

ModSim 2025– Day One, August 13, 2025

| | |
|---|---|
| 7:30-8:00 am | Registration and Welcome to ModSim |
| Introductions and Keynote Speaker | |
| 8:00-8:30 am | Introduction to the 2025 ModSim Workshop – Adolfy Hoisie |
| 8:30-9:15 am | Keynote Speaker; Norm Jouppi - Architectures for the AI Era, from a TPU to the Extreme Scale |
| Session: Novel ModSim Techniques – Session Lead: Jason Lowe-Power | |
| 9:15-9:45 am | Tushar Krishna - Chakra + ASTRA-sim: An open ecosystem for advancing benchmarking and co-design for diverse AI supercomputers |
| 9:45-10:15 am | Pradip Bose - ModSim Challenges in Secure and Resilient AI System Design |
| 10:15-10:45 am | Break |
| 10:45-11:15 am | Shashank Subramanian: Performance Modeling and System Design Insights for AI Foundation Models |
| 11:15-11:45 am | Lingda Li: PerfVec: Generalizable Representation Learning for Performance and Energy Modeling of Computer Systems |
| 11:45am-12:15 pm | Andreas Gerstlauer: Hierarchical Modeling and Exploration of Large-Scale Superconducting Supercomputer Architectures |
| Lunch Pickup: 12:15 – 1:15 p.m. | |
| 1:15-2:30 p.m. | <u>Panel Topic: Metrics for the AI Era: Moderator: Rob Hoekstra and Ian Karlin</u> <u>Panel Participants: Jens Domke, AI Geist, Jose Moreira, Richard Vuduc,</u> |
| 2:30-3:00 p.m. | Break |
| Session: Revolutionary Technologies – Session Lead: Jason Lowe-Power | |
| 3:00-3:30 p.m. | Xulong Tang - Towards High-Fidelity, Scalable, and Accessible Quantum Computing Systems |
| 3:30-4:00 p.m. | Brad Aimone - Is Neuromorphic Computing Ready for Prime Time? |
| 4:00-4:30 p.m. | Michael James – Scaling Physics Simulations on AI Hardware: ModSim on the Cerebras Wafer-Scale Engine |
| 4:30–5:00 p.m. | Marco Scrbak - Power-Aware Performance: Modeling for the Physical Limits of Extreme Computing |
| 5:00-5:10 p.m. | Closing Remarks |
| End Day One | |

ModSim 2025 – Day Two, August 14, 2025

| | |
|--|---|
| 8:20-8:30am | Opening Remarks – Adolfo Hoisie |
| Session: Workloads and Workflows in the AI Era – Session Lead: Adolfo Hoisie | |
| 8:30-9:00am | Timo Bremer - A National Security AI Platform: driving accelerated science with AI |
| 9:00-9:30am | Dan Ernst - How Should We Measure Supercomputers Targeting Modern Complex Science? |
| 9:30-10:00am | Arvind Ramanathan - Integrating generative AI, multi-scale modeling and simulations with experiments-in-the-loop for biotherapeutic design |
| 10:00-10:30am | Break (30 minutes) |
| 10:30-11:45am | Ad-Hoc Panel: To be Announced – Moderator: Bruce Childers |
| Lunch Pickup: 11:45 am - 12:45 pm | |
| Contributed Presentations Session: – RAPID-FIRE: Session Leaders: Martin Schulz | |
| 12:45-12:55pm | SimAI-Bench: A Benchmarking Suite for AI-coupled HPC Workflows: Harikrishna Tummalapalli |
| 12:55-1:05 pm | Concorde: Fast and Accurate CPU Performance Modeling with Compositional Analytical-ML Fusion: Arash Nasr-Esfahany |
| 1:05-1:15 pm | Virtual Benchmarking for HPC Systems Using ExaDigiT and Calculon: Srishti Kalepu |
| 1:15-1:25 pm | Simulation of Bit-Serial Digital Processing-In-Memory Systems: Siyuan Ma |
| 1:25 -1:35 pm | Leveraging Deep Learning for Architectural Cache Simulation: Pranjali Jain |
| 1:35-1:55 pm | Break (20 minutes) |
| 1:55-2:05 pm | GenZ: Simulating AI Platform Architectures for Distributed Inference of LLM - Ritik Raj |
| 2:05-2:15 pm | Agile Design Space Exploration Framework via Representation-based Performance Modeling: Kuan-Chieh Hsu |
| 2:15-2:25 pm | Q-DREAMER: A Quantum-aware Dynamic Runtime and Execution Adaptive Middleware Emulator for Quantum–HPC Workflows – Shantenu Jha |
| 2:25-2:35 pm | NoX: Technology-Aware Hierarchical Interconnect Modeling for Scalable 2.5D/3D Heterogeneous Chiplet Architectures: Jiho Kim |
| 2:35-2:45 pm | Twins in Action: Symbiotic Simulation for Intelligent HPC Scheduling: Zhiling Lan |
| 2:45-3:05 pm | Break (20 minutes) |
| 3:05-3:15 pm | Simulation-Driven Recommendation Framework for Scalable Distributed AI Training: Adrian Perez Dieguez |
| 3:15-3:25 pm | AI Data Center Power Curtailment - Potential, Challenges, and Implementation: Max Hawkins |
| 3:25-3:35 pm | SCALE-SIM v3: A modular cycle-accurate systolic accelerator simulator for end-to-end system analysis: Ritik Raj |
| 3:35-3:45 pm | A Gradient Descent based Design Space Exploration Framework for Multi-Level Hierarchical AI Accelerator Systems: Sairam Sri Vatsavai |
| 3:45-3:55 pm | Fluid Modeling for HPC Networks: Jason Liu |
| 4:00-5:00 pm | Poster Q&A Session |
| 4:00–6:00 pm | Reception with Refreshments |

ModSim 2025– Day Three, August 15, 2025

| | |
|---|---|
| 8:15-8:30 a.m. | <i>Dr. Sudhakar Yalamanchili – White Paper AWARD Presented by Hyesoon Kim</i> |
| Session: Extreme Computing beyond Exascale – Session Lead: Hyesoon Kim | |
| 8:30-9:00 a.m. | Keren Bergman - Sub-pJ/bit Optical Connectivity for AI Clusters |
| 9:00-9:30 a.m. | Satoshi Matsuoka - From Divergence to Deep Convergence: Charting a Shared Future for Modeling, Simulation, and AI |
| 9:30-10:00 a.m. | Heidi Poxon - Post Exascale: Workloads and Emerging Trends |
| 10:00-10:30 a.m. | Larry Kaplan - Zetta-what and Zetta-how: What comes next after Exascale? |
| 10:30-11:00 a.m. | Break |
| 11:00am - 12:15pm | <u>Panel Topic: Future of System Architecture</u>; <u>Moderator: Shekhar Borkar and Lizy John</u>; <u>Panel Participants: James Ang, Srilatha (Bobbie) Manne, Rajit Manohar, John Shalf, Jeff Vetter</u> |
| 12:15-12:30p.m. | Workshop Wrap-up |