IEEE/ACM

2026 INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN

ICCAD

45th Edition



45TH IEEE/ACM INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN

CONFERENCE PROGRAM

SPONSORS AND ORGANIZERS

IEEE | Association for Computing Machinery | CEDA IEEE Council on Electronic Design Automation | CAS IEEE CIRCUITS AND SYSTEMS SOCIETY | SIGDA | ELECTRON DEVICES SOCIETY

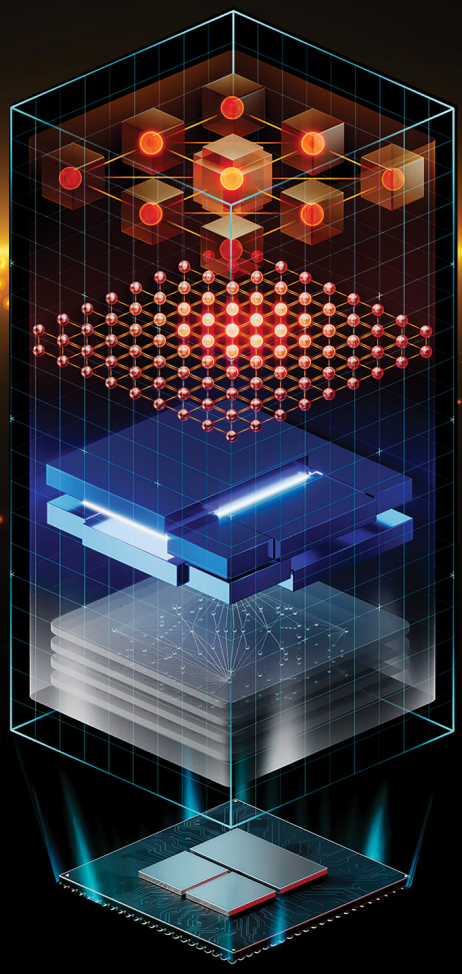
Please visit our website for more information!

ICCAD.COM/2026

cadence®

Design for AI and AI for Design

The Industry's Leading Full-Stack Agentic AI Platform from Chip to System



www.cadence.com

Table of Contents

Executive Committee 4
Venue Map & Information 5
Welcome Message 6
Awards 9
Plenary Speakers 14
Tutorial 16
Panel 17
Workshops 18
Workforce Development 23
Social Functions 24
Sponsors & Exhibitors 25
Program at a Glance 26
Technical Program – Nov 8 34
Technical Program – Nov 9 35
Technical Program – Nov 10 61
Technical Program – Nov 11 90
Technical Program – Nov 12 116

Executive Committee

General Chair

Deming Chen, University of Illinois, USA

Past Chair

Robert Wille, Technical University of Munich, Munich Quantum Software Company, and Software Competence Center Hagenberg GmbH

Program Chair

Ismail S. K. Bustany, AMD, USA

Vice Program Chair

Tsung-Yi Ho, The Chinese University of Hong Kong, China

Tutorial & Special Session Chair

Ron Duncan, Synopsys

Workshop Chair

Hai Helen Li, Duke University

Student Scholar Program Chair

Andy Yu-Guang Chen, National Tsing Hua University, TW

Industry Liaison

William Hung, Cadence, USA

IEEE CEDA Representative

Jiang Hu, Texas A&M University, USA

ACM SIGDA Representative

Christian Pilato, Politecnico di Milano, Italy

Asian Representative

Wei Zhang, The Hong Kong University of Science and Technology, China

European Representative

Ulf Schlichtmann, Technical University of Munich, Germany

Venue Map & Information

San Jose Marriott

Address: 301 S Market St, San Jose, CA 95113

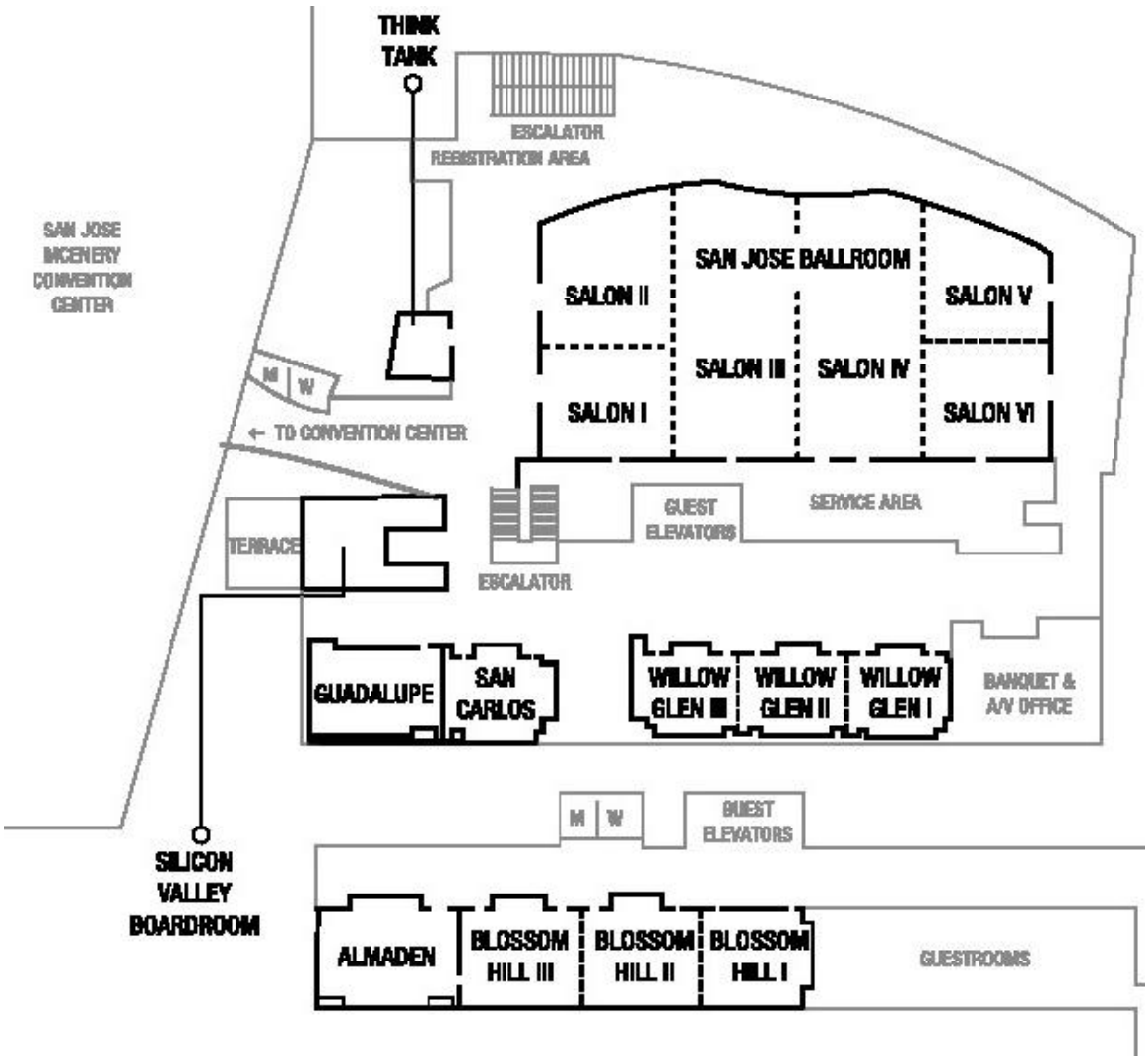
Parking at the Hotel

The San Jose Marriott offers valet parking services only.

The cost is $12.00 per hour, with a total of $36.00 for the first three (3) hours. For four (4) hours or more, a maximum daily rate of $53.00-$65.00 applies.

A second valet-parked vehicle is charged $53.00-$65.00, and oversized vehicles are $80.00 per night, subject to availability. An additional fee of $15.00 will apply per charging session for electric vehicles. For local parking options, please visit: https://parksj.org. Parking is also available at the San Jose Convention Center, which is connected to the San Jose Marriott, Rates are $30.00 per day, without in-and-out privileges, and payment is made directly through the automated machines in the parking structure.

Venue Map



SYNOPSYS®



Partnering with Academia to Accelerate Research and Innovation

At ICCAD, connect with Synopsys Academic & Research to explore research partnerships that connect academic expertise with industry-leading technology.

Meet with our team to discuss research collaboration opportunities.

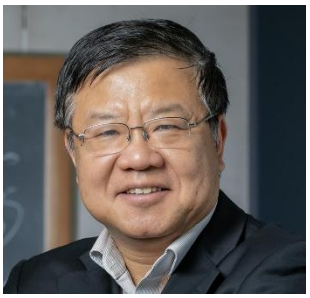
Office Hours:
Monday–Wednesday
Location: Silicon Valley Boardroom

Email **sara-leadership@synopsys.com** to schedule a time to connect.

Scan the QR code to explore research partnership opportunities, or visit:
Synopsys.com/SARA

Welcome Message

Dear Attendees of ICCAD 2026,



Welcome to San Jose, California—and welcome to the 45th International Conference on Computer-Aided Design (ICCAD). It is our immense pleasure to host this vibrant community for the next five days and to offer an exciting program filled with opportunities for exchange, networking, and professional growth.

Jointly sponsored by ACM and IEEE, ICCAD remains the premier venue for presenting cutting-edge research and exploring emerging technologies in computer-aided design. From devices and circuits to architectures, systems, as well as applications—especially in the age of AI and heterogeneous computing—ICCAD brings together thought leaders, innovators, and passionate researchers from around the world.

And ICCAD continues to thrive and evolve! For the seventh consecutive year, the number of submissions has increased. Since 2020, ICCAD has more than doubled its submission volume, a testament to the strength, vitality, and expanding reach of the EDA/CAD research community—and a guarantee of an outstanding technical program.

This year, ICCAD has set a new record for paper submissions, with 1968 abstract submissions across 21 technical tracks from 37 regions and countries. A total of 1434 full papers underwent a rigorous review process, supported by 499 exceptional expert reviewers on our Technical Program Committee. After a rigorous double-blind review process, ultimately, 341 papers were accepted, resulting in a competitive acceptance rate of 23.78%. These papers are organized into 63 sessions, forming the largest technical program in ICCAD's history. A set of outstanding papers were nominated for the IEEE/ACM William J. McCalla ICCAD Best Paper Award in the Front-End and Back-End categories. Following a rigorous evaluation by a panel of expert judges, two papers were ultimately selected to receive this prestigious award—one in each category. A separate panel of expert judges also selected the recipient of the ICCAD 10-Year Retrospective Most Influential Paper Award.

In addition to the regular sessions, we are proud to feature:

- 10 special sessions exploring the future of CAD—including microfluidics, chiplet systems, hardware-software co-design for AI, LLM foundations, neuromorphic computing, and more. It includes 30 invited papers from field matter experts.
- An embedded tutorial on hardware security and quantum architecture design.
- And a panel discussion on Thermal Co-Design for AI Chiplet Systems.

Our program is crowned by four distinguished keynotes from leaders in the field:

- On Monday morning, Ralph Wittig (AMD) will present "AI-Driven Chip Design and the Future of Computing."
- On Tuesday morning, Vaughn Betz (University of Toronto and Cerebras Systems) will present "Spatial Computing Architectures: Opportunities and CAD Flows in the Deep Learning and More-than-Moore Era."
- At noon on Tuesday, Chia-Lin Yang (National Taiwan University) will deliver the IEEE CEDA Luncheon Keynote, "Breaking the AI Memory Wall: A Paradigm Shift Toward Memory-Centric Computing."
- On Wednesday morning, Paul Cunningham (System Verification Group, Cadence) will present "Design for AI, AI for Design."

Welcome Message (cont)

Finally, on Thursday, we invite you to join several co-located workshops covering topics such as: Memory-Centric Computing for LLM Inference, QCAS Quantum Computing Applications and Systems, Safety and Security in Heterogeneous Open System-on-Chip Platforms, Sustainable Hardware Security, and Top Picks in Hardware & Embedded Security. These workshops offer a space for focused discussion and deeper engagement with emerging areas of research.

In addition, throughout the week, ICCAD will also host sessions dedicated to workforce development and student engagement. Our integrated Student Scholar Program supports young researchers through activities such as the CAD Contest, CADathlon, Student Research Competition (SRC), and the Job Fair. Thanks to generous support from both IEEE CEDA and ACM SIGDA, we are proud to welcome many first-time attendees and provide 40+ travel grants to help students connect with the ICCAD community.

None of this would be possible without the tireless efforts of our volunteers—including the executive committee, technical program committees for both the regular research tracks and special sessions track, workshop organizers, and student activity coordinators. We also want to extend our heartfelt thanks to our industrial sponsors for their continued support and strong belief in ICCAD's mission. This year, we are thrilled to have a record-breaking 11 industrial sponsors!

We also want to make ICCAD a venue for fun, connection, and community building! Monday's Welcome Reception and SRC Poster Session will offer great opportunities for networking and reconnecting with colleagues while learning about cutting-edge research from our outstanding student researchers. Tuesday night's Job Fair will provide an excellent opportunity for students and industry representatives to mingle, connect, and explore career opportunities over food and refreshments. And Wednesday's Gala Dinner will feature wonderful food, great company, and live music as we come together to celebrate our community!

Finally, the success of ICCAD hinges on your participation. We hope you will have a great time, gain new insights, share recent accomplishments, reconnect with old friends, and forge new connections. Whether you're here to present, learn, collaborate, or simply engage with the community—welcome to ICCAD 2026. We hope you enjoy the conference, the city of San Jose and the vibrant spirit that makes ICCAD so special.

Deming Chen
General Chair
University of Illinois Urbana-Champaign

Ismail S. K. Bustany
Program Chair
AMD Inc.

Tsung-Yi Ho
Vice Program Chair
The Chinese University of Hong Kong

Ron Duncan
Tutorial and Special Sessions Chair
Synopsys Inc.

Hai Helen Li
Workshop Chair
Duke University
on behalf of the entire Executive Committee of ICCAD 2026

Awards

Best Reviewer Award

Rongjian Liang, NVIDIA
Hiromitsu Awano, Nagoya University
Hongduo Liu, Huawei
Kristofer Vorwerk, Microchip Technology
Liang Chen, Shanghai University
Wei Xing, University of Sheffield
Zain Ul Abideen, University of Idaho
Pei-Yu Lee, Mediatek
Bingyao Li, UC Riverside
Marcello Traiola, Inria / IRISA

This award recognizes the dedicated service of reviewers and the outstanding quality of reviews received by TPC members. The awardees above were selected from a pool of candidates nominated by the TPC Track co-chairs, based on the following criteria:

- **Insightfulness and Clarity:** detailed comments with strong insights, e.g., feedback is clear and specific without vagueness.
- **Thoroughness and Quality:** thorough and high quality, e.g., comments are comprehensive and address both strengths and areas for improvement.
- **Originality of Contribution:** own review instead of relying on the secondary reviewer.
- **Consistency and Fairness:** the review scores should be consistent with the review content; overall tone and recommendations align with the content of the feedback.
- **Impact on Authors:** consider whether the review would genuinely help the author improve their work (a key sign of a high-quality review).
- **Professionalism:** ensure the review is respectful, free of bias, and professionally written.


DESIGN THE COMPUTE THAT WILL BRING PHYSICAL AI TO LIFE
Tomorrow's autonomous machines must turn perception into action under real-world constraints such as latency, power, safety, and determinism. AMD heterogeneous computing helps researchers address these challenges, providing a foundation to transform new ideas into real-world physical AI systems.
AMD
together we advance_
Research what's next.
Build what's possible with AMD.

