Kuan-Chieh Hsu

khsu@bnl.gov | https://www.bnl.gov/staff/khsu | https://kuan-chiehhsu.github.io/

Research Interest

High-performance computing; Heterogeneous/parallel computing; computer architecture/system; Scientific/engineering computing; AI for system

Education

University of California, Riverside *Riverside, CA*

PhD in Computer Science and Engineering 2019 - 2024

- GPA: 3.78 / 4.0
- Thesis: General-Purpose Tensor Processors: A Broader Computational Paradigm

National Cheng Kung University *Tainan, Taiwan*

M.S. in Computer and Communication Engineering 2014 - 2016

- GPA: 4.0 / 4.0
- Thesis: Performance Prediction Model on HSA-Compatible General-Purpose GPU System

National Cheng Kung University *Tainan, Taiwan*

B.S. in Electrical Engineering 2010 - 2014

- GPA: 3.91 / 4.0

Experiences

Research Associate *Upton, NY*

Computing and Data Sciences, Brookhaven National Laboratory 2024 - current

- performance modeling

Research Assistant *Riverside, CA*

Extreme Scale Computer Architecture Laboratory, UCR 2019 - 2024

- **SHMT: Simultaneous and Heterogeneous Multithreading**
  - Proposed a new parallel execution and programming model to allow simultaneous execution of heterogeneous accelerators including Jetson Nano's GPU and Google's Edge TPU to achieve 1.95x speedup and 51% energy saving compared to optimal GPU baseline for a set of parallelism-rich applications in C++.
  - the open source code: https://github.com/escalab/SHMT
- **GPTPU: Accelerating Applications using Edge Tensor Processing Units**
  - Developed a programming framework and the system library for Google's EdgeTPU AI/ML accelerators to achieve 2.46x speedup and 40% energy saving compared to high-performance CPU cores for a set of matrix applications using C++.
  - Redesigned the algorithm and implemented the library function for general matrix multiplications using 2D convolutions to more efficiently use the EdgeTPU hardware and achieve 41.2x speedup using C++.
  - the open source code: https://github.com/escalab/GPTPU

Graduate Research Program Intern *Oak Ridge, TN*

Oak Ridge National Laboratory (ORNL) Fall and Winter 2023

- Explored the general-purpose computing using tensor processing units for future grid simulation.

Performance Team Intern *Sunnyvale, CA*

Cerebras System Summer 2022

- Developed MLIR-based infrastructure including IR readers to support compilation flow integration with analytical performance models.

System Architect Intern *San Jose, CA*

Samsung Strategy and Innovation Center Summer 2019

- Developed a set of deep learning kernel functions on an FPGA-based hardware acceleration platform using OpenCL via SDAccel environment.

Research Assistant (transferred out after the one year PhD program studying) *Raleigh, NC*

Extreme Scale Computer Architecture Laboratory, NCSU 2018 - 2019

- **Autonomous Near-Data Processing on Intelligent SSD**
  - Implemented Financial and linear algebra applications in Python with C-wrapper Python functions to invoke GPU kernels. The implementation offers our proposed Python runtime to achieve the same level of performance of C runtime.

Research Assistant *Taipei, Taiwan*

Data Insights Research Lab, Institute of Information Science, Academia Sinica 2017

- Proposed a typhoon trajectory forecasting framework that augments limited real-world data and produces Euclidean distance prediction error within 100 kms using TensorFlow-based model.
- Organized the conference as a core staff including session scheduling and hardware facilities arrangement: 4th Taiwan Artificial Intelligence/Data Science Conference in Academia Sinica, Taiwan (2000+ participants)

Research Assistant *Tainan, Taiwan*

Computer Architecture and System Lab, NCKU 2014 - 2016

- Implemented an early-staged heterogeneous system simulation including scalable memory subsystem with hierarchical cache system, NoC module, and memory controllers for OpenCL and OpenGL applications using hardware-style abstraction by C/C++.
- Designed modularized class materials for a Computer Organization course based on a full-system CPU simulator with programs implementing Fibonacci sequence to orient students with function calls using MIPS.

Summer Intern *Hsinchu, Taiwan*

Nuvoton Technology corp. Summer 2014

- Participated in early-staged design of an universal protocol IC project via summarizing timing, port definition and frequency discrepancy among involved protocols including I2C, SMBus and UART.
- Implemented data processing automation and visualization in R language to avoid human intervention on weekly generated measurement data.

Publications

[1] Phani R V Marthi, Soumyajit Gangopadhyay, **Kuan-Chieh Hsu**, Jongchan Choi, Nilanjan Ray Chaudhuri, and Suman Debnath, "RE-INTEGRATE EMT Simulation Software: DAE Solvers and Automation," *IEEE PES General Meeting 2025, Austin, TX* [TP]

[2] **Kuan-Chieh Hsu**, Jongchan Choi, and Suman Debnath, "Sparse Linear Solvers for Large-scale Electromagnetic Transient Simulations," *in 50th Annual Conference of the IEEE Industrial Electronics Society, (IECON) 24'* [TP]

[3] **Kuan-Chieh Hsu**, and Hung-Wei Tseng, "Simultaneous and Heterogenous Multithreading: Exploiting Simultaneous and Heterogeneous Parallelism in Accelerator-Rich Architectures," *in IEEE Micro, vol. 44, no. 4, pp. 11-19, 24'* [Paper]

[4] **Kuan-Chieh Hsu**, and Hung-Wei Tseng, "Simultaneous and Heterogenous Multithreading, " *in IEEE/ACM International Symposium on Microarchitecture (MICRO), 23'* [Paper]

[5] Yu-Chia Liu, **Kuan-Chieh Hsu**, and Hung-Wei Tseng, "Rethinking Programming Frameworks for In-Storage Processing," *Proceedings of the 60th ACM/IEEE Design Automation Conference (DAC) 23'* [Paper]

[6] **Kuan-Chieh Hsu**, and Hung-Wei Tseng, "GPTPU: Accelerating Applications using Edge Tensor Processing Units," *In International Conference for High Performance Computing, Networking, Storage and Analysis, SC 21'* [Paper]

[7] Yun-Chi Huang, **Kuan-Chieh Hsu**, Wan-shan Hsieh, Chen-Chieh Wang, Chia-Han Lu, and Chung-Ho Chen, "Dynamic SIMD Reconvergence with Paired-Path Comparison," *the IEEE International Symposium on Circuit and System (ISCAS), 16'* [Paper]

[8] Heng-Yi Chen, Chung-Ho Chen, Yun-Chi Huang, **Kuan-Chieh Hsu**, and Chen-Chieh Wang, "An HSAIL ISA conformed GPU platform," *Applied System Innovation - Proceedings of the International Conference on Applied System Innovation, (ICASI) 15'* [Link]

Technical Skills

Professional	C/C++, CUDA, Python, OpenCL, MATLAB, OpenMP, and R
Version control	git
Systems	Linux, Mac, Windows
Deep Learning	Keras, TensorFlow, Pytorch
Educational	Scratch