Kevin Hsieh

4th Floor, Collaborative Innovation Center, 5000 Forbes Avenue, Pittsburgh, PA 15213

kevinhsieh@cmu.edu | kevintwhsieh

Summary

Kevin Hsieh is a Ph.D. candidate in the Department of Electrical and Computer Engineering at Carnegie Mellon University, advised by Professor Phillip B. Gibbons and Professor Onur Mutlu. His research interest lies at the intersection of machine learning, systems, and computer architecture, with a recent focus on high-performance machine learning systems over real-world data. He has published 10+ papers in top systems/architecture venues such as OSDI, NSDI, and ISCA. Before pursuing his PhD, he was an engineering manager of a 14-person team in Mediatek, Taiwan, where he worked on system architecture and performance for mobile SoCs. He has broad experiences in software and hardware, such as distributed systems, processor architecture, OS kernels, and low-power systems.

Education

CMU (Carnegie Mellon University) *Pittsburgh, USA*

Ph.D. *in progress* in Electrical and Computer Engineering *Sep. 2014 - present*

- Advisors: Phillip B. Gibbons and Onur Mutlu

NCTU (National Chiao-Tung University) *Hsinchu, Taiwan*

M.S. in Computer Science and Information Engineering *Sep. 2002 - Jun. 2004*

- Thesis: "A problem-approach-based software knowledge classification system"
- Advisor: Chyan-Goei Chung

NCTU (National Chiao-Tung University) *Hsinchu, Taiwan*

B.S. in Computer Science and Information Engineering *Sep. 1998 - Jun. 2002*

- Honorary member of The Phi Tau Phi Scholastic Honor Society, R.O.C., 2002 (Top 1% EE/CS undergrad students in Taiwan)

Publications

- *Focus: Querying Large Video Datasets with Low Latency and Low Cost* *OSDI, 2018*
  Kevin Hsieh, Ganesh Ananthanarayanan, Peter Bodik, Shivaram Venkataraman, Paramvir Balh, Matthai Philipose, Phillip B. Gibbons, Onur Mutlu
- *A Case for Richer Cross-layer Abstractions: Bridging the Semantic Gap with Expressive Memory* *ISCA, 2018*
  Nandita Vijaykumar, Abhilasha Jain, Diptesh Majumdar, Kevin Hsieh, Gennady Pekhimenko, Eiman Ebrahimi, Nastaran Hajinazar, Phillip B. Gibbons, Onur Mutlu
- *The Locality Descriptor: A Holistic Cross-Layer Abstraction to Express Data Locality in GPU* *ISCA, 2018*
  Nandita Vijaykumar, Eiman Ebrahimi, Kevin Hsieh, Phillip B. Gibbons, Onur Mutlu
- *Toward Standardized Near-Data Processing with Unrestricted Data Placement for GPUs* *SC, 2017*
  Gwangsun Kim, Niladrish Chatterjee, Mike O'Connor, Kevin Hsieh
- *Gaia: Geo-Distributed Machine Learning Approaching LAN Speeds* *NSDI, 2017*
  Kevin Hsieh, Aaron Harlap, Nandita Vijaykumar, Dimitris Konomis, Gregory R. Ganger, Phillip B. Gibbons, Onur Mutlu