Awards

William J. McCalla ICCAD Best Paper Award -

Front End: ***"Adapting DNNs to Embrace Voltage Scaling Induced Errors via Accumulation-Aware Quantization"***
Alexis Maras, Panagiotis Chaidos, Georgios Alexandris, Dimitrios Soudris, Sotirios Xydis

Backend: ***"NeuroAbs: A Neuro-Symbolic RTL Abstraction Framework for Property Checking Acceleration"***
Zhiyuan Yan, Xiaofeng Zhou, Ziyue Zheng, Ziyi Yang, Wenbin Che, Wei Zhang, Yangdi Lyu, and Hongce Zhang

ICCAD 10 Year Retrospective Most Influential Paper Award

"Reusability is FIRRTL ground: Hardware construction languages, compiler frameworks, and transformations"
Adam M. Izraelevitz, Jack Koenig, Patrick Li, Richard Lin, Angie Wang, Albert Magyar, Donggyu Kim, Colin Schmidt, Chick Markley, Jim Lawson, Jonathan Bachrach

2026 Outstanding Service Recognition Award



Robert Wille (Technical University of Munich & Software Competence Center Hagenberg GmbH)
"for outstanding service to the EDA Community as ICCAD General Chair in 2025"

Awards (cont.)

Best Reviewer Award

Rongjian Liang — NVIDIA

Hiromitsu Awano — Nagoya University

Hongduo Liu — Huawei

Kristofer Vorwerk — Microchip Technology

Liang Chen — Shanghai University

Wei Xing — University of Sheffield

Zain Ul Abideen — University of Idaho

Pei-Yu Lee — Mediatek

Bingyao Li — UC Riverside

Marcello Traiola — Inria / IRISA

This award recognizes the dedicated service of reviewers and the outstanding quality of reviews received by TPC members. The awardees above were selected from a pool of candidates nominated by the TPC Track co-chairs, based on the following criteria:

- **Insightfulness and Clarity:** detailed comments with strong insights, e.g., feedback is clear and specific without vagueness.
- **Thoroughness and Quality:** thorough and high quality, e.g., comments are comprehensive and address both strengths and areas for improvement.
- **Originality of Contribution:** own review instead of relying on the secondary reviewer.
- **Consistency and Fairness:** the review scores should be consistent with the review content; overall tone and recommendations align with the content of the feedback.
- **Impact on Authors:** consider whether the review would genuinely help the author improve their work (a key sign of a high-quality review).
- **Professionalism:** ensure the review is respectful, free of bias, and professionally written.



You don't want
to optimize
today's tools.
You want to
invent tomorrow's.

We are hiring.

imec

Plenary Speakers



AI-Driven Chip Design and the Future of Computing
Ralph Wittig, AMD

Monday | November 9, 2026 | 9:00 - 10:00
Room: San Jose Ballroom

Abstract: Artificial intelligence is beginning to transform computing at both ends of the design process: it is changing how we design chips, and it is changing the kinds of chips and systems we need to build. AI-assisted and agentic design methodologies promise to compress development cycles, expand architectural exploration, and automate increasingly complex tasks across design exploration, specification, RTL generation, verification, physical design, and system optimization. At the same time, rapidly evolving model architectures, inference workloads, and agentic applications are creating new requirements for latency, interactivity, energy efficiency, cost per token, memory systems, and programmability.

This talk will explore what comes next when hardware, software, and AI models are no longer optimized independently. Future systems will increasingly require continuous co-design across models, algorithms, architectures, packaging, interconnects, and software, with AI participating directly in the optimization loop. Open-weight models and enterprise-controlled AI platforms may further shift where innovation occurs and who controls data, deployment, and inference economics. Looking beyond today's data center, physical AI and emerging world-action models will introduce stringent real-time and energy constraints, while hybrid AI, HPC, and quantum systems may open new paths for scientific discovery. The central question for the design community is therefore broader than how AI can improve CAD: what happens when AI becomes both a workload we design for and an active participant in designing the next generation of computing systems?



Spatial Computing Architectures: Opportunities and CAD Flows in the Deep Learning and More-than-Moore Era
Vaughn Betz, University of Toronto, Cerebras Systems

Tuesday | November 10, 2026 | 8:30 - 9:30
Room: San Jose Ballroom

Abstract: To meet the stringent energy-efficiency and performance needs of deep learning workloads, minimizing data movement is crucial. This trend plays to the strengths of spatial architectures, including not only FPGAs but also processor arrays with mesh interconnects like the Cerebras Wafer Scale Engine and AMD AI Engine arrays. Spatial architectures also have characteristics that make them particularly well-suited to More-than-Moore integration that stacks multiple dice, connects them with interposers, or leverages wafer-scale integration. To explore this broad architecture space, we need flexible tools that can exploit the features of devices ranging from new FPGAs to novel spatial processor arrays. In this talk, I will discuss the strengths of these spatial platforms across a range of deep learning applications, outline how the different platforms are exploiting More-than-Moore integration, and give an overview of the tools we are creating to help enable and explore 3D FPGAs, new embedded FPGAs, and spatial processor arrays.

Plenary Speakers (cont.)



Breaking the AI Memory Wall - A Paradigm Shift Toward Memory-Centric Computing
Chia-Lin Yang, National Taiwan University

Tuesday | November 10, 2026 | 11:30 - 13:00
Room: San Jose Ballroom

Abstract: Rapid advancements in artificial intelligence are driving an exponential increase in model complexity and KV cache footprints, placing unsustainable demands on memory bandwidth and capacity. Overcoming this "memory wall" is imperative to scaling next-generation AI computing platforms. Bridging this gap demands a holistic approach spanning algorithmic efficiency, innovative hardware architectures, and emerging memory technologies. In this talk, I'll walk through recent developments reshaping system architecture to address the AI-era memory wall, and share our recent work on compute-in-memory architectures, computational storage, and memory orchestration for LLM serving.



Design for AI, AI for Design
Paul Cunningham, System Verification Group, Cadence

Wednesday | November 11, 2026 | 8:30 - 9:30
Room: San Jose Ballroom

Abstract: The AI Supercycle is driving unprecedented demand for semiconductors and systems, with some estimates putting the semiconductor industry at a value of over $2 Trillion by the end of this decade. In his keynote, Dr. Cunningham will discuss how EDA is providing solutions to address this surge in demand, enabling a "virtuous circle" for our customers: accelerating teams as they design the AI chips and systems to build out AI infrastructure, while embedding AI in the IC and system design tools to deliver greater productivity and automation than ever before.

Tutorial

Machine Learning for Quantum Circuit Synthesis, Mapping, and Architecture Search: Toward Quantum CAD

Monday | November 9, 2026 | 10:30 - 15:00

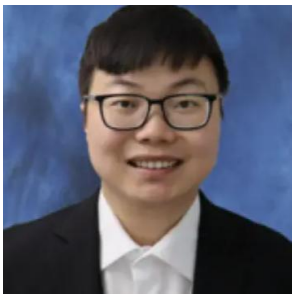
This tutorial presents a design-automation-oriented introduction to machine learning methods for quantum circuit synthesis, mapping, and architecture search. Rather than treating quantum machine learning (QML) as a standalone application area, the tutorial frames QML, quantum reinforcement learning (QRL), and search-based optimization as algorithmic tools for quantum CAD under realistic hardware constraints, including native gate sets, limited qubit connectivity, circuit depth, noise, and resource overhead. Participants will learn how parameterized quantum circuits define structured design spaces, how reinforcement learning, evolutionary optimization, and differentiable architecture search can be used for design-space exploration, and how these methods interact with compilation, mapping, and benchmarking workflows. Through hands-on examples using open-source quantum simulators such as Qiskit and PennyLane in Google Colab, attendees will explore hardware-aware ansatz discovery and circuit optimization using practical design metrics such as fidelity, two-qubit gate count, depth, and robustness. The tutorial concludes with an outlook on scalable Quantum CAD, including emerging self-programming and automated quantum design frameworks.



Kuan-Cheng Chen
Imperial College London



Samuel Yen-Chi Chen
Wells Fargo



Zhiding Liang
Chinese University of Hong Kong

Panel

Thermal Co-Design for AI Chiplet Systems: Simulation, Packaging Innovation, and System-Level Optimization

Tuesday | November 10, 2026 | 10:00 - 11:30

Advanced 2.5D and 3D heterogeneous AI chiplet packages have crossed a threshold where thermal management is a co-design challenge spanning architecture, packaging, and system-level optimization simultaneously. Chiplet partitioning sets the thermal envelope that packaging must manage; packaging architecture determines what simulation can predict; system-level power management shapes what the package must survive. No single team owns this problem, and no single methodology addresses it. Three forces have converged: simulation methods embedded in placement flows lose structural validity at manufacturing-resolution 2.5D/3D geometries—not merely precision; packaging innovation is moving faster than the tools that model its thermal implications; and system-level thermal optimization remains disconnected from the package co-design loop. This panel convenes experts across simulation, advanced packaging architecture, production AI thermal design, and system-level optimization to debate the path forward. Three positions structure the discussion: simulation first—physics-intelligence simulation is the missing link for fast, accurate thermal and warpage prediction at physical dimensions; packaging innovation first—advanced interposers, embedded thermal structures, and chiplet-aware stacking can reduce the thermal problem before simulation must solve it; and system-level ownership first—thermal co-design spans chip architects, package engineers, system designers, and foundry partners, requiring a closed-loop workflow so improvements propagate across every layer.



Ravi Mahajan
Intel Foundry



Rahima Mohammed
Vinci4D



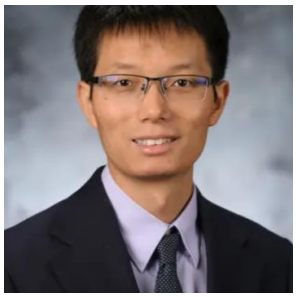
Maziar Raissi
University of California, Riverside



Gamal Refai-Ahmed
AMD



Mircea Stan
University of Virginia



Zheng Zhang
University of California, Santa Barbara

Workshops

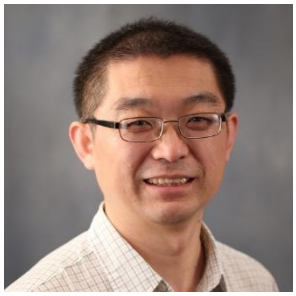
The 8th Workshop on Top Picks in Hardware and Embedded Security

Wednesday | November 11, 2026 | 10:00 – 17:45

Over the past two decades, research in hardware and embedded security has yielded thousands of papers and more than a dozen high-quality conferences and workshops dedicated to hardware security. Many of the techniques are being adopted by industry, (e.g., side-channel resistance techniques). Many have led to interesting intellectual contributions which impact go beyond hardware research community (e.g, physical-unclonable functions). The first workshop on Top Picks in Hardware and Embedded Security (Top Picks) was founded and co-located with ICCAD in 2018 to highlight the works that have made the greatest intellectual contributions. Since then, we have organized this workshop annually and co-located with ICCAD including the years when ICCAD was held virtually or abroad. We would like to continue this tradition by proposing Top Picks 2026.



Shuwen Deng
Tsinghua University



Gang Qu
University of Maryland

Workshops (cont.)

2026 QCAS Quantum Computing Applications and Systems Workshop

Thursday | November 12, 2026 | 8:30 – 18:00

The QCAS 2026 Workshop addresses the systems engineering challenges of building scalable, fault-tolerant quantum computers. The field has reached an inflection point: multiple hardware platforms, including superconducting, neutral atom, and trapped-ion systems, have demonstrated below-threshold quantum error correction performance with real-time decoding. Industry roadmaps now project modular multi-chip quantum processors within the next few years, and national programs such as DARPA HARQ are investing in heterogeneous architectures that combine distinct qubit types for processing, memory, and communication. These developments collectively shift the central question from whether fault-tolerant quantum computing is possible to how systems can be engineered to achieve it at useful scale. QCAS 2026 is organized around four tracks spanning the full system stack: quantum error correction at scale, fault-tolerant compilation and resource estimation, distributed and modular architectures, and quantum control and hardware-software co-design. A distinguishing emphasis of this edition is the elevation of resource estimation as a unifying design methodology, serving as the quantum analog of design space exploration in classical EDA.



Yufei Ding
University of California San Diego



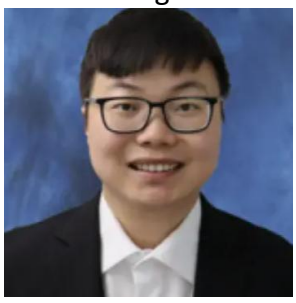
Zichang He
JPMorganChase



Tilas Kabengele
The Cincinnati Insurance Companies



Gilles Lamant
Cadence Design Systems



Zhiding Liang
Chinese University of Hong Kong



Pengyu Liu
Carnegie Mellon University



Yidong Zhou
Rutgers University

Workshops (cont.)

Memory-Centric Computing for LLM Inference

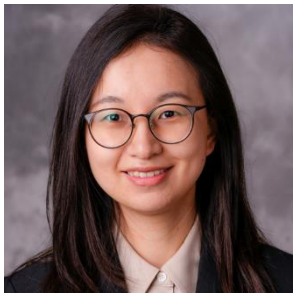
Thursday | November 12, 2026 | 8:30 – 12:10

The rapid growth of large language models (LLMs) is reshaping modern AI systems, driving unprecedented demands on memory capacity, bandwidth, energy efficiency, and scalability, particularly for resource-constrained edge AI deployments. As inference workloads become increasingly memory-bound, conventional GPU-centric platforms struggle to efficiently support long-context reasoning, high-throughput serving, and large-scale deployment. Emerging paradigms such as processing-in-memory (PIM), 3D-integrated computing, heterogeneous architectures, wafer-scale systems, and hardware-software co-design offer promising alternatives for overcoming these bottlenecks.

This workshop brings together researchers and practitioners from computer architecture, EDA, systems, compilers, machine learning, memory technologies, and industry to discuss the future of efficient AI inference. The workshop will highlight recent advances in memory-centric computing, including DRAM-based PIM, 3D-stacked PIM, and emerging non-volatile memory (NVM)-based computing architectures. Additional topics include heterogeneous CPU/GPU/NPU-PIM systems, wafer-scale AI computing, memory hierarchy optimization, compiler and runtime support, algorithm-system co-design, and EDA tools for scalable and energy-efficient LLM inference.



Chenchen Liu
Beihang University



Xiaoxuan Yang
University of Virginia

Workshops (cont.)

SSH-SoC: Safety and Security in Heterogeneous Open System-on-Chip Platforms

Thursday | November 12, 2026 | 8:00 – 16:30

The slowdown of technology scaling has shifted computer architecture toward heterogeneous, domain-specific systems. Modern SoCs increasingly combine custom accelerators with specialized interconnects, peripherals, and sensing components to meet demanding performance, efficiency, and real-time requirements. Open hardware and software ecosystems are accelerating this transition by enabling transparent design reuse, rapid prototyping, and reproducible validation.

A common strategy to manage SoC complexity is hierarchical development: critical blocks are built in-house at RTL, others are generated with HLS, and highly specialized IP is integrated from third parties. While effective for productivity, this approach makes system integration the key bottleneck. Integration errors, weak specifications, and limited system-level verification can introduce serious security, safety, and performance risks, especially in mission-critical deployments.

This workshop focuses on methods, tools, and early-stage ideas that improve trustworthy integration of next-generation heterogeneous SoCs using open platforms. We explicitly welcome AI-enabled approaches, including LLM-assisted design-space exploration, specification mining, IP integration support, verification/test generation, and security analysis. The goal is to advance open, safe, and high-performance SoC development through collaborative research and practical experimentation.



