# Fast Machine Learning for Science Workshop @ ICCAD 2023

Organizers: Javier Duarte, UC San Diego <<u>iduarte@ucsd.edu</u>> Nhan Tran, Fermilab and Northwestern University <<u>ntran@fnal.gov</u>> Paolo D'Alberto, AMD Ryan Kastner, UC San Diego Miaoyuan Liu, Purdue University Seda Ogrenci, Northwestern University

### Reminders

- Speakers, please send us your final slides for uploading to our workshop webpage: <a href="https://fastmachinelearning.org/iccad2023/program.html">https://fastmachinelearning.org/iccad2023/program.html</a>
  - Thank you to those that have already done so!
- Please follow the <u>IEEE code of conduct</u>:
  - Be respectful of others
  - Treat people fairly
  - Avoid injuring others, their property, reputation or employment
  - Refrain from retaliation
  - Comply with applicable laws
- Today's schedule:



2

#### Introduction

- Welcome to Fast Machine Learning for Science Workshop @ ICCAD!
- Builds on ideas in <u>"Applications and Techniques for Fast Machine Learning</u> <u>in Science" white paper</u> and corresponding Fast Machine Learning for Science conference series
  - 2023 @ Imperial College: <u>https://indico.cern.ch/e/fastml2023</u>
  - 2022 @ SMU: <u>https://indico.cern.ch/e/fml2022</u>
  - 2020 online: <u>https://indico.cern.ch/event/924283/</u>
  - 2019 @ Fermilab: <u>https://indico.cern.ch/event/822126/</u>









### Background

- Scientific applications across particle physics, astrophysics, material sciences, quantum information sciences, fusion energy (and beyond!) utilize data acquisition and in situ processing systems which require very low latency and high data bandwidth custom processing elements and real-time control modules
- Integrating data reduction and control applications with real-time machine learning algorithms can enable significant breakthroughs in the sciences









## Goals

- Address emerging challenges and explore innovative solutions in the field of computer-aided design (CAD) for integrated circuits and systems for ultra low latency and high bandwidth scientific applications
- Bring together researchers, practitioners, and industry experts to exchange ideas, share applications, and discuss the latest advancements in CAD methodologies, algorithms, and tools, and **forge new connections**



#### Program

Invited

- Community Vision, Needs, and Progress
- Design Tools Perspective: Catapult + HLS4ML for Inference at the Edge
- Designing Hardware for Machine Learning
- Design Tools Perspective: AMD
- Fast ML in the NSF A3D3 Institute
- Real-time ML at the Linac Coherent Light Source



#### Contributed

- Robust and Efficient Machine Learning for Mission-Critical Applications
- Quantifying the Efficiency of High-Level Synthesis for Machine Learning Inference
- TT-QEC: Transferable Transformer for Quantum Error Correction Code Decoding
- Benchmarking the Robustness of Neural Network-based Partial Differential Equation Solver
- Smart pixel sensors: towards on-sensor filtering of pixel clusters with deep learning
- FKeras: A Sensitivity Analysis Tool for Edge Neural Networks
- FPGA Deployment of LFADS for Real-time Neuroscience Experiments
- Toward Reinforcement Learning-based Rectilinear Macro Placement Under Human Constraints
- ResilienQ: Boosting Fidelity of Quantum State Preparation via Noise-Aware Variational Training