XIANZHONG DING

Email: dingxianzhong@lbl.gov
Phone Number: +1(979)739-3710

EDUCATION

University of California, Merced August 2018 - December 2023
Ph.D. in Electrical Engineering and Computer Science GPA: 4.0
Shandong University, China August 2015 - June 2018
M.S. in Computer Science and Technology GPA: 3.19
Taishan University, China August 2010 - June 2014
B.S. in Computer Science and Technology GPA: 3.25

RESEARCH INTERESTS

Cyber Physical Systems, Large Language Models, Deep Reinforcement Learning and ML for Systems.

SKILLS

Programming languages: Python, C/C++, Java, Matlab
Deep learning framework: Tensorflow, Pytorch
Tools and Libraries: Keras, Scikit-learn, OpenAI gym, Pandas, Jupyter, OpenCV

WORKING EXPERIENCE

Postdoctoral Researcher, Lawrence Berkeley National Laboratory (Berkeley Lab)
Feb 2024 - Present Supervisor: Dr. Bin Wang

1. Parking Lot Identification Using Low-Resolution Satellite Imagery [2]
2. Large-Scale Infrastructure Assessment for Medium- and Heavy-Duty Electric Vehicles: Performance Optimization and Scalability of the HEVI-LOAD Tool
3. AI-Driven Scenario Automation for National-Scale Charging Infrastructure Assessment: Leveraging Large Language Models and Retrieval-Based Systems in HEVI-LOAD
4. Optimizing Charging Load Peak Reduction for Electric Vehicles Using Deep Reinforcement Learning

Research Intern, Argonne Leadership Computing Facility, Argonne National Laboratory
May - Aug 2023 Supervisor: Dr. Murali Emani

1. Integrating Large Language Models to HPC-FAIR [10]:
- Introduce the HPC-GPT model, the first open-source HPC LLM tuned with instruction data.
- Integrate HPC knowledge, ensuring our model has accurate and domain-specific knowledge.
- Compared with state of the art, improving accuracy by 21.0%.

2. Data Race Detection Using Large Language Models [11]:
- Derive an innovative dataset from DataRaceBench for LLM fine-tuning.
- Fine-tune LLMs for the explicit task of data race detection to enhance their predictive accuracy.
- Detailed comparative study between traditional data race detection tools and LLM-based methods.

Research Intern, ByteDance Infrastructure System Lab, Mountain View, United States
May - Nov 2022 Supervisor: Dr. Tieying Zhang

1. Deep Reinforcement Learning for Virtual Machine (VM) Scheduling in Cloud Datacenter:
- Propose a DRL-based method to optimally place incoming VM requests.
- Design action as candidate pool of heuristic methods to solve high dimensional action space.
- Compared with state of the art, reducing fragment rate by 58.9%.

2. Deep Reinforcement Learning for Virtual Machine Rescheduling (VMR) in Cloud Datacenter [9]:
- Propose a DRL-based VMR framework towards a better trade-off between solution quality and speed.
- Design an attention-based agent architecture for sequential decisions on the migrated VM and PM.
- Incorporate imitation learning of the heuristic methods as a warm start to accelerate training.
- Achieve solution quality close to MIP method and computation speed close to heuristic methods.

3. ByteLife: Machine Learning algorithms for Ventricular fibrillation (VF) on Edge Device:
- Design different machine learning techniques including ensemble learning and hyper-parameter tuning to balance the detection accuracy, inference latency and memory occupation on edge device-NUCLEO.
- Achieve 32th over 150 teams on the 1st ICCAD TinyML Contest with 83% accuracy, 55.71ms latency and 68.92 KiB memory.

RESEARCH EXPERIENCE

Research Assistant, EECS, UC Merced *August 2018 - May 2022*

1. Chain-of-Thought (CoT) Prompting Order Does Matter for Large Language Models (LLM):
- Propose a dynamic CoT Prompting method based on the similarity between questions and CoT.
- Analyse the effect of the CoT Prompting numbers and orders to GPT-3.
- Experiments on ten benchmark reasoning tasks with GPT-3 show prompting order can generate 0-33% accuracy difference and the best CoT order can improve 2% accuracy.