Luca Benini
ETH-Zurich/University of Bologna



Brandon Reagen
NYU



Francesco Restuccia
UC San-Diego



Davide Rossi
University of Bologna



Biruk Seyoum
Columbia University

Workshops (cont.)

SUSHI'26: Sustainable Hardware Security

Thursday | November 12, 2026 | 8:00 – 17:00

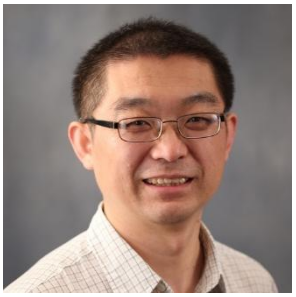
The pursuit of digital sovereignty, global semiconductor shortages, and geopolitical interests drive worldwide initiatives to bolster semiconductor technology and manufacturing. Hardware security is vital at the core of all computing systems, as insecure hardware puts critical systems and society at risk.

However, in recent years we are observing the discovery of a growing number of hardware design and implementation vulnerabilities that unprivileged software could exploit, leading to the potential exposure of sensitive data or compromise of the whole computing system. This new attack paradigm greatly shadows decades of system security research. Existing solutions are often ad-hoc, limited, inefficient, and specific.

Tackling these challenges calls for radically new approaches to the security-by-design for hardware, security-aware electronic design automation, resilient microarchitecture and architectures, and scalable assurance.



Jason Fung
Intel



Gang Qu
University of Maryland



JV Rajendran
Texas A&M



Sanjay Rekhi
NIST



Ahmad-Reza Sadeghi
TU Darmstadt

Workforce Development

Student Research Competition

Posters: Monday | November 9, 2026 | 18:30 - 20:00
Oral Presentations: Tuesday | November 10, 2026 | 13:00 - 14:30

The ACM Student Research Competition (SRC) provides undergraduate and graduate students who are ACM members with a unique opportunity to present their work, exchange ideas, and gain recognition from the international computing community. Organized by ACM SIGDA and hosted at ICCAD 2026, the competition features poster and presentation sessions where students showcase their research in design automation. Participants benefit from feedback by experts, networking with academic and industry leaders, and the chance to win prizes. Separate categories for undergraduate and graduate students ensure fair evaluation, with submissions judged on novelty, technical merit, and contributions to the field.

Job Fair

Tuesday | November 10, 2026 | 17:45 - 19:15

ACM SIGDA and IEEE CEDA will host the 5th annual EDA Job Fair at ICCAD 2026, providing students and professionals the opportunity to connect with leading companies and academic institutions. Attendees can learn about current openings, network directly with recruiters, and explore career opportunities in electronic design automation. Participants are encouraged to submit their CVs in advance, and ICCAD registration (minimum one-day) is required.

CADathlon@ICCAD 2026

Sunday | November 8, 2026 | 8:00 - 17:00

The CADathlon is a challenging, all-day programming competition focusing on practical problems at the forefront of Computer-Aided Design and Electronic Design Automation in particular. The contest emphasizes the knowledge of algorithmic techniques for CAD applications, problem-solving and programming skills, and teamwork.

As the “Olympic games of EDA,” the contest brings together the best and the brightest of the next generation of CAD professionals. It gives academia and the industry a unique perspective on challenging problems and rising stars, and it also helps attract top graduate students to the EDA field.

Social Functions

Welcome Reception

Monday | November 9, 2026 | 18:30 - 20:00 | Ballroom Foyer

All conference registrants are invited to join us for the Welcome Reception on Monday, November 9, at The San Jose Marriott. Start the conference on a high note with an evening of networking, hors d'oeuvres, and refreshments as you connect with fellow attendees in a relaxed and friendly setting. We look forward to welcoming you in Munich!

Gala Dinner

Wednesday | November 11, 2026 | 18:15 - 20:15 | San Jose Ballroom

Join us for an evening at the San Jose Marriott, where the rest of the conference is taking place. Enjoy live music, food, drinks, and lively conversation in a welcoming setting alongside fellow conference attendees. Access is included with your conference registration - no additional ticket required.



Sponsors & Exhibitors

Platinum Sponsors

cadence®

SYNOPSYS®

Gold Sponsors

altera™

AMD

Architect Labs

imec

Silver Sponsors

AGENTRYS

RICURSIVE INTELLIGENCE

SIEMENS

TENX SEMI

Conference Sponsors

CAS
IEEE CIRCUITS AND SYSTEMS SOCIETY

sigda
acm

CEDA®
IEEE Council on Electronic Design Automation

IEEE
Advancing Technology for Humanity

ELECTRON DEVICES SOCIETY®

	ICCAD 2026 Block Schedule - Sunday	
	11/8/2026	
Room	Almaden U Shape 25	Willow Glen I/II Custom Layout
8:00 AM	Breakfast	CADathlon Breakfast
9:00 AM	IEEE CEDA Executive Committee Meeting	CADathlon
10:30 AM	AM Break	
11:00 AM	IEEE CEDA Executive Committee Meeting	
12:00 PM		CADathlon Lunch
12:30 PM	Lunch	
1:00 PM		CADathlon
1:30 PM	IEEE CEDA Executive Committee Meeting	
15:00	PM Break	CADathlon Coffee Break
15:30	IEEE CEDA Executive Committee Meeting	CADathlon
17:00		
18:00		

	ICCAD 2026 Block Schedule - Monday							
	11/9/2026							
Room	Willow Glen I/II Theater 100	Willow Glen III Theater 55	Blossom Hill I/II Theater 100	Blossom Hill III Theater 55	Guadalupe Theater 50	Almaden Theater 50	San Jose Ballroom Theater 640	San Carlos Theater 55
7:30 AM	Registration Ballroom Foyer							
8:15 AM	Opening/ Awards Ceremony Room: San Jose Ballroom							
9:00 AM	Keynote - Ralph Wittig Room: San Jose Ballroom							
10:00 AM	Coffee Break | Exhibits Room: Ballroom Foyer							
10:30 AM	Tutorial: Machine Learning for Quantum Circuit Synthesis, Mapping, and Architecture Search: Toward Quantum CAD	The Physical Frontier: Analog Compute-in-Memory and Mixed-Signal Architectures	Special Session: Chiplets at Scale for GenAI and LLMs: Design, Reliability, and Security Challenges Ahead	Routing and Co-Optimization in Physical Design	Crafting Circuits: From Technology to Optimization Room: Guadalupe Theater: 50		Beyond Wirelength: Differentiable Placement for Timing, Routability, and PPA Closure	From Topology to Tapeout: AI-Driven Analog Circuit Sizing, Compact Modeling, and IP Retrieval
12:00 PM	Lunch Room: San Jose Ballroom							
13:30	Tutorial: Machine Learning for Quantum Circuit Synthesis, Mapping, and Architecture Search: Toward Quantum CAD		Special Session: AI-Assisted Chip Design Contest: 2026 ICCAD CAD Contest	Clock Network Design and Automation	Leakage in the Machine	Learning to Attack, Learning to Defend	Scalable Optimization for Floorplanning, Partitioning, and Chiplet-Based Design	Intelligent Design Optimization
15:00	Coffee Break | Exhibits Room: Ballroom Foyer							
15:30								

ICCAD 2026 Block Schedule - Monday

11/9/2026

Room	Willow Glen I/II Theater 100	Willow Glen III Theater 55	Blossom Hill I/II Theater 100	Blossom Hill III Theater 55	Guadalupe Theater 50	Almaden Theater 50	San Jose Ballroom Theater 640	San Carlos Theater 55
15:30	Special Session: Breaking the Memory Wall for LLMs: Computing-in-Memory as the Catalyst for Scalable and Efficient Generative AI	Pushing the Precision Floor of LLM Quantization	System-level Techniques for AI Acceleration	Special Session: 3D Nanofabric: A Pathway to Post-Denard CMOS Scaling	Accelerating Trustworthy Cryptography	LLM-Driven and Rule-Based Verification and Test Generation		Reimagining Logic Synthesis
16:30								
17:00	Transition							
17:15		Efficient and Reliable AI Models for Edge Inference	ACM SIGDA Town Hall	Thermal Design Across Scales: Interconnects, 3D Memory, and Mobile Systems	Making Quantum Resources Count: Search, Noise, and Simulation	Proofs in Practice: Zero-Knowledge Engines and Hardware Verification	Synopsys Sponsor Session	
18:15								
18:30	Welcome Reception & SRC Poster Session Room: Ballroom Foyer							
18:45								
19:00								
19:15								
19:30								
19:45								
20:00								

ICCAD 2026 Block Schedule - Tuesday

11/10/2026

Room	Willow Glen I/II Theater 100	Willow Glen III Theater 55	Blossom Hill I/II Theater 100	Blossom Hill III Theater 55	Guadalupe Theater 50	Almaden Theater 50	San Jose Ballroom Theater 640	San Carlos Theater 55
8:00 AM	Registration Ballroom Foyer							
8:30 AM	Keynote - Vaughn Betz Room: San Jose Ballroom							
9:30 AM	Coffee Break | Exhibits Room: Ballroom Foyer							
10:00 AM	Thermal Co-Design for AI Chiplet Systems: Simulation, Packaging Innovation, and System-Level Optimization Panel	Edge Intelligence: Novel Arithmetic, State-Space Models, and Agile Compilation	Special Session: Cross-Layer Co-Design for Multi-Chip Systems: Reliability, Security, and LLM-Centric Acceleration	Agentic AI for EDA Workflows and Design Automation	Special Session: From Copilots to AI-Native Systems: The Future of Agentic EDA	Emerging-Device Memory & Compute-in-Memory	Hardware-Accelerated Verification and Physical Design	Faster Than SPICE: Physics-Guided AI for RF Passive Modeling, EM Simulation, and Parasitic Extraction
11:30 AM	CEDA Luncheon - Chia-Lin Yang - RECORD THIS SESSION Room: San Jose Ballroom							
1:00 PM	Student Research Competition (oral)	Breaking the Memory Wall: XPU-PIM Orchestration and Digital CIM Architectures	Advances in Multi-Scale Power, Thermal, and PDN Optimization	LLMs for RTL Design, Verification, and Debugging	Beyond Determinism: Harnessing Approximate and Stochastic Computing in Silicon	AI for Security and Security for AI	Reliability and Testing at Scale	Automated at Last: Intelligent Frameworks for Analog and Mixed-Signal Layout Automation
14:30	Coffee Break | Exhibits Room: Ballroom Foyer							
15:00	Special Session: Design Automation for Early Fault-Tolerant Quantum Computing	Beyond Text: Hardware Architectures for Spatial, Temporal, and Vision Generation	Advances in Statistical, Multi-Corner, and AI-Driven Timing Analysis	AI for Physical Design and Semiconductor Manufacturing	Computing in Memory	AI Hardware Architectures and Intelligent Design Automation	Crafting Cells at Advanced Nodes: SAT, LLMs, and Timing-Aware Layout Automation	Beyond Logic Optimization

	ICCAD 2026 Block Schedule - Tuesday							
	11/10/2026							
Room	Willow Glen I/II Theater 100	Willow Glen III Theater 55	Blossom Hill I/II Theater 100	Blossom Hill III Theater 55	Guadalupe Theater 50	Almaden Theater 50	San Jose Ballroom Theater 640	San Carlos Theater 55
16:30	Transition							
16:45			AI-Driven PPA Modeling, EDA Tuning, and Physical Design Optimization	Chasing Speed: Accelerating AI and Beyond	Ising Machines, Photonic, Hyperdimensional Memory	Meeting the Deadline: Timing Guarantees for AI on Heterogeneous Platforms	Constraint-Aware Optimization for FPGA and PCB Design Automation	From Perception to Action: Efficient and Reliable Physical AI at the Edge
17:45	Job Fair Room: Ballroom Foyer							
19:15								

	ICCAD 2026 Block Schedule - Wednesday							
	11/11/2026							
Room	Willow Glen I/II Theater 100	Willow Glen III Theater 55	Blossom Hill I/II Theater 100	Blossom Hill III Theater 55	Guadalupe Theater 50	Almaden Theater 50	San Jose Ballroom Theater 640	San Carlos Theater 55
8:00 AM	Registration Ballroom Foyer							
8:30 AM	Keynote - Paul Cunningham Room: San Jose Ballroom							
9:30 AM	Coffee Break | Exhibits Room: Ballroom Foyer							
10:00 AM		Squeezing the Void: Hardware-Algorithm Co-Design for Attention and Sparsity	Multiphysics Electromigration and Reliability Analysis for Advanced Interconnects	Machine Learning for Circuit, Device, and Technology Modeling	Emerging Design Frontiers: Quantum Circuits, Photonics, and 3D Systems		Powering the Third Dimension: Co-Optimizing PDNs, Signal Integrity, and Thermal Performance	Top Picks Workshop
11:30 AM	Lunch Room: San Jose Ballroom							
1:00 PM	Special Session: GenAI for Chip Design Across the Lifecycle: Design, Optimization, Security, and Reliability	Next-Gen Scale: MoE Serving, Speculative Decoding, and Distributed Memory Hierarchies	Emerging System-Level Design Methodologies	Efficient AI Models and Learning Algorithms for EDA	Event-Driven Intelligence: Neuromorphic Architectures, SNN Compilation, and Energy-Efficient Edge Learning	Circuits and Accelerators for Secure and Private Computing		Top Picks Workshop
14:30	Coffee Break | Exhibits Room: Ballroom Foyer							

ICCAD 2026 Block Schedule - Wednesday

11/11/2026

Room	Willow Glen I/II Theater 100	Willow Glen III Theater 55	Blossom Hill I/II Theater 100	Blossom Hill III Theater 55	Guadalupe Theater 50	Almaden Theater 50	San Jose Ballroom Theater 640	San Carlos Theater 55
15:00	Special Session: Generative AI Meets Open-Source EDA: Accelerating Design Productivity and Talent Development	Special Session: From Ad Hoc to Standardized: Early Explorations of 3D IC Design Flows Across Integration Paradigms	From Layout Decomposition to Mask Optimization: Advances in Multiple Patterning Lithography	Foundation Models, Benchmarks, and Trustworthy AI for Hardware Design	Quantum Computing 1	Squeezing Intelligence In: Model Compression and Bespoke Hardware for the Extreme Edge	Learning the Heat: AI for Thermal, Timing, and Reliability in Advanced 2.5D/3D Integration	Top Picks Workshop
16:30	Transition							
16:45	Special Session: Hierarchical IR-Drop for Chiplets and 3D-ICs: Scalable PGA, Correlation, and Interoperability		Emerging Memory and Storage Architectures	Co-Design for Emerging Platforms: Neuromorphic Sensing and Programmable Biochips	Photonics & Optoelectronic Systems Design		Smarter by Design: Learning and Optimization for Chiplet and 3D Systems	Top Picks Workshop
17:45								
18:15	Gala Dinner Room: San Jose Ballroom							
20:00								

