



COLLEGE OF ENGINEERING

Department of Electrical & Computer Engineering
2120 Electrical and Computer Engineering Building, MC-702
306 N. Wright St.
Urbana, IL 61801-2991

December 30, 2023

To the Faculty Search Committee:

It is my great pleasure to recommend Dr. Mert Hidayetoglu for an Assistant Professor position in your department. Mert has demonstrated exceptional research abilities, and his Ph.D. thesis has already made significant contributions to the high-performance computing (HPC) community and beyond. After Mert received his Ph.D. degree in July 2022, he has been a postdoctoral scholar at Stanford University. My professional relationship with Mert as his Ph.D. thesis advisor has given me profound insight into his exceptional talents, innovative research approaches, outstanding research skills and his passion to advance the capabilities and applications of HPC systems.

Mert's Ph.D. thesis is in the field of large-scale supercomputing with an emphasis on solving inverse problems in computational imaging. He has a rare combination of solid expertise in mathematical physics, parallel algorithms, and parallel computing software infrastructure. He has well-positioned himself to tackle the computation and communication bottlenecks of applications such as full-wave image reconstruction using fast algorithms running on massively parallel supercomputing systems. His work pushes the frontiers of both computer science and computational physics. His dissertation makes two major contributions. First, he develops a novel sparse matrix tiling algorithm to accelerate repetitive Sparse-Matrix-Dense-Matrix Multiplication (SpMM) operations on GPUs. The applications include scientific, ML, and graph analytics workloads. This novel algorithm shifts these applications from memory-bandwidth bound towards compute bound, and can drastically improve their execution throughput on GPUs. Second, he proposes a novel hierarchical communication strategy to take advantage of multi-GPU node architectures with fast intra-node communication fabric like NVLinks, which significantly improves the scaling of communication-dominated sparse applications on exascale computers.

In his first years with me, Mert developed and implemented novel parallel algorithms on NCSA's Blue Waters supercomputer, where he obtains tomographic images at record-breaking scales (up to 4,096 GPUs), and in near-real time. He overcame the data communication bottleneck of Multiple Level Fast Multipole Algorithm (MLFMA) (due to its $O(N)$ computational complexity) by reformulating the MLFMA operations into matrix multiplication to enable sophisticated tiling and access pattern optimizations on GPUs. He also cleverly hid the MPI communications completely behind GPU computations for obtaining additional speedup. He made several conference presentations on various aspects of this work, such as multiple-scattering imaging and scalable parallelization of the distorted-Born iterative method. His computer science contributions led to a publication at IPDPS, a highly competitive venue, in 2018.

In the Summer of 2018, Mert was awarded with the prestigious Givens Fellowship by Argonne National Laboratory for a summer internship. At Argonne, he conducted interdisciplinary work and developed a novel memory-centric ultra-high-resolution X-ray image reconstruction method and implemented it on

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

217.333.2300 • ece.illinois.edu

Argonne's supercomputer. Mert's work is motivated by the need to process the extremely large measurements taken at the Advanced Photon Source (APS), which is a synchrotron light source to produce high-energy X-rays beams capable of imaging objects such as semiconductor chips and mouse brains at sub-micron-level resolution. Reconstructing images at such extreme level of resolution offers invaluable insights into the objects under study but requires efficient and accurate transformations of terabytes to petabytes of data. During his internship, he contributed to the development of efficient reconstruction algorithms and a journal paper published at Applied Optics. After his internship, with a great teamwork led by Mert, his Argonne work produced an excellent publication at the top supercomputing conference SC19. His code for the SC19 paper was also chosen as the reproducibility benchmark application for the student cluster competition at SC20. Teams of students from top universities worldwide used Mert's code as a benchmark to test their system performance while reproducing Mert's scientific results.

Mert's follow-up paper which he served as the lead author on the X-ray imaging application that he developed at Argonne was published at SC20 and won the Best Paper award, which is highly coveted recognition. This paper describes an innovative hierarchical communications scheme for large-scale computer systems composed of multi-GPU nodes (fat-node architecture), which is an integral part of his thesis, where he reduced the communication time cost by 60% on Summit in Oak Ridge National Laboratory. His SC20 paper demonstrated ultra-high-resolution 3D image reconstruction by solving an extremely large sparse equations at unprecedented scale on 24,576 GPUs of Summit, the world's fastest computer at the time. His success has been highlighted not only in Illinois and IBM newsletters, but also in magazines such as Scientific Computing World, HPC Wire, and EurekaAlert (AAAS). With the same work, he won 1st place at the ACM student research competition at SC20. Mert's production code has been used at Argonne for the image reconstruction with next-generation light source and with multiple exascale supercomputers. Mert is working with his collaborators not only at Argonne, but also with other DOE national laboratories to port and optimize his application to the most advanced supercomputers in the world.

Additionally, Mert contributed heavily to the IBM-Illinois Center for Cognitive Computing Systems Research (C3SR). As a senior Ph. D. student, Mert led the effort in C3SR for efficient sparse deep neural network (DNN) inference. It was a relevant and timely challenge since de-facto standard frameworks like PyTorch and TensorFlow lacked optimized implementations. The main bottleneck we faced was that GPUs' performance suffers in sparse computations defined in the form of a graph. Mert has the right skills to overcome this performance bottleneck as a pioneer. He designed and implemented novel techniques to maximize inference throughput with minimum power-delay product. He has also implemented a load balancing technique and has scaled sparse DNN inference on the Summit supercomputer. His paper at HPEC'20 won the MIT/Amazon/IEEE Graph Challenge Championship.

Mert has also been actively contributing to GPU education worldwide by assisting PUMPS summer school in Barcelona since 2017. He has been giving lectures and offering research clinic sessions on programming large-scale computing systems that consist of multi-GPU nodes. In 2021, Mert helped updating the curriculum and organizing lectures of the PUMPS summer school into PUMPS+AI 2021. At UIUC, he has also been giving guest lectures to the ECE 508 graduate-level class.

To summarize, Mert is a bright, knowledgeable, and highly skilled scholar who is passionate about advancing science and computing. He has already achieved impressive milestones in his research on large-scale computational imaging and is working towards an essential contribution to the development of generalized frameworks for GPU systems following his PhD thesis. He has already established fruitful collaborations with academia and several top labs. He has made invaluable contributions to our undergraduate and graduate courses. He is already making an impact on science communities around the world. Based on Mert's research achievements, contributions and excellent potential, I give his faculty application my strongest endorsement.

Sincerely,

A handwritten signature in blue ink, appearing to read 'Wen-Mei Hwu', with a stylized flourish at the end.

Wen-Mei Hwu

Professor and AMD Jerry Sanders Chair Emeritus

Department of Electrical and Computer Engineering, UIUC

Senior Distinguished Research Scientist and Senior Director of Research, NVIDIA

Email: w-hwu@illinois.edu

Wen-mei W. Hwu is a Professor and the Sanders-AMD Endowed Chair Emeritus of ECE at the University of Illinois at Urbana-Champaign. He is a Senior Distinguished Research Scientist of NVIDIA. He served as the Illinois co-director of the IBM-Illinois Center for Cognitive Computing Systems Research Center (c3sr.com). He received the ACM SigArch Maurice Wilkes Award, the ACM Grace Murray Hopper Award, the IEEE Computer Society Charles Babbage Award, the ISCA Influential Paper Award, the MICRO Test-of-Time Award, the IEEE Computer Society B. R. Rau Award, the CGO Test-of-Time Award, and the Distinguished Alumni Award in CS of the University of California, Berkeley. He is a Fellow of IEEE and ACM.