Z. JONNY KONG

+1(310) 498-9627 ◇ kong102@purdue.edu ◇ www.jonnykong.com

I am experienced in and enjoys building **networked systems, mobile systems, and systems for machine learning**. Recent focus is on building Machine-Learning-as-a-Service (MLaaS) serving systems for GPU clusters to serve a large amounts of concurrent requests with SLA guarantees.

EDUCATION

Purdue University — *West Lafayette, IN, U.S.*
Ph.D. in Electrical and Computer Engineering — *Aug 2020 - Present*

University of California, Los Angeles — *Los Angeles, CA, U.S.*
M.S. in Computer Science — *Sep 2018 - June 2020*

Beihang University — *Beijing, China*
B.E. in Automation — *Sep 2014 - June 2018*

RESEARCH AND PROFESSIONAL EXPERIENCE

Purdue University — *West Lafayette, IN, U.S.*
Research Assistant — *Aug 2020 - Present*
Advisor: Prof. Y. Charlie Hu

- Designed a machine-learning-as-a-service (MLaaS) framework for heterogeneous GPU clusters that exploits model parallelism to improve server capacities, improving serving throughput by 16.7%-52.8% [1]
- Designed an MLaaS framework for serving edge-assisted AR mobile apps, that maximizes the capacities of GPU servers and serves 1.7-6.9x more clients [2]
- Designed MLaaS frameworks that optimize the overall accuracy of an AR mobile app that offloads multiple tasks to an edge GPU server, improving the overall accuracy by 7.6%-14.3% [4]
- Performed measurement studies on next-generation wireless networks, e.g. 5G [3] [9] and 802.11ad [7]
- Designed edge-assisted AR mobile applications [10] [11], and conducted measurement studies on their performance over 5G networks [5] [6]

University of California, Los Angeles — *Los Angeles, CA, U.S.*
Research Assistant — *Oct 2018 - Jun 2020*
Advisor: Prof. Lixia Zhang

- Designed data synchronization protocols [8] [13], a transport-layer protocol for Named Data Networking (NDN)

PUBLICATIONS

Conference Papers

[1] **Z. Jonny Kong***, Qiang Xu*, Y. Charlie Hu. *"IPIPE: Enabling Effective DNN Serving on Heterogeneous GPU Clusters via Model Parallelism"*. Under submission. (* co-primary)

[2] **Z. Jonny Kong***, Qiang Xu*, Y. Charlie Hu. *"ARISE: An Accuracy-Aware Proactive Framework for Serving Concurrent Edge-Assisted AR Clients"*. Under submission. (* co-primary)

[3] Moinak Ghoshal*, Imran Khan*, **Z. Jonny Kong***, Phuc Dinh, Jiayi Meng, Y. Charlie Hu, Dimitrios Koutsonikolas. *"Performance of Cellular Networks on the Wheels"*. In **ACM IMC 2023**. (* co-primary)

[4] **Z. Jonny Kong***, Qiang Xu*, Jiayi Meng, Y. Charlie Hu. *"AccuMO: Accuracy-Centric Multitask Offloading in Edge-Assisted Mobile Augmented Reality"*. In **ACM MobiCom 2023**. (*co-primary)

[5] Moinak Ghoshal*, **Z. Jonny Kong***, Qiang Xu*, Zixiao Lu, Shivang Aggarwal, Imran Khan, Jiayi Meng, Yuanjie Li, Y. Charlie Hu, Dimitrios Koutsonikolas. *"Can 5G mmWave Enable Edge-Assisted Real-Time Object Detection for Augmented Reality?"*. In **IEEE MASCOTS 2023**. (*co-primary)

[6] Moinak Ghoshal, Pranab Dash, **Zhaoning Kong**, Qiang Xu, Y. Charlie Hu, Dimitrios Koutsonikolas, Yuanjie Li. *"Can 5G mmWave Support Multi-User AR Apps?"*. In **PAM 2022**. [pdf]

[7] Shivang Aggarwal, **Zhaoning Kong**, Moinak Ghoshal, Y. Charlie Hu, Dimitrios Koutsonikolas. *"Throughput Prediction on 60 GHz Mobile Devices for High-Bandwidth, Latency-Sensitive Applications"*. In **PAM 2021 (Best Dataset Award)**. [pdf]

[8] Tianxiang Li, **Zhaoning Kong**, Spyridon Mastorakis, Lixia Zhang. *"Distributed Dataset Synchronization in Disruptive Networks"*. In **IEEE MASS 2019**. [pdf]

Workshops & Posters

[9] Moinak Ghoshal*, **Z. Jonny Kong***, Qiang Xu*, Zixiao Lu, Shivang Aggarwal, Imran Khan, Yuanjie Li, Y. Charlie Hu, and Dimitrios Koutsonikolas. *"An In-Depth Study of Uplink Performance of 5G mmWave Networks"*. In **ACM SIGCOMM 5G-MeMU Workshop '22**. (* co-primary) [pdf]

[10] Jiayi Meng, **Z. Jonny Kong**, Y. Charlie Hu, Mun Gi Choi, Dhananjay Lal. *"Do We Need Sophisticated System Design for Edge-assisted Augmented Reality?"*. In **ACM EdgeSys 2022 (Best Paper Award)**. [pdf]

[11] Jiayi Meng*, **Zhaoning Kong***, Qiang Xu, Y. Charlie Hu. *"Do Larger (More Accurate) Deep Neural Network Models Help in Edge-assisted Augmented Reality?"*. In **ACM SIGCOMM NAI Workshop '21**. (*co-primary) [pdf]

[12] Lana Ramjit, **Zhaoning Kong**, Ravi Netravali, Eugene Wu. *"Physical Visualization Design (demo)"*. In **ACM SIGMOD 2020**. [pdf]

[13] Tianxiang Li, **Zhaoning Kong**, Lixia Zhang. *"Supporting Delay Tolerant Networking: A Comparative Study of Epidemic Routing and NDN"*. In **IEEE ICC '20 ICN-SRA workshop**. [pdf]

SELECTED AWARDS

Research Awards

- Best Paper Award, EdgeSys '22
- Best Dataset Award, PAM '21

Student Awards

- National Scholarship of China, 2017 (Top 0.2% nationwide)

PROFESSIONAL SERVICES

Journal Reviewers: IEEE Network, Computer Communications
Artifact Evaluation Committee (AEC): ACM MobiSys 2023, SOSP 2023

TEACHING ASSISTANT

ECE 26400 Advanced C Programming, Fall '20, Spring '21, Summer '21, Purdue University
CS 151B Computer Systems Architecture, Winter '20, UCLA
CS 217A Internet Architecture and Protocols, Fall '19, UCLA