ICCAD 2026 Block Schedule - Thursday

11/12/2026

Room	Willow Glen I/II Theater 100	Willow Glen III Theater 55	Blossom Hil I/II Theater 100	Blossom Hill III Theater 55	Guadalupe Theater 50
7:30 AM	Registration Room: Ballroom Foyer				
7:45 AM					
8:00 AM		SSH-SoC: Safety and Security in Heterogeneous Open System-on-Chip Platforms Workshop	Memory-Centric Computing for LLM Inference Workshop	SSH-SoC: Safety and Security in Heterogeneous Open System-on-Chip Platforms Workshop	SUSHI'26: Sustainable Hardware Security Workshop
8:30 AM	2026 QCAS Quantum Computing Applications and Systems Workshop				
9:30 AM	Coffee Break | Exhibits Room: Ballroom Foyer				
10:00 AM	2026 QCAS Quantum Computing Applications and Systems Workshop	SSH-SoC: Safety and Security in Heterogeneous Open System-on-Chip Platforms Workshop	Memory-Centric Computing for LLM Inference Workshop	SSH-SoC: Safety and Security in Heterogeneous Open System-on-Chip Platforms Workshop	SUSHI'26: Sustainable Hardware Security Workshop
11:30 AM	Lunch Room: TBD				
1:00 PM	2026 QCAS Quantum Computing Applications and Systems Workshop	SSH-SoC: Safety and Security in Heterogeneous Open System-on-Chip Platforms Workshop		SSH-SoC: Safety and Security in Heterogeneous Open System-on-Chip Platforms Workshop	SUSHI'26: Sustainable Hardware Security Workshop
14:30	Coffee Break | Exhibits Room: Ballroom Foyer				
14:45					
15:00	2026 QCAS Quantum Computing Applications and Systems Workshop	SSH-SoC: Safety and Security in Heterogeneous Open System-on-Chip Platforms Workshop		SSH-SoC: Safety and Security in Heterogeneous Open System-on-Chip Platforms Workshop	SUSHI'26: Sustainable Hardware Security Workshop
16:30					
17:00					
18:00					

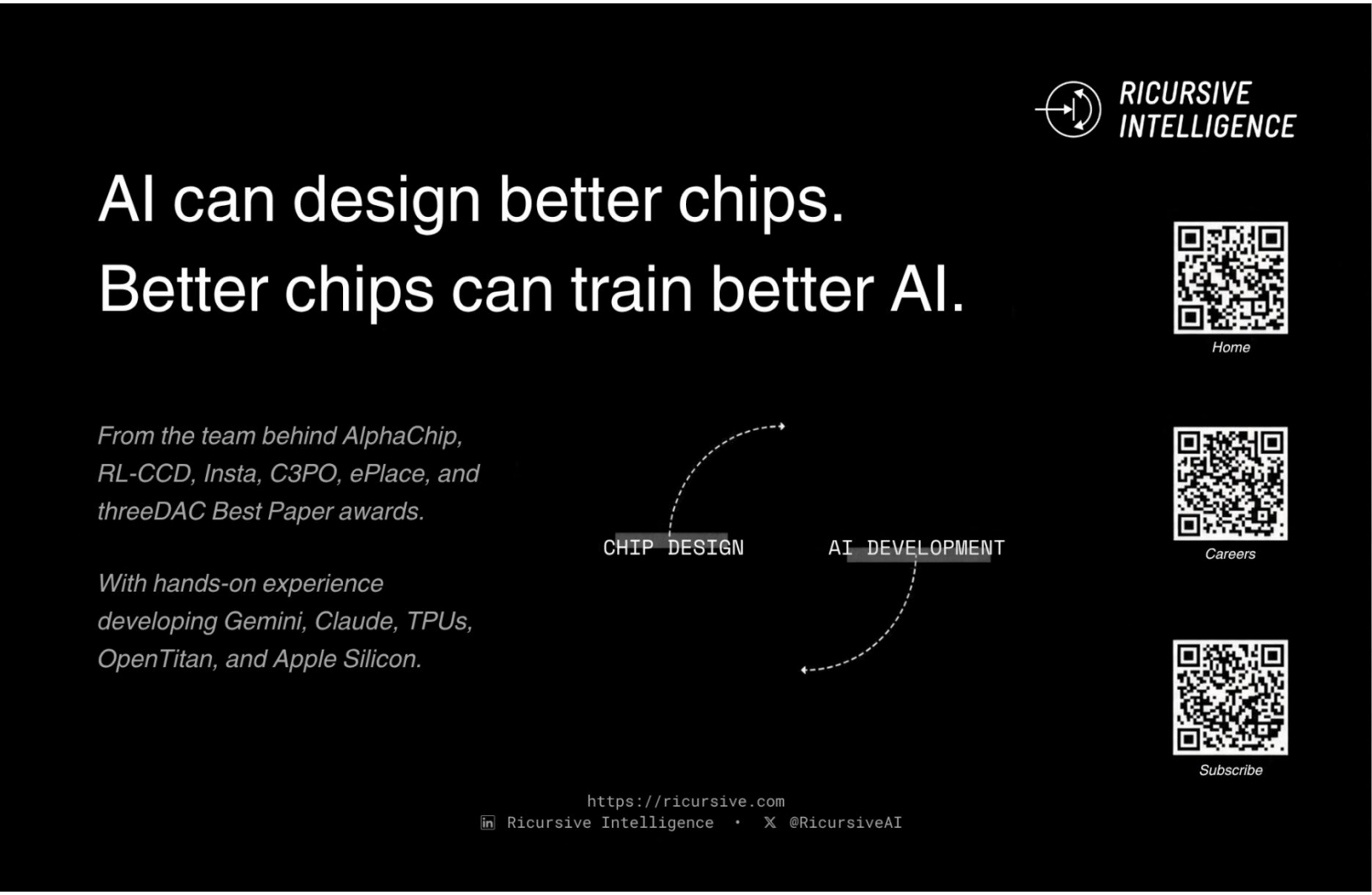
Technical Program – Nov 8

7:30 – 8:00
Registration
Room: Ballroom Foyer

8:00 – 9:00
CADathlon Breakfast
Room: Willow Glen I/II

9:00 – 17:00
CADathlon
Room: Willow Glen I/II

12:00 – 13:00
CADathlon Lunch
Room: Willow Glen I/II



Technical Program – Nov 9

7:30 – 8:15
Registration
Room: Ballroom Foyer

8:15 – 9:00
Opening/Awards Ceremony
Room: San Jose Ballroom

9:00 – 10:00
Keynote: AI-Driven Chip Design and the Future of Computing
Ralph Wittig, AMD
Room: San Jose Ballroom

10:00 – 10:30
Coffee Break | Exhibits
Room: Ballroom Foyer

10:30 – 15:00
Tutorial: Machine Learning for Quantum Circuit Synthesis, Mapping, and Architecture Search: Toward Quantum CAD
Room: Willow Glen I/II

This tutorial presents a design-automation-oriented introduction to machine learning methods for quantum circuit synthesis, mapping, and architecture search. Rather than treating quantum machine learning (QML) as a standalone application area, the tutorial frames QML, quantum reinforcement learning (QRL), and search-based optimization as algorithmic tools for quantum CAD under realistic hardware constraints, including native gate sets, limited qubit connectivity, circuit depth, noise, and resource overhead. Participants will learn how parameterized quantum circuits define structured design spaces, how reinforcement learning, evolutionary optimization, and differentiable architecture search can be used for design-space exploration, and how these methods interact with compilation, mapping, and benchmarking workflows. Through hands-on examples using open-source quantum simulators such as Qiskit and PennyLane in Google Colab, attendees will explore hardware-aware ansatz discovery and circuit optimization using practical design metrics such as fidelity, two-qubit gate count, depth, and robustness. The tutorial concludes with an outlook on scalable Quantum CAD, including emerging self-programming and automated quantum design frameworks.

10:30 - 12:00
Beyond Wirelength: Differentiable Placement for Timing, Routability, and PPA Closure
Room: San Jose Ballroom
Session Chair(s): Yao-Wen Chang, National Taiwan University; Hongxin Kong, Synopsys

Wirelength alone no longer predicts silicon quality. This session showcases the next generation of analytical and differentiable placement techniques that optimize what truly matters—post-route PPA, timing, routability, power domains, and pin assignment—while folding legalization, sizing, and buffering into GPU-accelerated gradient flows that close timing on real designs.

10:30
1192: PPAPlace: Differentiable Cross-Stage Objectives for Chip Placement Optimization
Ruogu Chen, Jie Han
University of Alberta

10:45
1094: MAGIA: Multi-domain Analytical Gradient-driven Placement with Island and Pin Assignment
Luis Peña Treviño{1}, Fady Abouzeid{2}, Lirida Naviner{1}, Philippe Roche{2}
{1}Télécom Paris; {2}STMicroelectronics

11:00
359: Congestion Source-Aware Global Placement for Routability Optimization
Jai-Ming Lin, Yu-Han Su, Jyun-Kai Huang, Jyun-Kai Huang
Department of Electrical Engineering, National Cheng Kung University

11:15
1133: A Probabilistic Differentiable Framework for Simultaneous Post-Placement Buffering, Sizing, and Cell Movement
Yu-Hsiang Huang{1}, Kuan-Yu Fang Chiang{1}, Yu-Chia Kuo{1}, Yun-Han Liang{2}, Iris Hui-Ru Jiang{1}
{1}National Taiwan University; {2}National Taiwan University of Science and Technology

11:30
555: GPU-Accelerated Multi-Objective Exploration for Timing- and Wirelength-aware Placement Legalization
Chen-Han Lu{1}, Wen-Hao Liu{2}, Ting-Chi Wang{1}
{1}National Tsing Hua University; {2}Nvidia

11:45
1292: Adaptive Rectangular Aggregation Methodology via Anisotropic Cohesion for Structural-Prior Placement
Mengen Chen{1}, Xu He{1}, Yao Wang{2}
{1}Hunan University; {2}Phytium Technology

10:30 - 12:00
Chiplets at Scale for GenAI and LLMs: Design, Reliability, and Security Challenges Ahead
Room: Blossom Hill I/II
Session Chair(s): Farshad Firouzi, Arizona State University

Chiplet-based architectures are rapidly transforming heterogeneous computing platforms for GenAI and LLMs, enabling modular scalability, technology heterogeneity, and high-performance integration. However, this shift introduces fundamental challenges across three critical dimensions: design, reliability, and security. Design challenges arise from the need for seamless interoperability, standardized interfaces, scalable design space exploration, and cross-layer hardware–software co-design across heterogeneous chiplet platforms. Reliability concerns encompass thermal management, power delivery, and advanced packaging techniques such as 2.5D/3D integration, all of which directly impact performance, yield, and long-term system robustness. At the same time, security becomes increasingly complex in disaggregated, multi-vendor environments, requiring new approaches to ensure trust, protect data movement, and mitigate emerging attack surfaces across chiplet boundaries. This special session brings together recent advances addressing these interconnected challenges, highlighting innovations in cross-layer design methodologies, reliability-aware system optimization, and secure architectures for heterogeneous AI systems. By unifying perspectives from academia and industry, the session aims to define a forward-looking roadmap for building scalable, reliable, and secure chiplet-based platforms tailored to the demands of GenAI and LLM workloads, spanning cloud, edge, and emerging embodied AI systems, and addressing timely challenges and opportunities central to the International Conference on Computer-Aided Design community.

10:30
1976: Invited Paper: ShatterQuant: Breaking Uniform Precision with Block-Wise Mixed-Precision on a Systolic Transformer Hardware Accelerator
Mikolaj Walczak{1}, Edward Humes{1}, Chao Fang{2}, Marian Verhelst{2}, Tinoosh Mohsenin{1}
{1}Johns Hopkins Whiting School of Engineering; {2}KU Leuven

10:40
1978: Invited Paper: High Performance Silicon Photonic Network-on-Interposers for Dense and Mixture-of-Experts LLMs
Gaurav Narang{1}, Krishnendu Chakrabarty{2}, Farshad Firouzi{3}, Krishnendu Chakrabarty
{1}University of South Florida; {2}Arizona State University; {3}Johns Hopkins University

11:00
1975: Invited Paper: NoWLeak: Side-Channel Attacks on Network-on-Wafer
Dao Xian Ding{1}, Sneha Swaroopa{2}, Farshad Firouzi{3}, Boris Vaisband{4}, Nader Sehatbakhsh{1}
{1}UCLA; {2}University of California, Los Angeles; {3}Johns Hopkins University; {4}University of California, Irvine

11:20
Invited Talk: Efficient HW and Systems for AI Inference at the Edge
Evgeni Gousev
Qualcomm Technologies, Inc.

10:30 - 12:00
Crafting Circuits: From Technology to Optimization
Room: Guadalupe
Session Chair(s): Nan Wu, George Washington University; Peipei Zhou, Brown University

This session highlights foundational advances in reconfigurable computing, spanning FPGA technologies, design automation, compiler frameworks, hardware generation, and system optimization. The accepted papers explore innovations from technology-aware FPGA fabrics and intelligent mapping to CGRA compilation and programming frameworks, enabling more efficient and capable reconfigurable systems.

10:30
1476: GLO-Map: A Globally-Aware Reinforcement Learning Framework for FPGA Technology Mapping
Huyuan Ren{1}, Bingcai Sui{1}, Qinglin Wang{1}, Yaohua Wang{1}, Jie Liu{2}
{1}National University of Defense Technology; {2}National University Of Defense Technology

10:45
1170: Flexible Non-linear Function Hardware Generator with Error-constrained Optimization
Jingmin Zhao{1}, Yizhao Gao{1}, Hayden So{1}, Jingmin Zhao{2}
{1}University of Hong Kong; {2}The University of Hong Kong

11:00
762: ViViD: A Versatile CGRA-Based Framework for Visual Rendering Acceleration on VR Devices
Yanzhou Tang{1}, Tianhua Xia{1}, Cheng Tan{2}, Sai Qian Zhang{1}
{1}New York University; {2}Google/ASU

11:15
146: NUKE: An Automatic Pattern Fusion Framework for CGRAs using E-graphs
Jiajun Qin{1}, Yuhao Ji{2}, Zhaotan Lin{3}, Cheng Tan{4}, Xinyun Zhang{5}, Bei Yu{1}
{1}The Chinese University of Hong Kong; {2}CUHK; {3}The Chinese University of Hongkong; {4}Arizona State University; {5}ShanghaiTech University

11:30
397: SpatialMR: A Data-Parallel Framework on FPGAs
Mingzhe Zhang{1}, Yanzheng Cai{1}, Zhenbo Sun{1}, Shengqi Chen{1}, Yunyi Chen{1}, Bowen Yu{2}, Wenguang Chen{1}
{1}Tsinghua University; {2}Beijing Houdu Technology Co., Ltd.

11:45
752: SHiFT: Enabling Switch Heterogeneity in Technology-aware FPGA Fabric Optimization
Navya Agarwal{1}, Moritz Brunion{2}, Dawit Abdi{2}, Chang Gao{1}, Siva Satyendra Sahoo{2}
{1}TU Delft; {2}IMEC

10:30 - 12:00
From Topology to Tapeout: AI-Driven Analog Circuit Sizing, Compact Modeling, and IP Retrieval
Room: San Carlos
Session Chair(s): Sachin Sachdeva, UC Riverside, USA

Designing analog circuits demands navigating vast multi-objective spaces with scarce data. This session showcases breakthroughs in meta-RL for cross-topology sizing, mixed-variable Bayesian optimization, LLM-guided symbolic regression for compact device models, physics-embedded variability prediction, and cross-modal analog IP retrieval. These papers make a step towards redefining AI-driven analog circuit design automation.

