CHIEN-CHIN HUANG

(347)221-2653
fegins@gmail.com
fegin.github.io
github.com/fegin

Education

2013–Present **Ph.D, Computer Science**, *New York University*.
- Advisor: Dr. Jinyang Li
- Research Topic: Machine Learning Systems with Large Tensors Support

2005–2007 **M.S., Computer Science**, *National Tsing Hua University*.
- Advisor: Dr. Jenq-Kuen Lee
- Thesis Topic: Microkernel Design and Dual-core Supports for PAC VLIW DSP Processors.

2001–2005 **B.S., Computer Science**, *National Tsing Hua University*.

Publications (selected)

Support Very Large Models using Dataflow Graph Partitioning.
Minjie Wang, Chien-Chin Huang, Jinyang Li.
The European Conference on Computer Systems (EuroSys'19), March 2019

Unifying Data, Model and Hybrid Parallelism in Deep Learning via Tensor Tiling.
Minjie Wang, Chien-Chin Huang, Jinyang Li.
arXiv, 2018

Fast Image Clustering on Network of Mobile Phones.
Jorge Ortiz, Chien-Chin Huang, Supriyo Chakraborty.
ICML 2016 Workshop for On-Device Intelligence

Spartan: A Distributed Array Framework with Smart Tiling.
Chien-Chin Huang, Qi Chen, Zhaoguo Wang, Russell Power, Jorge Ortiz, Jinyang Li, Zhen Xiao.
USENIX Annual Technical Conference (ATC'15), July 2015

Garbage Collection for Multiversion Index in Flash-based Embedded Databases.
Po-Chun Huang, Yuan-Hao Chang, Kam-Yiu Lam, Jian-Tao Wang, Chien-Chin Huang.
ACM Transactions on Design Automation of Electronic Systems, June 2014

Research Projects

2018-present **SwapAdvisor: Support Large Deep Learning Models via Static Planning Tensor Swap**.
SwapAdvisor helps users to explore large deep learning models by statically analyzng the dataflow graph and memory constraints to provide a good tensor swapping plan that incurs minimized swapping overhead.

2016-2018 **Tofu: Distributing Tensor Computation Automatically for Large-scale Machine Learning**.
Tofu is a system that partitions very large deep neural network models across multiple GPU devices to reduce per-GPU memory footprint via automatically choosing the partition strategies among different parallelisms.

2014-2015 **Spartan: Distributed Array Programming Framework**.
Spartan is a distributed array framework which provides dataflow operators for users to implement distributed array programs. Spartan then partitions arrays across machines by analyzing the combination of the operators.

Professional Experience

2017-2018 **Recitation Instructor (Computer System Organization)**, *NYU*, New York, NY, USA.

2016 **Summer Software Engineer Intern**, *Google Inc.*, Mountain View, CA, USA.

2015 **Summer Research Intern**, *IBM T.J. Watson Research Center*, Yorktown Heights, NY, USA.

2013–Present **Research Assistant**, *NYU*, New York, NY, USA.

2012–2013 **Research Assistant**, *Institute of Information Science, Academia Sinica*, Taipei, Taiwan.

2011–2012 **Senior Software Engineer**, *MediaTek Inc.*, Hsinchu, Taiwan.

2008–2011 **Software Engineer**, *MediaTek Inc.*, Hsinchu, Taiwan.

Honors, Awards and Grants

2015 **ATC'15 Student Travel Grant**.

2014 **OSDI'14 Student Travel Grant**.

2007 **1st Place**, *Best Thesis Award*, Institute of Information Computing Machinery, Taiwan.

2006 **1st Place**, *Embedded System Design Contest*, Ministry of Education, Taiwan.

2006 **Selective Preference**, *SiliconAward, Embedded Software Group*, MXIC Inc..

2004 **7th Place**, *ACM International Collegiate Programming Contest (ICPC)*, Kaohsiung Station.