2. Deep Reinforcement Learning for Water Resource Optimization [15]:
- Propose a DRL-based irrigation framework for agricultural water usage saving.
- Design a safety irrigation module to evaluate whether the RL algorithm outputs a safe action.
- Build an irrigation testbed with customized sensing and actuation nodes, and irrigation system.

3. Model-Based Deep Reinforcement Learning for Multi-zone Building Control [17]:
- Propose an efficient model-based DRL building control system.
- Develop a weighted ensemble learning to solve building neural network model uncertainty.
- Adopt a model predictive path integral control method to perform building control.

4. Real-Time Object Detection on Mobile Devices [18]:
- Build a deep neural network-based object detection method on mobile devices without offloading.
- Propose a parallel detection and tracking pipeline to fully utilize the computation resource on current mobile devices for high detection accuracy.
- Design an algorithm to adapt the DNN models according to the change rate of video content.

5. DRL for Holistic Smart Building Control [19]:
- Leverage DRL to balance the trade-off between energy use and human comfort for smart buildings.
- Adopt a special reward function and a novel neural network architecture to tackle the challenges imposed by the combined joint control of four subsystems with a very large action space.
- Tackle data training requirements by adopting a simulation strategy for data generation, and spending effort in calibrating the simulations to make them as close as possible to the target building.

Research Assistant, EECS, Shandong University *August 2016 - May 2017*

6. Improve Packet Matching Performance for Software-Defined Networking Switch [20]:
- Propose a hybrid TCAM Storage architecture to make full use of high density of nvTCAM and efficient access of sTCAM.
- Design a novel rule migration strategy to improve the rule update performance.
- Contribute a replacement value algorithm to choose the best rules to be evicted in both nvTCAM and sTCAM Storage to further improve packet matching performance.

PUBLICATIONS

1. [**EuroSys'25**] **Xianzhong Ding**, Yunkai Zhang, Binbin Chen, Donghao Ying, Tieying Zhang, Jianjun Chen, Lei Zhang, Alberto Cerpa, Wan Du, "Towards VM Rescheduling Optimization Through Deep Reinforcement Learning", Proceedings of the Twentieth European Conference on Computer Systems, Rotterdam, April 2025.

2. [**HICSS'25**] **Xianzhong Ding**, Wanshi Hong, Zhiyu An, Bin Wang, Wan Du, "Deepot: Parking Lot Identification Using Low-Resolution Satellite Imagery", Hawaii International Conference on System Sciences (HICSS-58), 2025. **Best Paper Nominations**

3. [**TOSN'24**] **Xianzhong Ding**, Wan Du, "Optimizing Irrigation Efficiency Using Deep Reinforcement Learning in the Field", ACM Transactions on Sensor Networks, July 2024.

4. [**TASE'24**] **Xianzhong Ding**, Alberto Cerpa, Wan Du, "Multi-zone HVAC Control with Model-based Deep Reinforcement Learning", IEEE Transactions on Automation Science and Engineering, June 2024.

5. [**DAC'24**] Zhiyu An, **Xianzhong Ding**, Wan Du, "Go Beyond Black-box Policies: Rethinking the Design of Learning Agent for Interpretable and Verifiable HVAC Control", The 61st Design Automation Conference (DAC), February 2024.

6. [**ICLR'24**] Zhiyu An, **Xianzhong Ding**, Wan Du, "Reward Bound for Behavioral Guarantee of Model-Based Planning Agents", To be published in ICLR 2024 Tiny Paper Track.

7. [**TOSN'24**] **Xianzhong Ding**, Alberto Cerpa, Wan Du, "Exploring Deep Reinforcement Learning for Holistic Smart Building Control", ACM Transactions on Sensor Networks, 2024.

8. [**arXiv'24**] Zhiyu An, **Xianzhong Ding**, Yen-Chun Fu, Cheng-Chung Chu, Yan Li, Wan Du, "Golden-Retriever: High-Fidelity Agentic Retrieval Augmented Generation for Industrial Knowledge Base", arXiv preprint arXiv:2408.00798, July 2024.

9. [**Mlsys'23**] **Xianzhong Ding** et al., "Reinforcement Learning for Virtual Machines Rescheduling in Cloud Data Centers", Workshop on ML for Systems at NeurIPS 2023, December 16, New Orleans. **Oral**

10. [**SC-W'23**] **Xianzhong Ding** et al., "HPC-GPT: Integrating Large Language Model for High-Performance Computing", In Workshops of The International Conference on High Performance, Computing, Network, Storage, and Analysis (SC-W 2023).