10:30
1651: MetaSizing: Meta-Reinforcement Learning for Generalizable Analog Circuit Sizing
Shikai Wang{1}, Jiyuan Duan{2}, Houbo He{2}, Gerald Topalli{2}, Shaoyi Huang{3}, Taiyun Chi{2}, Weidong Cao{4}
{1}George Washington University; {2}Rice University; {3}Stevens Institute of Technology; {4}The George Washington University

10:45
258: Enabling Efficient Many Objective Analog Circuit Synthesis via Learning-Augmented Hypervolume Approximation
Aidong Zhao, Xihan Yang, Hanwen Guo, Zhaori Bi
Fudan University

11:00
692: Hierarchical Adaptive Bayesian Optimization for Mixed-Variable Analog Circuit Design
Jindong Tu{1}, Junchen Liu{1}, Liu Wang{2}, Yapeng Li{1}, Benyang Qi{1}, Tongkai Wu{1}, Zushuai Xie{2}, Guoqing Li{3}, Tinghuan Chen{1}
{1}The Chinese University of Hong Kong, Shenzhen; {2}Nanjing University of Posts and Telecommunications; {3}Shandong Yunhai Guochuang Cloud Computing Equipment Industry Innovation Co., Ltd

11:15
754: PESA: Physics-Embedded Neural Network for Data-Efficient Variability Prediction in DRAM Sense Amplifiers
Dongyeong Kim, Hyerin Lee, Ryun Kang, Suyeon Kim, Jewon Park, Sinwook Kim, Geon Kim, Myoungjin Lee
Chonnam National University

11:30
1353: CMLSR: Automated Compact Modeling through (Knowledge-enhanced) LLM-Guided Symbolic Regression
Tianye Yu, Gongteng Xiao, Haobo Huang, Xuan Zeng, Shisheng Xiong, Ye Lu
Fudan University

11:45
1774: AnalogRetriever: Learning Cross-Modal Representations for Analog Circuit Retrieval
Yihan Wang{1}, Lei Li{2}, Yao Lai{3}, Jing Wang{4}, Yan Lu{1}
{1}Tsinghua University; {2}The University of Hong Kong; {3}University of Cambridge; {4}Nanjing University of Posts and Telecommunications

10:30 - 12:00
Routing and Co-Optimization in Physical Design
Room: Blossom Hill III
Session Chair(s): Jiang Hu, Texas A&M University

This session explores routing and automated flow optimization in physical design, as well as FPGAs and PCBs. The session highlights innovative routing strategies, featuring multi-FPGA co-optimization, pattern routing, routability prediction, and congestion-aware double-sided signal routing.

10:30
317: Routing and Reconfigurable Topology Co-optimization for Multi-FPGA Systems
Hui Xu{1}, Junhong Qiao{1}, Jiaming Chang{1}, Peng Zou{2}, Jianli Chen{1}
{1}College of Integrated Circuits & Micro-Nano Electronics, Fudan University, Shanghai, China; {2}Shanghai LEDA Technology Co., Ltd., Shanghai, China

10:45
515: StateTune: Transforming LLM-Assisted EDA Flow Tuning into a Stateful, Closed-Loop Process
Kunlong Li{1}, Shangshang Yao, Su Zheng{2}, Lingli Wang{1}
{1}Fudan University; {2}Chinese University of Hong Kong

11:00
602: Routing Dense Layouts with History-Aware Offline Reinforcement Learning using LSTM
Afsara Khan, Austin Rovinski
New York University

11:15
809: AxialNet: Orthogonally Decoupled Layout Learning with Manhattan Bias for Routability Prediction
Zhiyan Liu{1}, Jiawei Liu{2}, Jianwang Zhai{1}, Chuan Shi{1}
{1}Beijing University of Posts and Telecommunications; {2}The Chinese University of Hong Kong

11:30
1147: From Discrete Choices to Continuous Bend Positions: A Differentiable Framework for Global Pattern Routing
Lin Chen, Hu Ding
University of Science and Technology of China

11:45
1464: Timing-Driven Bridging Cell Planning via Congestion-Aware Region Selection for Double-Sided Signal Routing
Bo-Ying Huang, Yi-En Wu, Yao-Wen Chang
National Taiwan University

10:30 - 12:00
The Physical Frontier: Analog Compute-in-Memory and Mixed-Signal Architectures
Room: Willow Glen III
Session Chair(s): Hyun Kim, Seoul National University of Science and Technology; Andrea Galimberti, Politecnico di MIlano

Analog computing offers unparalleled energy efficiency but demands robust noise tolerance. This session unveils breakthroughs in SRAM-RRAM heterogeneous fabrics, single-cut chopping MACs, and write-and-verify programming. Discover how analog CIM is overcoming variability to efficiently power integer-only diffusion and recommendation systems.

10:30
206: HIMAS: A Heterogeneous SRAM–RRAM In-Memory Accelerator for Scale-Adaptive PointKAN Inference
Xuliang Yu{1}, Xinao Zhao{2}, Zewen Cao{2}, Yi Zhou{3}, Qi Wang{2}, Liang Zhao{1}
{1}Zhejiang University, information science and electronic engineering college; {2}Institute of Microelectronics of the Chinese Academy of Sciences, Beijing, China; University of Chinese Academy of Sciences, Beijing, China; {3}AI Thrust, Hong Kong University of Science and Technology (Guangzhou)

10:45
261: HARP: Hadamard-Domain Write-and-Verify for Noise-Robust RRAM Programming
Ilhuan Choi, Jiwon Yoo, Yoona Lee, Yewon Jeong, Jason Jaesung Lee, Woo-Seok Choi
Seoul National University
11:00
528: Algorithm-Hardware Co-Design of a Robust Large-Scale Analog MAC for Neural Network Acceleration using Single-Cut Chopping
Maysara Hamada{1}, Youngrae Kim{1}, Sreetama Sakar{1}, Chee-An Yu{1}, Jake Hyun{2}, Huixuan Yin{1}, Hsiang-Chun Cheng{1}, Yifan He{2}, Jian Meng{2}, Jae-sun Seo{3}, C.-C. Jay Kuo{1}, Peter A. Beerel{1}, Chen Shuo-Wei{1}, Jian Meng{2}
{1}University of Southern California; {2}Cornell University; {3}Cornell Tech

11:15
663: IMC-Diff: An Activation-Centric IMC Architecture for True Integer-Only Diffusion Inference
Donghyeon Yi{1}, Donghwan Kwon{1}, Yoontae Jung{2}, Ik-Joon Chang{2}, Minkyu Je{1}
{1}Korea Advanced Institute of Science and Technology, KAIST; {2}Kyung Hee University

11:30
1201: D2C-CIM: Dual-Domain Cascaded Analog Compute-in-Memory Processor Featuring Eliminated Intermediate ADCs and Memory Access for Depthwise Separable Convolution
Liang Zhang{1}, Yang Liu{2}, Jing Kou{2}, Junda Zhao{2}, Sizhe Niu{2}, He Zhang{2}, Wang Kang{1}, Junzhan Liu{1}
{1}Beihang Univeristy; {2}Beihang University

11:45
1901: A Reconfigurable RRAM CIM Accelerator with Inner-Outer-Mixed Product and BCH-Compatible Correction for Energy-Throughput-Accuracy Co-Optimization in Recommendation Systems
Hongyang Hu{1}, Zengji Li{1}, Bohan Wang{1}, Jinshan Yue{2}, Jing Liu{3}, Xiaoxin Xu{4}, Dashan Shang{1}, Chunmeng Dou{1}
{1}Institute of Microelectronics, Chinese Academy of Sciences; {2}Institute of Microelectronics of the Chinese Academy of Sciences; {3}Beijing Memsilicon Technology Co., Ltd.; {4}State Key Laboratory of Fabrication Technologies for Integrated Circuits

12:00 – 13:30
Lunch
Room: San Jose Ballroom

13:30 - 15:00
AI-Assisted Chip Design Contest: 2026 ICCAD CAD Contest
Room: Blossom Hill I/II
Session Chair(s): Shao-Yun Fang, National Taiwan University of Science and Technology

The CAD Contest at ICCAD is a challenging, multi-month, research and development competition, focusing on advanced, real-world problems in the field of Electronic Design Automation (EDA). It is open to multi-person teams world-wide. Each year the organizing committee announce three challenging problems in different topic areas provided by industrial companies. Contestants can participate in one or more problems. The prizes will be awarded at an ICCAD special session dedicated to this contest.

Since its inaugural year of 2012, the CAD Contest at ICCAD has been attracting more than a hundred teams per year, fostering productive industry-academia collaborations, and leading to hundreds of publications in top-tier conferences and journals. The contest undoubtedly boosts EDA research and keeps enhancing its impact.

1998: Invited Paper: Overview of 2026 CAD Contest at ICCAD
Shao-Yun Fang{1}, Yi-Yu Liu{1}, Chung-Kuan Cheng{2}, Tsun-Ming Tseng{3}
{1}National Taiwan University of Science and Technology; {2}Computer Science and Engineering Department, UC San Diego; {3}Technical University of Munich

1999: Invited Paper: 2026 ICCAD CAD Contest Problem A: LLM-Assisted Netlist Exploration and Transformation
Chung-Han Chou{1}, Hung-Chun Chiu{1}, Chih-Jen (Jacky) Hsu{1}, Kai-Chiang Wu{2}, Zhuo Li{1}, Shao-Yun Fang{3}
{1}Cadence Design Systems, Inc.; {2}National Yang Ming Chio Tung University; {3}National Taiwan University of Science and Technology

1994: Invited Paper: 2026 ICCAD CAD Contest Problem B: Regression Failure Bucketing
Gung-Yu Pan, Yen-Ju Chen
Synopsys Taiwan

1997: Invited Paper: 2026 ICCAD CAD Contest Problem C: The FloorSet Challenge: Data-Driven SoC Floorplanning
Uday Mallappa{1}, Sachin Bhat{1}, Pei Chun Ch'ng{1}, Yuan Lu{2}, Yusu Wang{2}, Somdeb Majumdar{3}, Mariano Phielipp{4}, Uday Mallappa{1}
{1}Intel Corporation; {2}UC San Diego; {3}DeepL; {4}All3

2007: Invited Paper: 2026 Extensions of IEEE CEDA DATC's Robust Design Flow and Other Activities
Alec Aversa{1}, David Chinnery{2}, Amur Ghose{3}, Aryan Jain{4}, Andrew Johnson{2}, Yugesh Kothari{4}, Jakang Lee{3}, Yiting Liu{3}, Davit Markarian{3}, Sahrima Jannat Oishwee{2}, Fairooz Saiyara{2}, Ioannis Savidis{1}, Pratik Shrestha{1}, Paul Sirois{2}, Zhiang Wang{5}
{1}Drexel University; {2}Siemens EDA; {3}UC San Diego; {4}Ergodex AI, Inc.; {5}Fudan University

13:30 - 15:00
Clock Network Design and Automation
Room: Blossom Hill III
Session Chair(s): Wenchao Qian, AMD

This session targets CTS for traditional 2D IC as well as 3D IC topologies. Papers include techniques on evolutionary structural optimization for CTS, timing-aware Steiner tree generation for buffer insertion, and comprehensive BS-CDN synthesis utilizing GNN clustering paired with MILP-based optimization.

13:30
1737: TAT: Timing-Aware Steiner Tree Generation for Buffer Insertion
Xiang Li, Zuodong Zhang, Yibo Lin
Peking University
13:45
1298: EvoCTS: Differentiable Clock Tree Synthesis Based on Evolutionary Structural Optimization
Han Zhang{1}, Jiaqi Lyu{2}, Junming Jiao{2}, Xilin Qian{2}, Peng Cao{2}
{1}Southeast university; {2}Southeast University

14:00
689: Clock Tree Synthesis with Cross-Tier Skew and Wirelength Co-Optimization for Face-to-Face Hybrid Bonding 3D ICs
Yuxi Lin{1}, Yibo Lin{2}, Zepeng Li{1}, Weijie Fang{1}, Guangyong Chen{1}, Genggeng Liu{1}
{1}Fuzhou University; {2}Peking University

14:15
1678: CLK-3D: RC-Calibrated Differentiable Clock-Tree Optimization for 3D ICs
Yen-Hsiang Huang{1}, Sung Kyu Lim{2}
{1}Georgia Institute of Technology; {2}University of Southern California

14:30
560: Comprehensive BS-CDN Synthesis Framework with GNN Clustering and MILP-Based Optimization
Ikkyum Kim{1}, Junghyun Yoon{2}, Gyumin Kim{1}, Jooyeon Jeong{3}, Heechun Park{1}
{1}UNIST; {2}Ulsan National Institute of Science and Technology (UNIST); {3}UCLA

13:30 - 15:00
Intelligent Design Optimization
Room: San Carlos
Session Chair(s): Zanhe Qi, Ritsumeikan University; Shigeru Yamashita, Ritsumeikan University

This session highlights intelligent optimization techniques spanning multiple stages of the EDA flow, from quantum circuit optimization and logic synthesis to layout restructuring, library tuning, and high-level synthesis. Leveraging reinforcement learning, graph neural networks, large language models, and knowledge-driven methods, these works showcase the growing impact of AI on EDA.

13:30
338: Quantum Oracle Circuit Optimization via T-Count-Aware Subcircuit Synthesis and Ancilla–T Count Balancing
Hsiang-Chun Yang{1}, Chen-Ying Tung{2}, Chung-Yang (Ric) Huang{1}
{1}Graduate Institute of Electronics Engineering, National Taiwan University; {2}Department of Electrical Engineering, National Taiwan University

13:45
45: LevelSyn: Physical-Aware Logic Synthesis via Level-Asynchronous Graph Neural Networks
Jingyi Zhou{1}, Zhengyuan Shi{2}, Ziyang Zheng{2}, Qiang Xu{2}, Zhou Jingyi
{1}Tsinghua University; {2}The Chinese University of Hong Kong

14:00
1436: LibPilot: Knowledge-Constrained Dual-Agent Reasoning for Standard Cell Library Tuning
Yanfang Liu{1}, Zijin Cheng{2}, Mingjun Wang{1}, Rongliang Fu{1}, Chao Wang{2}, Bei Yu{1}, Rongliang Fu{1}
{1}The Chinese University of Hong Kong; {2}Southeast University

14:15
268: E-Layout: Layout-Stage Restructuring via RL-Optimized Equality-Saturated Graph Retrieval-Augmented Generation
Wenji Fang{1}, Rongjian Liang{2}, Yi-Chen Lu{3}, Cunxi Yu{4}, Chen Chen{4}, Shang Liu{1}, Chenhui Deng{2}, Wenkai Li{1}, Haoxing Ren{5}, Zhiyao Xie{1}, Chenhui Deng{6}
{1}Hong Kong University of Science and Technology; {2}Nvidia; {3}Ricursive Intelligence; {4}University of Maryland, College Park; {5}Agentrys; {6}NVIDIA

14:30
431: HEGNN-RCSAC: Scaling HLS Design Space Exploration with Hyperbolic Embeddings and Reward-Centered Reinforcement Learning
Langchang Zhang, Lei Xu, Shanshan Wang, Chenglong Xiao
Shantou University

14:45
662: LiteCut: Lightweight Context-Enhanced Cut Delay Prediction for AIG Rewriting via Knowledge Distillation
Lingfeng Zhou{1}, Yilong Zhou{1}, Xiaoyan Yang{2}, Yue Wu{3}
{1}Hangzhou Dianzi University; {2}Hangzhou Dianzi University, Shaoxing Integrated Circuit Institute of Hangzhou Dianzi University; {3}HangZhou DianZi University, Shaoxing Integrated Circuit Institute of Hangzhou Dianzi University

13:30 - 15:00
Leakage in the Machine
Room: Guadalupe
Session Chair(s): Jitendra Bhandari, Synopsys; Saion Roy, Northeastern University College of Engineering

Hardware isolation is failing across modern compute boundaries. From penetrating GPU multi-tenancy and confidential VM extensions to weaponizing network-on-chip contention, stealthy microarchitectural leakage threatens critical software secrets. This session delivers urgent answers: uncovering novel cross-domain side channels while deploying pre-silicon information-flow tracking and router-level ML detectors to stop them.