- *Zorua: A Holistic Approach to Resource Virtualization in GPUs* *MICRO, 2016*
  Nandita Vijaykumar, Kevin Hsieh, Gennady Pekhimenko, Samira Khan, Ashish Shrestha, Saugata Ghose, Adwait Jog, Phillip B. Gibbons, Onur Mutlu
- *Accelerating Pointer Chasing in 3D-Stacked Memory: Challenges, Mechanisms, Evaluation* *ICCD, 2016*
  Kevin Hsieh, Samira Khan, Nandita Vijaykumar, Kevin K. Chang, Amirali Boroumand, Saugata Ghose, Onur Mutlu
- *Transparent Offloading and Mapping (TOM): Enabling Programmer-Transparent Near-Data Processing in GPU Systems* *ISCA, 2016*
  Kevin Hsieh, Eiman Ebrahimi, Gwangsun Kim, Niladrish Chatterjee, Mike O'Connor, Nandita Vijaykumar, Onur Mutlu, Stephen W. Keckler
- *LazyPIM: An Efficient Cache Coherence Mechanism for Processing-in-Memory* *CAL, 2016*
  Amirali Boroumand, Saugata Ghose, Minesh Patel, Hasan Hassan, Brandon Lucia, Kevin Hsieh, Krishna T. Malladi, Hongzhong Zheng, Onur Mutlu
- *Understanding Latency Variation in Modern DRAM Chips: Experimental Characterization, Analysis, and Optimization* *SIGMETRICS, 2016*
  Kevin Chang, Abhijith Kashyap, Hasan Hassan, Samira Khan, Kevin Hsieh, Donghyuk Lee, Saugata Ghose, Gennady Pekhimenko, Tianshi Li, Onur Mutlu
- *Fast Bulk Bitwise AND and OR in DRAM* *CAL, 2015*
  Vivek Seshadri, Kevin Hsieh, Amirali Boroumand, Donghyuk Lee, Michael A. Kozuch, Onur Mutlu, Phillip B. Gibbons, Todd C. Mowry

Academic/Research Experience

Carnegie Mellon University *Pittsburgh, USA*
Ph.D Student *Aug. 2014 - Present*
Conducting research on distributed machine learning systems and near-memory computing architectures.

Microsoft Research *Redmond, USA*
Research Intern *May. 2017 - Aug. 2017*
Conducted research on interactive video querying on large video sets. The work was published at OSDI'18.

NVIDIA Research *Austin, USA*
Research Intern *Jun. 2015 - Sep. 2015*
Conducted research on enabling near-data computing for GPUs. The work was published at ISCA'16.

Work Experience

Mediatek inc. *Hsinchu, Taiwan*
Manager / Technical Manager / Senior Engineer *Nov. 2008 - Jul. 2014*

- Led system software teams (9-16 people) to develop the system boot-up code, OS kernel, debugging tools, platform drivers, low-power drivers, and chip verification suites on application processors, baseband processors, and DSPs across 13 mobile SoCs.
- Coordinated and worked with software and hardware teams to define processor/system architecture and optimize system performance. The work improved 15%-50% system performance with 20%-30% on-chip memory reduction across 7 mobile SoCs. The holistic approach was published as a 2011 Mediatek Best Paper.
- Developed an embedded RTOS kernel for a proprietary DSP. It's 30% smaller and 20%-30% faster than the previous generation. The RTOS kernel has been shipped with tens of millions chips without any bug reported.

Realtek Semiconductor Corp. *Hsinchu, Taiwan*
Senior Software Engineer *Oct. 2004 - Oct. 2008*

- Designed and implemented an OpenAL library from scratch to enable 3D sound effects in PC games.
- Designed and implemented a Windows GUI on Windows Vista from scratch, which consists of more than 30K lines of code and has been used for more than 7 years.

Teaching Experience

Teaching Assistant at Carnegie Mellon University *Pittsburgh, USA*
Advanced Operating and Distributed Systems (15-712), Graduate *Sep. 2015 - Dec. 2015*

Teaching Assistant at Carnegie Mellon University *Pittsburgh, USA*
Advanced Cloud Computing (15-719), Graduate *Jan. 2018 - May. 2018*

Service

Reviews: FAST 2019, IEEE Micro Top Picks 2018, IEEE/ACM Transactions on Networking 2018, PLDI 2017, ISCA 2015-2017, ICS 2017, HotStorage 2017, MICRO 2017, IEEE Transactions on Cloud Computing 2016, DAC 2016, ASPLOS 2016, OSDI 2016, HPCA 2015-2017

Skills

Programming Languages	C/C++, Python, C#, CUDA, Java, JavaScript, assembly (ARM, x86), Perl, PHP, ASP, HTML
Frameworks / OS	MFC, .NET, Linux kernel
Tools / Simulators	Microsoft CNTK, TensorFlow, Caffe, OpenCV, gem5, GPGPU-Sim, MATLAB, Octave, GDB, VTune, TRACE32, CodeViser