

YUAN MENG

(+1) 518 961 3768 ✧ ymeng643@usc.edu

Personal Website: <https://meng-yuan-usc.com/>

EDUCATION

University of Southern California

Ph.D. candidate, Computer Engineering

Ming Hsieh Department of Electrical and Computer Engineering

Fall 2019 - Present

Overall GPA: 3.86/4.0

Rensselaer Polytechnic Institute

B.S., Dual Major: Electrical Engineering & Computer and Systems Engineering

Department of Electrical, Computer, and Systems Engineering

Fall 2015 - Spring 2019

Overall GPA: 3.8/4.0

RESEARCH INTERESTS

- Learning in the loop for dynamic systems on parallel and/or heterogeneous platforms
- Hardware accelerators for deep learning

RESEARCH EXPERIENCE

Accelerating Reinforcement Learning (RL)

Graduate Research Assistant

Ongoing

@FPGA/Parallel Computing Lab, USC

- Proposed an algorithm-hardware co-optimized accelerator for dynamic tree management in MCTS (FPL '22, FPGA '23)
- Proposed the first CPU-FPGA system design for accelerating parallel Monte-Carlo Tree Search (MCTS) that is portable to various applications, and achieved better scalability to larger number of parallel workers than state-of-the-art CPU-only implementations. (FPL '22)
- Proposed a novel task mapping framework for Deep RL on CPU-GPU-FPGA Heterogeneous Platform, and designed a pipelined accelerator for prioritized replay operations based on K-ary Sum Tree data structure (CF '22, TPDS)
- Developed a software toolkit for interfacing Pytorch host with High-Level Synthesis for bench-marking FPGA-accelerated Reinforcement Learning (FPL '21)
- Implemented pipelined architecture on FPGA to accelerate Table-based Q learning (IPDPSW '20)
- Designed and developed a systolic-array-based architecture to accelerate Proximal Policy Optimization algorithm on CPU-FPGA heterogeneous platforms (FCCM '20, TPDS)

Acceleration and Design Automation of Deep Learning

Graduate Research Assistant

Ongoing

@FPGA/Parallel Computing Lab, USC

- Designed a unified hardware overlay supporting dynamic mapping of different convolution algorithms across CNN layers for low-latency inference (FPGA '21, IC3 '22)
- Proposed a hybrid accelerator design for streaming Deformable Convolution Network inference, and proposed an associated DSE engine for mapping model to pipeline stages using Tandem Repeat Sequence (SiPS '22)
- Proposed a polynomial-time PBQP (Partitioned Boolean Quadratic Programming) optimal solution to the algorithm mapping problem on series-parallel CNN graphs (FPGA '21)
- Proposed a compute-saving kn2row-based algorithm hardware co-design methodology for accelerating fractionally strided convolution in upsampling and generative CNNs (HiPC '21)

Mesh Adaptation using FPGAs

Undergraduate Research Assistant

Fall 2018

@Scientific Computation Research Center, RPI

- Implemented and optimized a parameterized 2-D mesh adaptation accelerator on FPGAs with High Level Synthesis, evaluated algorithm efficiency and optimization methods
- Developed and explored parameterized OpenCL kernels implementing 1-D tensor operations including Array Reductions and Sliding Average on cloud FPGA nodes

PUBLICATIONS

Journal Papers:

Chi Zhang, **Yuan Meng** and Viktor Prasanna “A Framework for Mapping DRL Algorithms with Prioritized Replay Buffer onto Heterogeneous Platforms.” Under review in the IEEE Transactions on Parallel and Distributed Systems (TPDS)

Yuan Meng, Sanmukh Kuppannagari, Rajgopal Kannan and Viktor Prasanna “PPOAccel: A High-Throughput Acceleration Framework for Proximal Policy Optimization.” IEEE Transactions on Parallel and Distributed Systems (TPDS).

Conference Papers:

Yuan Meng, Rajgopal Kannan and Viktor Prasanna “A Framework for Monte-Carlo Tree Search on CPU-FPGA Heterogeneous Platform via Dynamic Tree Management.” The 2023 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA 2023)

Yuan Meng, Hongjiang Men and Viktor Prasanna “Accelerator Design and Exploration for Deformable Convolution Networks.” The 36th IEEE Workshop on Signal Processing Systems (SiPS 2022)

Haomei Liu, **Yuan Meng**, Sanmukh Kuppannagari and Viktor Prasanna “End to End Framework for CNN Acceleration on FPGAs with Dynamic Algorithm Mapping.” The 2022 International Conference on Contemporary Computing (IC3 2022)

Yuan Meng, Rajgopal Kannan and Viktor Prasanna “Accelerating Monte-Carlo Tree Search on CPU-FPGA Heterogeneous Platform.” 22nd International Conference on Field Programmable Logic & Applications (FPL 2022)

Yuan Meng, Chi Zhang, and Viktor Prasanna, “FPGA acceleration of deep reinforcement learning using on-chip replay management.” Proceedings of the 19th ACM International Conference on Computing Frontiers (CF 2022). **(Best Paper Award)**

Yuan Meng, Sanmukh Kuppannagari, Rajgopal Kannan and Viktor Prasanna “How to Avoid Zero-Spacing in Fractionally-Strided Convolution? A Hardware-Algorithm Co-Design Methodology.” High Performance Computing, Data, and Analytics (HiPC 2021).