13:30
105: RekSai: Covert Channel Attacks via GPU Last-Level Caches
Lingfeng Pei{1}, Changxi Liu{2}, Xiaoyu Cheng{3}, Trevor E. Carlson{1}, Lingfeng Pei{1}
{1}National University of Singapore; {2}The Hong Kong University of Science and Technology (Guangzhou); {3}Southeast University

13:45
312: From Noisy to Practical: Enabling Robust NoC Side and Covert Channels on AMD Processors
Junho jo, Wonjun song
Kangwon National University

14:00
480: LazyLeak: Software Covert Channels based on Lazy Memory Allocation Mechanisms
Congcong Chen, Jiliang Zhang
Nanjing University of Posts and Telecommunications

14:15
1404: AMXposed: Extracting Transformer Architectures and IP Exfiltration from Confidential VMs via Intel AMX Side-Channel
Upasana Mandal{1}, Yagnik Fadadu{1}, Nimish Mishra{1}, Shubhi Shukla{1}, Sarani Bhattacharya{1}, Debdeep Mukhopadhyay{2}
{1}Indian Institute of Technology, Kharagpur; {2}Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur

14:30
390: SPARC: Automated Root-Cause Analysis of Pre-Silicon Power Side-Channel Leakage in the Processor Design Flow
Andrija Neskovic{1}, Christian Ewert{1}, Mladen Berekovic{2}, Saleh Mulhem{2}, Andrija Neskovic
{1}Universität zu Lübeck; {2}University of Luebeck

14:45
926: Decanter: Detecting Cache Side-Channel Attacks via ML-inspired NoC Traffic Analyzer
Pengyang Li{1}, Yuxiang Wang{1}, Pengfei Gou{2}, Deshan Zhang{3}, Zhenhua Guo{3}, Fu Xiao{1}, Xiaobai Chen{1}, Jieming Yin{1}
{1}Nanjing University of Posts and Telecommunications; {2}RISCV Computing Co.; {3}IEIT SYSTEMS Co., Ltd.

13:30 - 15:00
Learning to Attack, Learning to Defend
Room: Almaden
Session Chair(s): Muhammed Jukaku, Synopsys

AI models are rapidly becoming both high-value hardware targets and frontline defensive sentinels. Discover how subtle bit-flip faults silently manipulate LLM costs and behaviors without gradient access, while machine learning fights back—demolishing resilient NVM-based PUFs, detecting adversarial inputs on GPUs, and scaling RTL Trojan localization to millions of gates.

13:30
254: BitHydra: Towards Bit-flip Inference Cost Attack against Large Language Model
Xiaobei Yan, Yiming Li, Tianwei Zhang
Nanyang Technological University

13:45
509: Invisible Hands: Gray-Box Bit Flip Attack for Steering LLMs Without Knowledge of Gradients, Data, and Weights
Abeer Matar A Almalky{1}, Ziyan Wang{2}, Mohaiminul Al Nahian{3}, Li Yang{2}, Adnan Siraj Rakin{1}
{1}Binghamton University (SUNY); {2}University of North Carolina at Charlotte; {3}Binghamton University

14:00
1264: GPUrify: Exploiting GPU Kernel-level Metric Signatures in Vision Transformers for Black-Box Detection of Adversarial Examples
Shubhi Shukla{1}, Manaar Alam{2}, Michail Maniatakos{2}, Pabitra Mitra{1}, Debdeep Mukhopadhyay{3}
{1}IIT Kharagpur; {2}New York University Abu Dhabi; {3}Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur

14:15
731: BREACH: Breaking Resilient NVM-based PUFs via Tailored ML Attacks
Mohamed Alsharkawy{1}, Hassan Nassar{2}, Kuan-Hsun Chen{3}, Jörg Henkel{2}
{1}CES - Chair for Embedded Systems Department of Computer Science KIT - Karlsruhe Institute of Technology; {2}Karlsruhe Institute of Technology; {3}University of Twente

14:30
747: ADVERSARIAL: And-Inverter Graph-Assisted Hardware Trojan Detection At Scale
Yaroslav Popryho{1}, Debjit Pal{2}, Inna Partin-Vaisband{3}
{1}University of Illinois Chicago; {2}University of Illinois at Chicago; {3}University of Illinois, Chicago

14:45
1604: TrojanLoC: LLM-based Framework for RTL Trojan Localization and Classification
Weihua Xiao{1}, Zeng Wang{1}, Minghao Shao{1}, Raghu Vamshi Hemadri{1}, Chang Meng{2}, Johann Knechtel{3}, Muhammad Shafique{3}, Ozgur Sinanoglu{4}, Siddharth Garg{1}, Ramesh Karri{1}
{1}New York University; {2}Eindhoven University of Technology; {3}New York University Abu Dhabi; {4}New York University, Abu Dhabi

---

13:30 - 15:00
Scalable Optimization for Floorplanning, Partitioning, and Chiplet-Based Design
Room: San Jose Ballroom
Session Chair(s): Zhengqi Gao, Massachusetts Institute of Technology; Patrick Madden, SUNY Binghamton CSD

This session presents scalable optimization techniques for modern physical design, spanning analytical floorplanning, legalization, chiplet placement, and hypergraph partitioning. The accepted papers address complex SoC and multi-reticle chiplet constraints through force-directed methods, multilevel partitioning, portfolio optimization, and learning-guided placement, highlighting emerging directions for high-quality, constraint-aware design automation.

13:30
1430: Toward Near-Optimal Floorplanning with Complex Design Constraints for Modern SoCs
Fuxing Huang{1}, Hao Ai{2}, Qiyuan Chen{1}, Hao Wu{1}, Wenxing Zhu{2}, Xinning Liu{1}, Ziran Zhu{1}
{1}Southeast University; {2}Fuzhou University

13:45
553: Analytical Chiplet Placement with Force-Directed Guidance and Chiplet Bridging for Multi-Reticle Interposers
Jiayao Li{1}, Lang Feng{2}, Jinhong Huang{3}, Keren Zhu{3}, Zhongfeng Wang{4}
{1}Nanjing University; {2}Sun Yat-sen University; {3}Fudan University; {4}Nanjing University, Sun Yat-sen University

14:00
559: HyperPart: A Multilevel Hypergraph Partitioner with Dual-Path Variational Embeddings
Magi Chen{1}, Ting-Chi Wang{2}
{1}National Changhua University of Education; {2}National Tsing Hua University

14:15
585: Learning Module-Level Placement Guidance for Timing Optimization
Heng-Yu Chang{1}, Wai-Ting Bai{1}, Ta-Yu Kuan{2}, Tung-Chieh Chen{2}, Wai-Kei Mak{1}, Ting-Chi Wang{1}
{1}National Tsing Hua University; {2}Synopsys Incorporated

14:30
39: Quality-Preserving Legalization for Analytical Floorplanning
Yunqi He, You Li, Ruofan Huang, Hai Zhou
Northwestern University

14:45
1695: FastPart: A Powerful Portfolio-based Hypergraph Partitioner
Amur Ghose{1}, Andrew Kahng{2}
{1}UC San Diego; {2}UCSD CSE and ECE Depts.

15:00 – 15:30
Coffee Break | Exhibits
Room: Ballroom Foyer

15:30 - 17:00
3D Nanofabric: A Pathway to Post-Denard CMOS Scaling
Room: Blossom Hill III
Session Chair(s): Giovanni Ansaloni, EPFL

Chip technology underpins progress across sectors such as automotive, telecommunications, finance, and healthcare. However, continued improvements in performance, efficiency, and functionality are increasingly constrained by the limits of traditional CMOS scaling and TSV-based 3D integration, challenged by scaling, memory, power, sustainability, and cost. Increasing integration density has long been a key enabler of progress in logic design. However, conventional two-dimensional CMOS scaling faces fundamental physical and economic challenges. Expanding integration into the third dimension offers a natural path forward. 3D Nanofabric introduces a disruptive alternative by enabling the simultaneous processing of N identical stacked layers, independent of the number of layers. This unique capability dramatically reduces fabrication cost and hardware footprint while unlocking new possibilities for dense, energy-efficient, and high-performance systems. This new fabrication paradigm has profound implications across the entire semiconductor design stack—from process design kits (PDKs) and physical design methodologies to logic and system architecture. Traditional EDA tools must evolve to address these cross-layer co-design challenges and opportunities. This session will bring together experts across the full design stack, spanning technology, design automation, and system architecture, to discuss the methodologies and innovations required to realize the full potential of 3D Nanofabric and shape the next era of semiconductor design. As this session covers the technology and the entire design stack, from EDA, circuit design, memory, and system architecture, it should be relevant to a large audience at ICCAD from both academia and industry.

1988: Invited Paper: 3D Nanofabric, A Pathway to Continued CMOS Cost Scaling
Christopher Kessler{1}, Maah Paara{2}, Stefano Albini{3}, Enzo Rafinesque{4}, Dawit Abdi{5}, Juergen Boemmels{5}, Ahmed Hemani{6}, MEHDI B TAHOORI{7}, David Novo{8}, Pascal Nouet{4}, Francky Catthoor{9}, Giovanni Ansaloni{3}, Denisa Constantinescu{3}, David Atienza{3}, Julien Ryckaert{5}, Serge Biesemans{5}
{1}Karlsruher Institut for Technology (KIT); {2}KTH Royal Institute of Technology; {3}Embedded Systems Laboratory, EPFL; {4}LIRMM, Univ. of Montpellier, CNRS; {5}imec; {6}KTH - Royal Institute of Technology; {7}KIT, imec; {8}LIRMM, Univ Montpellier, CNRS; {9}Microlab, NTUA

15:30 - 17:00
Accelerating Trustworthy Cryptography
Room: Guadalupe
Session Chair(s): Jitendra Bhandari, NYU

Deploying post-quantum cryptography, FHE, and zero-knowledge proofs requires conquering massive memory bottlenecks, physical faults, and side-channel vulnerabilities. This session showcases breakthrough hardware architectures—including compute-in-memory NTT engines, fault-tolerant modular ALUs, and GAN-driven FHE accelerators—that slash area and latency while establishing robust, silicon-level defenses for next-generation zero-trust systems.

15:30
330: CRSCIM-NTT: An Algorithm-CIM Co-Optimized Architecture for Compact and Security-Enhanced NTT Acceleration
Jiguang Kan, Chenghua Wang, Bei Wang, Yuhao Shu, Jiansheng Chen, Jiatong Tian, Yijun Cui, Weiqiang Liu
Nanjing University of Aeronautics and Astronautics

15:45
353: LEAP_LMSD: A Low-Overhead Side-Channel-Resistant Unified Polynomial Accelerator for HQC
Congwei Chen{1}, Jianxiong Zhang{1}, Haolin Mai{1}, Zijian Lu{1}, Guohong Liu{1}, Beihai Tan{1}, Yun Chen{2}, Xiaoming Xiong{1}, Shuting Cai{1}
{1}Guangdong University of Technology; {2}HKUST(GZ)

16:00
887: CMALU: Compact Fault-Tolerant Modular Arithmetic Logic Unit for Post-Quantum Cryptography
YoungBeom Kim{1}, Malik Imran{2}, Zain Ul Abideen{3}, Ciara Rafferty{2}, Ayesha Khalid{2}, Máire O'Neill{4}, Seog Chung Seo{1}
{1}Kookmin University; {2}Queen's University, Belfast; {3}University of Idaho; {4}Queen's University Belfast

16:15
400: Odin: Rethinking FHE Accelerator Architecture via Computation–Storage Co-Design
Silin Liu, Yinghao Yang, Fuyao Liu, Xiaowei Li, Hang Lu
Institute of Computing Technology, Chinese Academy of Sciences

16:30
437: YUNDING: Fully Homomorphic Encryption Accelerator with Reconfigurable Engines for Privacy-Preserving Inference
Jian Wei, Hao Li, Guohao Zheng, Bo Chen, Xueyang Wang, Yang Cao, Lei Bao
Alibaba Group

16:45
1528: A Resource-Efficient Hardware Accelerator for Large-Size NTT via Algorithm–Architecture Co-Design
Kaixuan Wang{1}, Yifan Yanggong{1}, Xiaoyu Yang{2}, Chenti Baixiao{2}, Lei Wang{1}
{1}Shanghai Jiao Tong University; {2}Chipltech

15:30 - 17:00

Breaking the Memory Wall for LLMs: Computing-in-Memory as the Catalyst for Scalable and Efficient Generative AI

Room: Willow Glen I/II

Session Chair(s): Jian-Jia Chen, RWTH Aachen University

The rapid emergence of large language models (LLMs) has transformed modern information processing, enabling unprecedented capabilities in reasoning, content generation, and human–machine interaction. However, their deployment is fundamentally constrained by the "memory wall" — the excessive cost of data movement between memory and compute units in conventional von Neumann architectures. As LLMs continue to scale in both model size and context length, this bottleneck increasingly dominates system performance, energy consumption, and scalability. Computing-in-memory (CIM) has long been viewed as a promising paradigm to address this challenge by performing computation directly within memory arrays, thereby minimizing data movement. Yet, despite its potential, CIM has traditionally been considered ill-suited for LLMs due to the inherent mismatch between the massive parameter size of LLMs and the limited capacity, precision, and reliability of emerging memory devices. This has led to a prevailing belief that CIM cannot effectively support modern generative AI workloads. This special session challenges this conventional wisdom and presents a new cross-layer perspective on how CIM can not only support but fundamentally reshape LLM system design. The key insight unifying this session is that CIM should not be viewed as a monolithic replacement for digital accelerators, but rather as a specialized substrate for selectively accelerating memory-intensive components of LLM workloads. By rethinking the mapping of LLM execution — offloading retrieval, context storage, and stochastic transformations into CIM fabrics — recent advances demonstrate that the perceived limitations of CIM can be transformed into unique advantages. The session begins with a broad architectural perspective on machine learning in memory, establishing the foundation for memory-centric computing paradigms. It then presents concrete system realizations showing how memristor-based CIM can efficiently support retrieval-augmented generation and long-context modeling in LLMs. Finally, the session highlights a deeper algorithm–hardware co-design insight: intrinsic device-level randomness, traditionally viewed as a drawback, can be directly exploited to implement advanced LLM compression techniques such as the recent breakthrough TurboQuant. Together, these talks provide a coherent and forward-looking vision in which CIM evolves from a niche accelerator into a key enabler for scalable, energy-efficient, and deployable generative AI systems. By bridging the gap between device-level characteristics and system-level design, this session aims to inspire new directions in EDA, architecture, and AI hardware research.

1984: Invited Paper: Efficient LLM Inference with Memristor-Based Computing-in-Memory: From Retrieval-Augmented Generation to 3D-Integrated LLM Acceleration

Mengyun Zhang, Peng Yao, Jiaming Li, Jingtai Wu, Xueqi Li, Yuan He, Yiming Zhou, Tao Liu, Qingtian Zhang, Jianshi Tang, He Qian, Bin Gao, Huaqiang Wu

Tsinghua University

1995: Invited Paper: Interface-Aware KV Cache Quantization for Dense On-Chip NVM in Long-Context LLM Decoding

Jiahao Zheng, Yifan Qin, Xiaobo Sharon Hu, Yiyu Shi

University of Notre Dame

15:30 - 17:00
LLM-Driven and Rule-Based Verification and Test Generation
Room: Almaden
Session Chair(s): Takahide Yoshikawa, Fujitsu; Kamesh Chandrasekar, AMD

This session's papers ask large language models to draft the assertions, abstractions, coverage models, tests, and layouts that verification and sign-off depend on, or skip the model for fixed rules instead. Every output is then checked independently: formal model checking, SMT solving, structural validation, or differential testing.