11. [**Correctness'23**] Le Chen, **Xianzhong Ding** et al., "Data Race Detection Using Large Language Models", In Proceedings of Seventh International Workshop on Software Correctness for HPC Applications (Correctness 2023). ACM, New York, NY, USA, 2023.

12. [**BuildSys'23**] Hamid Rajabi, **Xianzhong Ding**, Wan Du, Alberto Cerpa, "TODOS: Thermal sensOr Data-driven Occupancy Estimation System for Smart Buildings", ACM BuildSys, 2023.

13. [**BuildSys'23**] Zhiyu An, **Xianzhong Ding**, Arya Rathee, Wan Du, "CLUE: Safe Model-Based RL HVAC Control Using Epistemic Uncertainty Estimation", ACM BuildSys, 2023. **Best Paper Runner-Up Award**

14. [**E-Energy'22**] Hamid Rajabi, Zhizhang Hu, **Xianzhong Ding**, Shijia Pan, Wan Du, Alberto Cerpa, "MODES: Multi-sensor Occupancy Data-driven Estimation System for Smart Buildings", ACM International Conference on Future Energy Systems, 2022.

15. [**IPSN'22**] **Xianzhong Ding** and Wan Du, "DRLIC : Deep Reinforcement Learning for Irrigation Control. The International Conference on Information Processing in Sensor Networks, 2022.

16. [**BuildSys'21**] Devanshu Kumar, **Xianzhong Ding**, Wan Du and Alberto Cerpa, "Building Sensor Fault Detection and Diagnostic System, ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Coimbra, Portugal, 2021.

17. [**BuildSys'20**] **Xianzhong Ding**, Wan Du, and Alberto Cerpa, "MB^{2}C : Model-Based deep reinforcement learning for Multi-zone Building Control", ACM BuildSys, 2020.
Best Paper Runner-Up Award, **Best Presentation Award**

18. [**ICDCS'20**] Miaomiao Liu, **Xianzhong Ding**, Wan Du, "Continuous, Real-Time Object Detection on Mobile Devices without Offloading", IEEE ICDCS, 2020.

19. [**BuildSys'19**] **Xianzhong Ding**, Wan Du, and Alberto Cerpa, "OCTOPUS: Deep Reinforcement Learning for Holistic Smart Building Control", ACM BuildSys, 2019.

20. [**LCTES'17**] **Xianzhong Ding**, Zhiyong Zhang, Zhiping Jia and others, "Unified nvTCAM and sTCAM Architecture for Improving Packet Matching Performance", ACM LCTES, 2017.

PROFESSIONAL EXPERIENCE

Research Supervisor, EECS, UC Merced *Jan 2023 - July 2023*

- Supervised and mentored 1 Ph.D. student in the domain of HVAC control using DRL.
- Facilitated brainstorming sessions to generate innovative research ideas and topics.
- Organized regular group discussions to foster collaboration and exchange of ideas.
- Provided guidance and feedback on paper writing, resulting in 1 paper accepted to BuildSys 2023.

PROFESSIONAL ACTIVITIES

- PC members: ATC'25, ATC'24.
- Conference Reviewer: AISTATS'24, ICLR'24, NeurIPS'24, SenSys'23, ICDCS'22, IoTDI'21, INFOCOM'21, Globecom'21, Buildsys'20, ICDCS'20, MASS'20, MSN'19.
- Journal Reviewer: Transactions on Network Science and Engineering '24, IEEE Internet of Things Journal '24, ACM Transactions on Internet of Things'24, Applied Artificial Intelligence'24, IEEE Transactions on Sensor Networks '23, IEEE Internet of Things Journal '22.
- Artifact Evaluation Committee: OSDI'24, ATC'24, OSDI'23, ATC'23.

HONORS AND REWARDS

Best Paper Award Runner-Up at BuildSys 2023	*2023*
Bobcat Summer Fellowship, EECS, UC Merced	*2018, 2020, 2021*
Best Paper Award Runner-Up at BuildSys 2020	*2020*
Best Presentation Award at BuildSys 2020	*2020*
ACM SenSys 2019 NSF student travel grant	*2019*