

Yu-Ju Huang

yh885@cornell.edu
<https://www.cs.cornell.edu/~yjhuang>

- Education**
- Ph.D. Computer Science, Cornell University, NY, USA** Aug. 2019 – May. 2025 (expected)
- Advisor: Prof. Robbert van Renesse
- M.S. Computer Science, National Chiao Tung University, Hsinchu, Taiwan** Sep. 2013 - Jul. 2015
- Advisor: Prof. Wei-Chung Hsu
- B.S. Computer Science, National Chiao Tung University, Hsinchu, Taiwan** Sep. 2009 - Jul. 2013
-

- Publications**
- Disaggregated Applications Using Nanoservices**
Xinwen Wang, **Yu-Ju Huang**, Tiancheng Yuan, Robbert van Renesse
Workshop On Resource Disaggregation and Serverless (WORD'21), April 2021
- Building a KVM-based Hypervisor for a Heterogeneous System Architecture Compliant System**
Yu-Ju Huang, Hsuan-Heng Wu, Yeh-Ching Chung, Wei-Chung Hsu.
12th International Conference on Virtual Execution Environments (VEE'16), April 2016
Best paper award!
-

- Research**
- My general interest is to see how systems (on one or multiple machines and all layers of computer systems) work and how to make them work even better. My research thus spans operating systems, virtualization, cloud computing, distributed computing, and computer architecture.

Ziplog (ongoing)

Near-realtime stream processing is getting more important in data-intensive applications. A key building block of it is the message queue (or message broker, event stream), such services include Corfu, Apache Kafka, Amazon Kinesis Streams, ... etc. While heavily used, state-of-the-art shard log designs don't provide all the properties that application developers usually require, which are cross-shard total order, scalable throughput, and low latency. The Ziplog project tries to achieve all three: it guarantees total order, achieves scalable throughput, and experiences low latency!

Fluid OS (ongoing)

With the proposal of hardware resource disaggregation, we believe that the paradigm of computer abstraction is shifting. While we used to think a computer consists of a fixed set of hardware resources, the new computer abstraction is becoming fluid, where the hardware resources are abundant, heterogeneous, and elastic.

We aim to build an OS that is also fluid, i.e., it views hardware resources as a dynamic pool. It can boot with a small amount of CPU, memory and can grow more powerful (including GPU and accelerators!) upon requests from the applications

HSA virtualization (past)

- Built a hypervisor that virtualizes a new heterogeneous computing architecture called Heterogeneous System Architecture (HSA).
- Allowed guest processes to use a GPU on the HSA-compliant machine (AMD's Kaveri) and achieves nearly 95% of native performance.
- Multiple guest processes on different guest OSes and are able to share one GPU.

Professional Service

Program Committee Member and Session Chair

Xian, China. April. 2017

13th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments (VEE).

Employment

System Software Engineer - MediaTek, Office of CTO

Hsinchu, Taiwan. Dec. 2015 – Jun. 2019

* *Compiler and runtime software stack for edge AI*

- Built software to support various frontend AI frameworks and backend computing devices.
- Compiler development for in-house AI processor.

* *Android runtime (ART) and compiler optimization*

- Built a staged Java compiler using ART's optimising compiler and LLVM to optimize Java applications.
- Improve performance by around 10% on a plenty of benchmarks.

* *QoS (quality of service)-based framework for optimizing Android runtime*

- Enhanced the Android runtime to be aware of QoS hints and adjust the system accordingly.
- Extended the Android runtime to profile Java applications and export QoS hints.
- Improve user experience by removing some frame drops in game scenario.

Intern Engineer - Marvell Taiwan

Hsinchu, Taiwan. Jul. 2014 - Aug. 2014

* *Visualization tool for cache system*

- Used Gen5 to generate cache profiles and created a visualizer to show the complex behavior.
- This tool helps the architecture group analyze the performance issues more clearly.

Intern Engineer - ITRI, Cloud Computing Center

Hsinchu, Taiwan. Feb. 2014 - Jul. 2014

* *KVM-ARM I/O virtualization optimization*

- Explore different kinds of I/O virtualization techniques, including VirtIO and device pass-through, and use them to optimize ARMv8 foundation model as a practice.
-

Awards

Cornell University Fellowship, 2019-2020

Best paper award, 12th International Conference on Virtual Execution Environments (VEE'16)

Teaching

Cornell University

Grad TA for CS 3410: Computer System Organization and Programming, Fall 2021

Head TA for CS 4411: Practicum in Operating Systems, Spring 2020

Head TA for CS 4410: Operating Systems, Fall 2020

Extracurricular Activities

President of Cornell Taiwanese Student Association

Jun. 2020 – Jun 2021

Captain of Kendo Team at National Chiao Tung University

Sep. 2010 - Feb. 2012