15:30
1060: CHARGE: Leveraging CWE Hierarchies for Hardware Security SystemVerilog Assertion Generation
Xiao Tan, Cynthia Sturton
University of North Carolina at Chapel Hill

15:45
1703: A Rule-based Automated Assertion Generation Approach for Detecting Hardware Vulnerabilities
Ruochen Dai{1}, Dakota Cappel{1}, Chen Chen{2}, Mohamadreza Rostami{3}, Yoel Rodrigues{3}, Ahmad-Reza Sadeghi{3}, Jeyavijayan Rajendran{2}, Tuba Yavuz{1}, Ruochen Dai{1}
{1}University of Florida; {2}Texas A&M University; {3}Technical University of Darmstadt

16:00
543: NeuroAbs: A Neuro-Symbolic RTL Abstraction Framework for Property Checking Acceleration
Zhiyuan Yan{1}, Xiaofeng Zhou{2}, Ziyue Zheng{1}, Ziyi Yang{1}, Wenbin Che{1}, Wei Zhang{3}, Yangdi Lyu{4}, Hongce Zhang{4}
{1}The Hong Kong University of Science and Technology (Guangzhou); {2}The Hong Kong University of Science and Technology; {3}Hong Kong University of Science and Technology; {4}Hong Kong University of Science and Technology (Guangzhou)

16:15
660: ChatFCM: Comprehension-Synthesis Decoupling for LLM-Based Functional Coverage Model Generation
Yuan Pu{1}, Yudong He{2}, Zhenghao Chen{3}, Mingjun Wang{1}, Bingkun Yao{4}, Zhaotan Lin{5}, Qin Chen{6}, Bing Li{6}, Song Chen{2}, Zhuolun He{1}, Wenjian Yu{7}, Yanfeng Li{6}, Bei Yu{1}, Yanfeng Li{6}
{1}The Chinese University of Hong Kong; {2}University of Science and Technology of China; {3}Beijing University of Posts and Telecommunications; {4}Chinese University of Hong Kong; {5}CUHK; {6}ChangXin Memory Technologies; {7}Tsinghua University

16:30
221: GI-RTL-Fuzz: Grammar-Informed RTL Generation with LLM for End-to-End Differential Simulator Fuzzing
Ziyue ZHENG, Zhi QU, Yangdi Lyu, Ziyue ZHENG
Hong Kong University of Science and Technology (Guangzhou)

16:45
644: VioGen: Targeted Single-Type DRC Violation Case Generator with Agentic Orchestration for Debugging and Tool Benchmarking
Xinran Li, Youran Wu, Junqi Wang, Weifan Li, Shunjie Chang, Haodong Lu, Jianli Chen, Jun Yu, Kun Wang
Fudan University

15:30 - 17:00
Pushing the Precision Floor of LLM Quantization
Room: Willow Glen III
Session Chair(s): Ruokai Yin, Meta; Seungkyu Choi, Yonsei University

Escaping the memory bandwidth bottleneck requires extreme quantization. This session explores sub-8-bit innovations, dynamic outlier flattening, and bit-plane sharing. Discover how next-gen hardware handles asymmetric quantization and low-bit formats, empowering massive language models to run efficiently without sacrificing critical reasoning accuracy.

15:30
208: High-Coverage Outlier Representation in Low-Bit MX Formats for Efficient LLM Inference
Youngmin Cho{1}, Yoontae Lee{2}, Rojin Park{3}, Sunjae Lee{4}, Jonghyeok Park{5}, Jimin Lee{6}, Young H. Oh{1}, Jeongwoo Park{7}
{1}Ajou University; {2}University College London; {3}KAIST; {4}Sungkyunkwan University; {5}Seoul National University; {6}Carnegie Mellon University; {7}Sungkyunkwan university

15:45
262: FlatLLM: An Efficient Low-Bit LLM Accelerator via Cluster-aware Lightweight Outlier Flattening
Yifan Song{1}, Changxu Liu{1}, Baoyu Fan{2}, Fan Yang{1}, yifan song{1}
{1}Fudan University; {2}IEIT Systems Co Ltd

16:00
282: BitWeave: A Bit-Plane Sharing Accelerator for Efficient Mixed-Precision Parallel Scaling in Reasoning Models
Yanmin Li{1}, Lin Shu{2}, Long Xiao{1}, Liuqing Chen{3}, Xiru Lei{1}, Jie Hao{2}
{1}Institute of Automation, Chinese Academy of Sciences; School of Artificial Intelligence, University of Chinese Academy of Sciences; {2}Institute of Automation, Chinese Academy of Sciences; {3}Institute of Automation, Chinese Academy of Sciences; School of Industry-education Integration, University of Chinese Academy of Sciences

16:15
799: ADD-LLM: An Add-Only LLM Accelerator via Ultra Low-bit Asymmetric Power-of-Two Quantization
Nam Joon Kim, Young Chan Kim, Wonjun Hwang, Hyun Kim
Seoul National University of Science and Technology

16:30
806: OmniAcc: A Heterogeneous Memory-Centric LLM Accelerator with Reconfigurable Fixed-Point Nonlinear Operations and Layer-Agnostic Compute Fabric
Yimin Wang, Zixuan Wang, Yue Jiet Chong, Xuanyao Fong
National University of Singapore

16:45
1587: PRESTE: Preserving Tiny Exponent Precision for Efficient Sub-8-bit LLM Inference and Fine-Tuning
Jehun Lee, Hyesung Jeon, Juchan Lee, Jaejoon Kim
Seoul National University

15:30 - 17:00
Reimagining Logic Synthesis
Room: San Carlos
Session Chair(s): Yi-Yu Liu, National Taiwan University of Science and Technology; Rongliang Fu, The Chinese University of Hong Kong

This session explores emerging paradigms that are reshaping logic synthesis, including GPU acceleration, differentiable optimization, AI-guided synthesis, and advanced technology mapping. The presented works demonstrate how modern algorithms and learning-based techniques are redefining synthesis flows to achieve superior quality of results, scalability, and computational efficiency.

15:30
28: UniMap: Unified Sizing and Remapping for ASIC Technology Mapping on GPU
Zhenxuan Xie{1}, Lixin LIU{1}, Tianji Liu{2}, Evangeline F.Y. Young{1}, Tianji Liu{3}
{1}The Chinese University of Hong Kong; {2}Institute of Computing Technology, CAS; {3}SKLP, Institute of Computing Technology, CAS

15:45
151: DiaMap: A Differentiable LUT Mapper for Area Optimization
Zhenxuan Xie, Evangeline F.Y. Young
The Chinese University of Hong Kong

16:00
499: YOCO: You Only Cut Once for Technology Mapping
Xin Ning{1}, Ruogu Ding{1}, Ulf Schlichtmann{2}, Weikang Qian{1}
{1}Shanghai Jiao Tong University; {2}Technical University of Munich

16:15
878: ChoMerge: Dual-Output LUT Optimization via LUT-Level Choice Network and Routing-Aware Merging
Xingyu Qin, Guande Dong, Chengshen Lin, Jianwang Zhai, Kang Zhao, Chengshen Lin
Beijing University of Posts and Telecommunications

16:30
580: Efficient Hybrid-action Search Framework with QoR Prediction for Logic Synthesis Sequence Optimization
Mingyang Chen{1}, Yunfei Dai{1}, Kaixiang Zhu{1}, Yuanqi Wang{1}, Wai-Shing Luk{2}, Lingli Wang{2}
{1}State Key Laboratory of Integrated Chips and Systems, Fudan University; {2}Fudan University

16:45
1309: EncyLogic: Logic Synthesis as Sequence Generation with Learned Priors and Physical Alignment
Zijian Jiang{1}, Chenyu Li{2}, Cunqing Lan{1}, Hongyang Pan{1}, Jintao Li{3}, Zhiang Wang{4}, Keren Zhu{1}
{1}Fudan University; {2}University of Science and Technology of China; {3}Shenzhen Institute for Advanced Study; {4}State Key Laboratory of Integrated Chips and Systems, Fudan University, Shanghai, China

15:30 - 17:00
System-level Techniques for AI Acceleration
Room: Blossom Hill I/II
Session Chair(s): Xiang Cheng, Peking University; Caiwen Ding, University of Minnesota Twin Cities

As AI models continue to grow in size and complexity, efficient inference requires coordinated advances across hardware, software, and system design. This session highlights recent advances in hardware and software techniques for accelerating modern AI workloads, showcasing the power of cross-layer optimizations in delivering high-performance, energy-efficient, and secure AI systems.

15:30
625: SMT-PartiSched: Automated DNN Operator Scheduling with SMT and Flexible Partitioning
Jiajia Huang, Kuan Jiang, Yili Guo, Fuyang Zhao, Wanli Chang
Hunan University

15:45
953: MatMoE: Matryoshka Mixture-of-Experts with Dynamic Mixed-Precision Quantization for Efficient Inference
Renjie Wei{1}, Haochen Huang{2}, Chuyu Qiu{3}, Jinqi Wen{2}, Dehao Xu{4}, Meng Li{5}
{1}Institute for Artificial Intelligence and School of Integrated Circuits, Peking University; {2}Peking University; {3}Sun Yat-sen University, Peking University; {4}Southeast University, Peking University; {5}Institute for Artificial Intelligence and School of Integrated Circuits, Peking University, Beijing Advanced Innovation Center for Integrated Circuits

16:00
1673: ExaGEMM: Exploration Framework for CPU-Driven ML Inference via Associative In-Register Computing for Low-Bit GEMM
Hyunwoo Oh{1}, Suyeon Jang{1}, Hanning Chen{1}, Sanggeon Yun{1}, Ryozo Masukawa{1}, Mohsen Imani{2}
{1}University of California, Irvine; {2}University of California Irvine

16:15
1770: TEA: Taylor-Expansion-Inspired Unified Hardware Acceleration for High-Precision Non-Linearity in Edge AI
Xingzhi Zhou{1}, Wei Jiang{1}, Jinyu Zhan{1}, Wenjie Yan{1}, Lin Zuo{1}, Yalan Ye{2}
{1}School of Information and Software Engineering, University of Electronic Science and Technology of China; {2}School of Computer Science and Engineering, University of Electronic Science and Technology of China

16:30
1146: HALO: A Layout-aware Framework for Efficient and Scalable Homomorphic Transformer Inference
Zhaoting Gong{1}, Ran Ran{1}, Jinjun Xiong{2}, Wujie Wen{1}
{1}North Carolina State University; {2}UT San Antonio

16:45
162: PAG: An Automated Generation Framework for Precision-Scalable Accelerators
Zhongdi Luo{1}, Zeyu Xue{1}, Jing Feng{1}, Yang Shi{1}, Zhongdi Luo{2}
{1}Key Laboratory of Advanced Microprocessor Chips and Systems, National University of Defense Technology; {2}Shahe Advanced Computing SIG

17:00 – 17:15
Transition

17:15 - 18:15
Efficient and Reliable AI Models for Edge Inference
Room: Willow Glen III
Session Chair(s): Tsung-Yi Ho, CUHK

This session focuses on algorithmic advances that improve the efficiency and reliability of DNNs and LLMs through novel quantization techniques, adaptive optimization, and reliability-aware learning. Topics span calibration-free LLM compression, low-bit representations, hardware-aware quantization, fault tolerance, and automated accelerator generation techniques that bridge model and hardware co-design.

17:15
427: FASQ: Flexible Accelerated Subspace Quantization for Calibration-Free LLM Compression
Ye Qiao{1}, Yian Wang{1}, Zhiheng Chen{1}, Hyoukjun Kwon{2}, Sitao Huang{2}
{1}University of California, irvine; {2}University of California, Irvine

17:30
891: PeLog: Rotation-Assisted Logarithmic Encoding for Low-Bit Large Language Model Quantization
Xiangfei Hu{1}, Lancheng Zou{2}, Xilong Kang{1}, Yuchen Liu{1}, Jiajie Xu{1}, Tairan Cheng{2}, Yuyang Ye{2}, Chang Meng{3}, Hao Yan{1}, Bei Yu{2}
{1}Southeast University; {2}The Chinese University of Hong Kong; {3}Eindhoven University of Technology

17:45
1026: Adapting DNNs to Embrace Voltage Scaling Induced Errors via Accumulation-Aware Quantization
Alexios Maras, Panagiotis Chaidos, Georgios Alexandris, Dimitrios Soudris, Sotirios Xydis
National Technical University of Athens

18:00
1152: MetaRel: Toward Efficient and Generalizable Reliability Evaluation of Neural Networks Under Bit-Flip Faults Empowered by Meta-Learning
Chao Chen, Jilong Zhang, Yan Li, Xiaoyang Zeng
Fudan University

17:15 - 18:15
Making Quantum Resources Count: Search, Noise, and Simulation
Room: Guadalupe
Session Chair(s): Robert Wille, TUM

How can quantum workloads do more with limited resources? This session pairs logarithmic encoding for qubit-efficient hyperdimensional search with tunable error correction that puts device noise to work in open-system simulation. It also examines when compile-once GPU state-vector simulation accelerates variational quantum workflows. Together, the papers address qubit cost, hardware noise, and the overhead of repeated circuit execution.

17:15
519: Qubit-Efficient Quantum Search for Hyperdimensional Decomposition via Logarithmic Encoding
Sanggeon Yun{1}, Hyunwoo Oh{1}, Ryozo Masukawa{1}, Raheeb Hassan{1}, Mohsen Imani{2}
{1}University of California, Irvine; {2}University of California Irvine

17:30
232: VQCSim: When Does Compile-Once Statevector Simulation Beat Generic Quantum Frameworks?
Anton Firc{1}, Martin Perešíni{1}, Vojtech Mrazek{1}, Kamil Malinka{1}, Vojtech Stanek{1}, Zbynek Licka{1}, Nouhaila Innan{2}, Walid El Maouaki{2}, Alberto Marchisio{2}, Muhammad Shafique{2}, Vojtech Mrazek{1}
{1}Brno University of Technology; {2}New York University Abu Dhabi

17:45
1702: TuneQEC: Noise-Aided Open Quantum Simulation via Tunable Quantum Error Correction
Lingjun Xiong{1}, Jinyang Li{2}, Lei Yang{1}, Yu Zhang{3}, Weiwen Jiang{1}
{1}George Mason University; {2}Miami University; {3}Los Alamos National Laboratory

17:15 - 18:15
Proofs in Practice: Zero-Knowledge Engines and Hardware Verification
Room: Almaden
Session Chair(s): William Hung, Cadence; Jiafeng Xie, Villanova University

From zero-knowledge proofs to CPUs and GPUs, trustworthy hardware demands both efficient computation and rigorous checking. This session brings together optimized point-addition hardware for proof generation, formal verification of constant-time CPUs, symbolic simulation that strengthens testbenches, and online differential verification of SIMT GPUs. Together, the papers connect specialized hardware with practical methods for building confidence in complex computing systems.

17:15
1038: Locus: A Framework for Exploring and Optimizing Point Addition Hardware for Zero-Knowledge Proofs
Gaurav Kuwar, Alhad Daftardar, Jianqiao Mo, Siddharth Garg, Brandon Reagen
New York University

17:30
1865: Inconstant: Refining Assumptions in Automated Formal Constant-Time CPU Verification
Flavien Solt{1}, Chris Fletcher{2}, Flavien Solt{1}
{1}National University of Singapore; {2}UC Berkeley

17:45
841: Forbench: Symbolic Simulation Helps Make Your Testbench More Formal
Ziyi Yang{1}, Wenbin Che{1}, Ziyue Zheng{1}, Guangyu Hu{2}, Hongce Zhang{1}
{1}The Hong Kong University of Science and Technology (Guangzhou); {2}The Hong Kong University of Science and Technology

18:00
618: GVM: Online Differential Verification for SIMT GPGPUs
Wenxuan Xie{1}, Mingyuan Ma{1}, Yuhan Wang{1}, Haonan Sun{1}, Wei Liu{1}, Hu He{1}, Mingyuan Ma{2}
{1}School of Integrated Circuits, Tsinghua University; {2}Tsinghua University

17:15 - 18:15
Thermal Design Across Scales: Interconnects, 3D Memory, and Mobile Systems
Room: Blossom Hill III
Session Chair(s): Yukai Chen, IMEC, Belgium; Naehyuck Chang, Samsung Electronics Device Solutions Research America (DSRA)

Thermal behavior affects designs from individual interconnects to complete systems. This session covers coupled electro-thermomigration analysis, package-induced temperature-sensor deviations in 3D-stacked DRAM, multiscale chip thermal simulation, and coordinated thermal management for mobile devices. The papers show how thermal modeling, measurement, and control support reliable design and operation across these scales.