Nathaniel Peura, **Yuan Meng**, Sanmukh Kuppannagari and Viktor Prasanna “FGYM: Toolkit for Benchmarking FPGA based Reinforcement Learning Algorithms.” Field-Programmable Logic and Applications (FPL 2021) (Demo Abstract).

Yuan Meng, Sanmukh Kuppannagari, Rajgopal Kannan and Viktor Prasanna, “DYNAMAP: Dynamic Algorithm Mapping Framework for Low Latency CNN Inference.” The 2021 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA 2021).

Yuan Meng, Sanmukh Kuppannagari, Rajgopal Kannan and Viktor Prasanna, “How to Efficiently Train Your AI Agent? Characterizing and Evaluating Deep Reinforcement Learning on Heterogeneous Platforms.” 24th IEEE High Performance Extreme Computing Conference (HPEC 2020). **(Outstanding Student Paper Award)**

Yuan Meng, Sanmukh Kuppannagari, and Viktor Prasanna. “Accelerating Proximal Policy Optimization on CPU-FPGA Heterogeneous Platforms.” 28th IEEE International Symposium on Field-Programmable Custom Computing Machines (FCCM 2020).

Yuan Meng, et al. “QTAcel: A Generic FPGA based Design for Q-Table based Reinforcement Learning Accelerators.” 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW 2020).

AWARDS & HONOURS

- | | |
|---|-----------|
| - Finalist, Ming Hsieh Institute Scholarship
<i>University of Southern California</i> | 2022-2023 |
| - Best Paper Award
<i>ACM International Conference on Computing Frontiers</i> | 2022 |
| - Outstanding Student Paper Award
<i>IEEE High Performance Extreme Computing Conference</i> | 2020 |
| - Annenberg Fellowship
<i>University of Southern California</i> | 2019-2020 |
| - Dean's Honor List
<i>Rensselaer Polytechnic Institute</i> | 2015-2019 |

MENTORING

- Samuel Wiggins (PhD Student) - Acceleration of communication in multi-agent reinforcement learning
- Zongle Huang (Undergraduate Student) - FPGA acceleration of deep learning guided Monte-Carlo Tree Search
- Peter Wang, Tianxin Zhu (Undergraduate Student) - Multi-Core Acceleration of AlphaZero using Adaptive Parallelism
- Haomei Liu (Undergraduate Student) - End to End Framework for CNN Acceleration on FPGAs with Dynamic Algorithm Mapping (Accepted to IC3 2022)
- Hongjiang Men (Master Student) - Accelerator Design For Deformable Convolution Networks using HBM (Accepted to SiPS 2022)
- Nathaniel Peura (Master Student) - FGYM: Framework for benchmarking FPGA-accelerated Reinforcement Learning under VITIS software development flow (FPL 2021 DEMO)
- Sarah Chow (Undergraduate Student) - HLS-accelerated Deep Neural Network Inference

WORK EXPERIENCE

- | | |
|--|---------------------------|
| Graduate Teaching Assistant
<i>University of Southern California</i> | Fall 2020 - Spring 2021 |
| <ul style="list-style-type: none">• Parallel and Distributed Computing• Accelerated Computing using FPGAs• Parallel Programming | |
| Electronics Engineer Co-op
<i>Hasbro, Inc.</i> | January 2018 - June 2018 |
| <ul style="list-style-type: none">• Prototyping for Animatronics and games• Research on embedded voice recognition and computer vision applications in toys | |
| Undergraduate Teaching Assistant/Mentor
<i>Rensselaer Polytechnic Institute</i> | Spring 2017 - Spring 2019 |

- Embedded Control
- Foundation of Computer Science

TECHNICAL STRENGTHS

Programming Languages

Python, C/C++, PostgreSQL Scripting

Parallel Programming

CUDA, OpenMP, MPI, OneAPI

Hardware Design

VIVADO & VITIS HLS, OpenCL, Verilog

Embedded Prototyping

Arduino, C8051 & stm32 microprocessors, Raspberry Pi

Other Software & Tools

OpenAI Gym, PyTorch, PyOpenCL, CAD (NX, AutoCAD), SPICE