems.

Recycle: My latest work [11] has shown that not all applications need the latest and greatest hardware; in fact, using older, deprecated hardware can be more sustainable, as it avoids the carbon emissions associated with manufacturing new components. This insight paves the way for innovative approaches to integrating refurbished and recycled hardware into modern datacenter hardware. One of my postdoctoral projects has revealed that there are opportunities to develop new scheduling strategies that improve sustainability in Cloud applications without compromising performance. For example, by building on the applied ML tools in my previous work, I plan to explore how to effectively apply it towards sustainable schedulers that take advantage of characteristics in old and new hardware platforms.

Reuse: To date, there have been upwards of hundreds of research systems developed to improve the performance of various fields within computer science [12]. However, the actual impact of these advancements on the technology industry remains unclear, largely due to differing incentive structures. With the growing need for sustainability in computing as a whole, I believe there is a huge opportunity to reuse and recontextualize many of the advantages shown in these systems in a new light. One of my interests is to explore the development of a general purpose OS that can seamlessly integrate techniques from these diverse systems into existing software ecosystems, making them more practical and easier to adopt. This would be a long-term, ongoing project that could potentially be part of upper-level course projects, allowing students to gain hands-on experience in cataloging and reproducing results from advanced research systems.

Based on my previous experience in undergraduate and graduate mentorship, my goal is to develop similar self-contained class projects derived from my research agenda that can be stitched into the various courses I am prepared to teach. Through these projects, I aim to instill in my students' a rigorous and quantitative approach to understanding and measuring computer systems, as well as the skills needed to explore optimization and applied ML approaches to systems. Given the challenges of sustainable computing, where even small efficiency gains can significantly impact future carbon emissions, I believe that fostering a sustainability mindset in students will be a tremendous asset to their future careers.

REFERENCES

[1] **Han Dong**, Shujia Zhou, and David Grove. X10-enabled MapReduce. In Proceedings of the Fourth Conference on Partitioned Global Address Space Programming Model (PGAS '10). Association for Computing Machinery, New York, NY, 2010.

[2] Jeff Willert, C. T. Kelley, D. A. Knoll, **Han Dong**, Mahesh Ravishankar, Paul Sathre, Michael Sullivan, William Taitano. Hybrid Deterministic/Monte Carlo Neutronics using GPU Accelerators. Proceedings of 11th International Symposium on Distributed Computing and Applications to Business, Engineering & Science, Guilin, Guangxi, China, 2012.

[3] **Han Dong**, Dibyajyoti Ghosh, Fahad Zafar, Shujia Zhou. Cross-Platform OpenCL Code and

Performance Portability for CPU and GPU Architectures Investigated with a Climate Physics Model. Proceedings of Fifth International Workshop on Parallel Programming Models and Systems Software for High-End Computing, Pittsburgh, PA, 2012.

[4] **Han Dong**, Milton Halem and Shujia Zhou. Social Media Data Analytics Applied to Hurricane Sandy. 2013 International Conference on Social Computing, VA, USA, 2013.

[5] Yin Huang, **Han Dong**, Yelena Yesha and Shujia Zhou. A Scalable System for Community Discovery in Twitter During Hurricane Sandy. 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, IL, USA, 2014.

[6] Andreu Badal-Soler, Fahad Zafar, **Han Dong**, Aldo Badano. A real-time radiation dose monitoring system for patients and staff during interventional fluoroscopy using a GPU accelerated Monte Carlo simulator and an automatic 3D localization system based on a Kinect depth camera. Proceedings of SPIE (The International Society for Optics and Photonics) Medical Imaging. Florida, USA, 2013.

[7] Dan Schatzberg, James Cadden, **Han Dong**, Orran Krieger, and Jonathan Appavoo. EbbRT: A Framework for Building Per-Application Library Operating Systems. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), Savannah, GA, USA, 2016.

[8] James Cadden, Thomas Unger, Yara Awad, **Han Dong**, Orran Krieger, and Jonathan Appavoo. SEUSS: skip redundant paths to make serverless fast. In Proceedings of the Fifteenth European Conference on Computer Systems (EuroSys '20). Association for Computing Machinery, New York, NY, USA, 2020.

[9] **Han Dong**, Sanjay Arora, Yara Awad, Tommy Unger, Orran Krieger, Jonathan Appavoo. Slowing Down for Performance and Energy: An OS-Centric Analysis of Network Applications. https://arxiv.org/abs/2112.07010, 2021.

[10] **Han Dong**. Tuning Linux kernel policies for energy efficiency with machine learning. https://research.redhat.com/blog/article/tuning-linux-kernel-policies-for-energy-efficiency-with-machine-learning/. May, 2023.

[11] **Han Dong**, Sanjay Arora, Orran Krieger, and Jonathan Appavoo. Can OS Specialization give new life to old carbon in the cloud? In Proceedings of the 17th ACM International Systems and Storage Conference (SYSTOR '24). Association for Computing Machinery, New York, NY, USA, 2024.

[12] National Science Foundation. CNS: Computer Systems Research (CSR). https://www.nsf.gov/awardsearch/advancedSearchResult?ProgEleCode=735400&BooleanElement=Any&BooleanRef=Any&ActiveAwards=true#results

[13] López PO, Dauer LT, Loose R, et al. ICRP Publication 139: Occupational Radiological Protection in Interventional Procedures. Annals of the ICRP. 2018;47(2):1-118. doi:10.1177/0146645317750356

[14] Charlotte Freitag, Mike Berners-Lee, Kelly Widdicks, Bran Knowles, GordonS. Blair, and Adrian Friday. 2021. The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. Patterns 2, 9 (2021), 100340. https://doi.org/10.1016/j.patter.2021.100340