Chang-Hong Hsu

http://web.eecs.umich.edu/~hsuch
E: hsuch@umich.edu T: +1-(734)2392423

Profile

- A PhD candidate in Computer Science and Engineering, with an anticipated graduation time of Dec., 2017
- Worked on designing intelligent runtime systems that coordinate emerging hardware technologies to improve the scalability and efficiency of large-scale datacenters and modern datacenter workloads, such as large-scale cloud applications and machine learning pipelines
- Interested in positions whose jobs are to optimize datacenter latency and throughput, or develop large-scale highly-scalable systems for machine learning or cloud applications

Education

Ph.D., Computer Science and Engineering, University of Michigan, Ann Arbor 2012 – present
- Thesis title: Towards Power- and Energy-efficient Datacenters
- GPA: 4.0/4.0

M.S., Electrical Engineering, National Taiwan University 2008 -- 2011
- GPA: 95.67/100.0 (Ranking: 1/74)

B.S., Electrical Engineering, National Taiwan University 2004 -- 2008

Internship Experiences

Research Intern, Facebook Inc.
- Helped design a datacenter power-capping runtime and validate the design decisions of it
- Developed a framework that help deriving service placement for highly efficient power budget utilization

Intern, Cadence Taiwan
- Designed primitive modules that help accelerating circuit verification

Skills

Programming & Markup Languages
- PL & HDL: C/C++, Python, Java, JavaScript, Shell script, Verilog
- Markup: LaTeX, HTML, CSS

Tools and Simulators
- Compilers: LLVM
- Simulator: BigHouse Simulator, gem5
- Distributed computing: Apache Spark, Apache Storm, Apache Hadoop
- Design automation tools and solvers: EDA/FPGA toolchains, Solver: MiniSAT, boolector

Selected Projects

Architectural and System Implication of Largescale Video Search System (Work in progress)
- Collaborating with UM Transportation Research Institute (UMTRI) to build a video search engine for a large dash-cam video dataset
- Integrating deep learning pipelines to analyze video and sensor data and answer complex user queries
- Investigating where system bottlenecks are and how to use distributed-computing engines or capabilities (i.e., Spark, Hadoop, etc.) to improve system throughput

Fine-grain Resource Scheduling for Reconfigurable Datacenter Hardware
- Investigated how to leverage the partial reconfiguration capability of modern tightly-coupled CPU-FPGA platforms (e.g., Intel HARP) to help improve datacenter performance and efficiency
- Designed a runtime system to accurately schedule requests and manage FPGAs resources in fine granularity on-the-fly
- Authored a simulator for CPU-Accelerator heterogeneous system on top of BigHouse, an event-driven queueing-theoretic datacenter simulator

Combating Power Budget Fragmentation Problem in Large-Scale Datacenters
- Identified root cause of suboptimal power budget utilization in large-scale datacenters
- Leveraged temporal heterogeneity of the power consumption patterns among different services and designed a clustering-based service placement framework to optimize power utilization
- Increased the power budget utilization significantly without changing the power-delivery infrastructure

Pinpointing and Reining in Tail Queries with Quick Voltage Boosting

- Developed a framework that improve tail latency of important cloud applications (Memcached and Web Search) by more than 4× with high energy efficiency
- Identified query-level indicator to predict and pinpoint tail queries, and design low-overhead DVFS policy to utilize quick voltage switching circuits to boost system performance for tail queries

High-Performance Post-Silicon Architectural Checking via Event Digests

- Speeded up the data-intensive, acceleration-platform-based post-silicon validation process by proposing a sequence-by-sequence checking approach on digests of logged architectural events
- Largely reduces the amount of offloaded data by >90% without losing detection accuracy

Teaching Experiences

Graduate Student Instructor, EECS 583 Advanced Compiler, University of Michigan

- Graduate-level compiler course. Guided students to learn LLVM and design backend compiler passes that effectively takes advantage of architectural characteristics of the underlying hardware systems

Teaching Assistant, Logic Synthesis and Verification S'10, National Taiwan University

- Graduate-level EDA course. Guided students with their projects to learn and practice the knowledge about logic synthesis, optimization, and verification

Teaching Assistant, Data Structure and Programming F'09, National Taiwan University

- Undergraduate-level course. Guided students to establish solid background in C/C++ programming and data structure through building and optimizing a binary decision diagram (BDD) library

Publications

[MICRO'17, to appear] Addressing Compute and Memory Bottlenecks for DNN Execution on GPUs
P. Hill, A. Jain, M. Hill, B. Zamirai, M. Laurenzano, C.-H. Hsu, S. Mahlke, L. Tang, J. Mars

[IEEE IC] Thermal Time Shifting: Decreasing Datacenter Cooling Costs with Phase Change Materials
M. Skach, M. Aurora, C.-H. Hsu, Q. Li, D. Tullsen, L. Tang, J. Mars

[ACM TOCS] Achieving Short Tail Latency with High Energy Efficiency for Warehouse-scale Computers with Adrenaline
C.-H. Hsu, Y. Zhang, M. A. Laurenzano, D. Meisner, T. Wenisch, R. G. Dreslinski, J. Mars, L. Tang

[ISCA'16] Dynamo: Facebook's Data Center-Wide Power Management System
Q. Wu, Q. Deng, L. Ganesh, C.-H. Hsu, Y. Jin, S. Kumar, B. Li, J. Meza, Y. J. Song

[ISCA'15] Thermal Time Shifting: Leveraging Phase Change Materials to Reduce Cooling Costs in Warehouse-Scale Computers
M. Skach, M. Arora, C.-H. Hsu, D. Tullsen, J. Mars, L. Tang

[HPCA'15] Adrenaline: Pinpointing and Reining in Tail Queries with Quick Voltage Boosting
C.-H. Hsu, Y. Zhang, M. A. Laurenzano, D. Meisner, T. Wenisch, J. Mars, L. Tang, R. G. Dreslinski

[CCS'14] Verifying Curve25519 Software
Y.-F. Chen, C.-H. Hsu, H.-H. Lin, P. Schwabe, M.-H. Tsai, B.-Y. Wang, B.-Y. Yang, S.-Y. Yang

[DATE'14] ArChiVED: Architectural Checking via Event Digests for High-Performance Validation
C.-H. Hsu, D. Chatterjee, R. Morad, R. Gal, V. Bertacco

[ASP-DAC'11] A Robust ECO Engine by Resource-constraint-aware Technology mapping and Incremental Routing Optimization
S.-L. Huang, C.-A. Wu, K.-F. Tang, C.-H. Hsu, C.-Y. Huang

[ICCAD'10] Formal Deadlock Checking on High-level SystemC Designs
C.-N. Chou, C.-H. Hsu, Y.-T. Chao, C.-Y. Huang

Selected Courses

Computer Architecture-related

- UM: EECS470-Computer Architecture (A+), EECS570-Parallel Computer Architecture (A), EECS578-Correct Operation for Processors and Embedded Systems (A+), EECS583-Advanced Compiler (A)
- NTU: Digital System Design (91/100), SoC Design Lab (99/100)

Artificial Intelligence-/Machine Learning-related

- UM: EECS492-Introduction to AI (A)
- NTU: Data Mining (92/100)