17:15
1783: X-DTM: Intelligent Coordination of Cross-Layer Dynamic Thermal Management for Reliable Mobile Experience
Seonghyeon Jeon{1}, Hyegang Son{1}, Seung Hun Choi{2}, Young Geun Kim{1}
{1}Korea University; {2}Soongsil University

17:30
958: VULCAN: Unified Multi-Scale Chip Thermal Simulation System for Design-Technology Co-Optimization
Zhenkun Lin, Taoyu Zhou, Qipan Wang, Tianxiang Zhu, Yibo Lin, Meng Li, Runsheng Wang, Yandong He, Gang Du
Peking University

17:45
1234: An Electro–Thermomigration Coupled Solver via Physics-Informed GraphONet for Multisegment Interconnects
Zhaoyuan Liu{1}, Jiahao Wang{2}, Haodong Lu{1}, Kun Wang{2}
{1}Fudan university; {2}Fudan University

18:00
1307: Stress-Aware Modeling and Mitigation of Package-Induced Temperature Sensor Deviations in 3D-Stacked DRAM
Wenjie Zhu, Fenze Feng, Xunsheng Tan, Pengpeng Ren, Zhigang Ji, Hai-Bao Chen
Shanghai Jiao Tong University

17:15 – 18:15
ACM SIGDA Town Hall
Room: Blossom Hill I/II

17:15 – 18:15
Synopsys Sponsor Session: What to expect from Physical AI
Brandon Wang, VP of Technology Strategy
Room: San Jose Ballroom

The talk will cover the technology landscape of Physical AI, from technology Stack to Commercial Scale, it addresses where and how the capital and technology are converging, and what EDA companies like Synopsys can play a key role in the Physical AI Ecosystem.

18:30 – 20:00
Welcome Reception & SRC Poster Session
Room: Ballroom Foyer

8:00 – 8:30
Registration
Room: Ballroom Foyer

8:30 – 9:30
Keynote: Spatial Computing Architectures: Opportunities and CAD Flows in the Deep Learning and More-than-Moore Era
Vaughn Betz, University of Toronto, Cerebras Systems
Room: San Jose Ballroom

9:30 – 10:00
Coffee Break | Exhibits
Room: Ballroom Foyer

10:00 – 11:30
Panel: Thermal Co-Design for AI Chiplet Systems: Simulation, Packaging Innovation, and System-Level Optimization
Room: Willow Glen I/II

10:00 - 11:30
Agentic AI for EDA Workflows and Design Automation
Room: Blossom Hill III
Session Chair(s): Shantanu Telharkar, Advanced Micro Devices, Inc

Large language models are evolving from programming assistants into autonomous design agents for EDA. This session highlights agentic AI systems that combine domain expertise, retrieval-augmented generation, optimization, and collaborative reasoning to automate logic synthesis, physical implementation, and board-level design, advancing AI-native environments for end-to-end chip development.

10:00
398: GUALO: A Generalizable Uncertainty-Aware AI Agent for Logic Optimization
Jingxin Wang{1}, Sinian Zhang{2}, Yun Liang{3}, Yibo Lin{3}, Pengpeng Ren{1}, Runsheng Wang{3}, Weikang Qian{1}
{1}Shanghai Jiao Tong University; {2}University of Minnesota; {3}Peking University

10:15
1561: LEGO 2.0: A Unified Agentic Platform for Digital Chip Design and Optimization
Jincheng Lou{1}, Ruohan Xu{1}, Jiecheng Ma{2}, Boyang Wu{1}, Yuhao Ji{3}, Chunyuan Zhao{4}, Xinyu Qu{1}, Yuxiang Zhao{1}, Runzhe Tao{1}, Bei Yu{3}, Zhixiong Di{5}, Yibo Lin{1}
{1}Peking University; {2}Faculty of Integrated Circuits, Xidian University; {3}The Chinese University of Hong Kong; {4}School of Integrated Circuits, Peking University; {5}Southwest JiaoTong University

10:30
456: A Multi-Granularity Retrieval Augmented Generation Framework for EDA Tasks Leveraging Knowledge Graph
Yuheng Huang{1}, Yuhan Qin{2}, Yuan Pu{2}, Zhaotan Lin{2}, Yuntao Nie{1}, Zhuolun He{2, 3}, Jianwang Zhai{1}, Kang Zhao{1}, Bei Yu{2}
{1}Beijing University of Posts and Telecommunications; {2}the Chinese University of HongKong; {3} Tsinghua University

10:45
1701: Automated QoR Improvement in OpenROAD with Coding Agents
Amur Ghose{1}, Junyeong Jang{2}, Andrew Kahng{3}, Jakang Lee{2}
{1}UC San Diego; {2}University of California, San Diego; {3}UCSD CSE and ECE Depts.

11:00
1558: OpenLayout: An Open and Agentic-driven Platform for Holistic VLSI Physical Design and Benchmarking
Qiufeng Li{1}, Sizhe Tang{1}, Quan Cheng{2}, Zirui Wang{3}, Wantong Li{3}, Tian Lan{1}, Weidong Cao{1}, Quan Cheng
{1}The George Washington University; {2}Brown University; {3}University of California, Riverside

11:15
661: PCBclaw: Hierarchical Multi-Agent Schematic Synthesis and Verification for MCU System Boards
Yiming Zhang{1}, Yangbo Wei{1}, Zhanfei Chen{2}, Jinlong Yan{2}, Min Gao{3}, Ting-Jung Lin{2}, Chen Wu{2}, Lei He{4}, Chen Wu{5}, Yangbo Wei{1}
{1}Shanghai Jiao Tong University; {2}Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, China; {3}Google; {4}Eastern Institute of Technology, Ningbo; {5}Ningbo Institute of Digital Twin

10:00 - 11:30
Cross-Layer Co-Design for Multi-Chip Systems: Reliability, Security, and LLM-Centric Acceleration
Room: Blossom Hill I/II
Session Chair(s): Krishnendu Chakrabarty, Arizona State University

The explosive growth of large language models (LLMs) is reshaping modern computing, driving unprecedented demands on performance, energy efficiency, scalability, and reliability. At the same time, the emergence of 2.5D and 3D heterogeneous, chiplet-based multi-die architectures is redefining the hardware substrate for next-generation systems. While these architectures offer a compelling path forward, they also introduce significant challenges, including cross-die communication bottlenecks, thermal coupling, process variability, and increased susceptibility to reliability and security threats. This special session focuses on the design of multi-chip systems that can efficiently and robustly support LLM workloads. It explores cross-layer EDA and co-design strategies for heterogeneous architectures that integrate compute, memory, and interconnect fabrics to meet the unique computational and data-movement characteristics of LLMs. By bringing together experts in EDA, testing, and hardware security, the session highlights emerging methodologies that span device, architecture, and system levels. It will examine how architectural heterogeneity and advanced integration technologies can be harnessed to deliver scalable performance, while addressing critical concerns such as testability, fault tolerance, and security in complex multi-die systems. Through this perspective, the session aims to advance a holistic vision for reliable, secure, and high-performance multi-chip computing platforms tailored for the demands of AI workloads.

10:00
1990: Invited Paper: Alleviating the Memory and Routing Bottlenecks in LLM Mixture of Experts
Pratyush Dhingra, Jana Doppa, Partha Pratim Pande
Washington State University

10:30
1983: A Workload-Driven Test Framework for FeFET-Based CiM LLM Accelerators
Dhruv Thapar{1}, Ashish Reddy Bommana{1}, Farshad Firouzi{2}, Krishnendu Chakrabarty{1}
{1}Arizona State University; {2}Johns Hopkins University

11:00
Invited Talk: Hardware Security for Chiplet Based SoCs Powering Large Language Models
Amitabh Das
AMD

10:00 - 11:30
From Copilots to AI-Native Systems: The Future of Agentic EDA
Room: Guadalupe
Session Chair(s): Qi Sun, Zhejiang University

Existing EDA environments have been predominantly organized around human interpretation, intervention, and control. Consequently, large language model (LLM)-based methods have thus far shown their greatest practical utility as copilots layered atop established tools and design flows. As chip design becomes increasingly heterogeneous, tool-intensive, and optimization-driven, however, there is growing interest in AI systems that can coordinate tasks across abstraction levels and operate in a closed loop with simulation, verification, and optimization. This shift is not merely a matter of incremental automation; rather, it raises a more fundamental question of how design intent, tool state, and optimization objectives should be represented, propagated, and acted upon across the EDA stack. Agentic AI is emerging as a consequential direction for EDA, pointing toward systems that can plan, orchestrate tools, and iteratively refine solutions across complex design workflows. This special session brings together experts from academia and industry to examine the opportunities and challenges of Agentic EDA across the design stack. Topics of interest include agent architectures, tool integration, human-agent collaboration, and related system and workflow considerations. More broadly, the session aims to interrogate how EDA systems and workflows may need to evolve to support increasingly capable AI agents, and what new forms of representation, interface, feedback, and evaluation may be required across the design flow. Agentic EDA is not only a new application of AI to existing tools, but also a lens through which to reconsider the foundations of future design automation, pointing toward next-generation AI-native systems for EDA.

10:00
1991: From Pragmas to Partners: A Symbiotic Evolution of Agentic High-Level Synthesis
Niansong Zhang{1}, Sunwoo Kim{1}, Andrew Butt{1}, Shreesha Srinath{2}, Zhiru Zhang{1}
{1}Cornell University; {2}Cerebras Systems

10:20
Invited Talk: Design for AI, AI for Design: The Agentic Era of Electronic Design Automation
Ben Gu
Corporate Vice President of R&D, Cadence Design Systems

10:40
Invited Talk: FluxEDA: A Unified Execution Infrastructure for Stateful Agentic EDA
Cheng Zhuo
Zhejiang University

11:00
2008: EDA-env: A Reinforcement Learning Environment for RTL-to-GDS Design with Independent Scoring
Amur Ghose{1}, Andrew Kahng{1}, Yugesh Kothari{2}, Umang Pandey{2}, Souktik Roy{2}
{1}UC San Diego; {2}Ergodex AI, Inc.

10:00 - 11:30
Edge Intelligence: Novel Arithmetic, State-Space Models, and Agile Compilation
Room: Willow Glen III
Session Chair(s): Varun Manjunath, University of Southern California; Hiromitsu Awano, Nagoya University

Deploying complex models on constrained edge devices demands architectural creativity. This session explores Fermat NTT arithmetic, CORDIC-based FlashAttention, and semantic compiler lifting. Learn how novel addressing structures and State-Space Duality models are unlocking high-performance AI on FPGAs and embedded processors.

10:00
139: TensorLift: Automatic Extraction of Tensor-Level ISA Semantics from Accelerator RTL via MLIR Semantic Lifting
Ruijie Gao{1}, Haoran Jin{1}, Jirong Yang{1}, Nathan Bleier{1}, Jirong Yang{2}
{1}University of Michigan; {2}University of Texas at Austin

10:15
244: FermaLog: Fermat NTT Arithmetic for DSP-Free LLM Inference on FPGAs
Guanchen Li, Hao Yu, Sizhao Li
Harbin Engineering University

10:30
770: FALCON: A CORDIC-based FlashAttention Accelerator for On-Device Long-Context LLM Inference
Junyoung Park{1}, Seongjoon Cho{1}, Seungkyu Choi{2}
{1}Kyung Hee University; {2}Yonsei University

10:45
981: HEMERA: A Heterogeneous Memory-Centric Accelerator with Recursive Dataflow for Edge-Constrained State-Space-Duality Models Inference
Hao Ding{1}, Ling Liang{2}, Ruitong Qiao{2}, Dongxue Zhao{2}, Xiantong Qiu{2}, Jinshan Li{2}, Meng Li{2}, Lei Jin{3}, Zhiliang Xia{3}, Zongliang Huo{3}, Zongwei Wang{4}, Yimao Cai{5}
{1}School of Integrated Circuits, Peking University; Unigroup Guoxin Microelectronics Co., Ltd.; {2}School of Integrated Circuits, Peking University; {3}Yangtze Memory Technologies Co., Ltd.; {4}School of Integrated Circuits, Peking University; Beijing Advanced Innovation Center for Integrated Circuits; YanXin MicroElectronics Co., Ltd. (YXME); {5}School of Integrated Circuits, Peking University; Beijing Advanced Innovation Center for Integrated Circuits

11:00
1188: SpiderViT: Multi-Task ViT Accelerator with Sparse Attention Decoupling and Dynamic Expert Routing
Zheqi He, ZhenDong Zheng, Teng Wang, Wenqi Lou, Lei Gong, Chao Wang, Xuehai Zhou
University of Science and Technology of China

11:15
1387: XDA: Representation-Aware Exponent-Direct Addressing for Efficient Shared Nonlinear Units
Min Wu Jeong, Chae Eun Rhee
Department of Electronic Engineering, Hanyang University

10:00 - 11:30
Emerging-Device Memory & Compute-in-Memory
Room: Almaden
Session Chair(s): Ahmedullah Aziz, University of Tennessee Knoxville; Jonathan Birjiniuk, Oboe

This session brings together novel memory and in-memory computing architectures built on ferroelectric, memristor, and RRAM devices for dense, energy-efficient acceleration. Papers cover charge-domain and content-addressable memories, in-situ attention computation, and analog compute-in-memory co-design.

10:00
1131: TFCM-CAM: A Charge-Domain CAM Based on Highly Scaled Trench Ferroelectric Capacitive Memory for Dense and Energy-Efficient Associative Computing
Yi Tong{1}, Jingguang Lu{1}, Yan Le{1}, Kai Ni{2}, Chen Sun{1}, Xiao Gong{1}, Yi Tong{1}
{1}National University of Singapore; {2}University of Notre Dame

10:15
1827: Write-Free Self-Attention Acceleration for Transformer Models via In-Situ Q·KT Computation in Memristor Crossbar Arrays
Sanghyuk Kim, Jinwoo Park, Hyungjin Kim
Hanyang University

10:30
892: FADE-PBF: An Energy-Efficient FeFET-based Adaptively Deletable Partitioned Bloom Filter
Chenyu Ni{1}, Yu Qian{1}, Jingchao Hu{1}, Jiayi Wang{2}, Cheng Zhuo{1}, Xunzhao Yin{1}
{1}Zhejiang University; {2}Zhejiang University, Hangzhou, China

10:45
1810: BRQN: Blockwise Random Quantized Networks for RRAM-based Non-Deterministic Analog Computing-in-Memory via Co-Design
Pengbin Liu{1}, Chengyue Li{1}, Danian Dong{2}, Wenxuan Sun{2}, Dashan Shang{2}, Xu Zheng{2}, Xiaoxin Xu{2}
{1}School of Advanced Interdisciplinary Sciences, University of Chinese Academy of Sciences; {2}State Key Laboratory of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics of Chinese Academy of Sciences, Beijing, China

11:00
1418: DACAM: A High-Density One-Base-per-Cell Content Addressable Memory for Diversity-Aware Genomic Filtering
Yuxiao Jiang{1}, Ranran Wang{1}, Xuchu Huang{1}, Yifei Zhou{1}, Thomas Kämpfe{2}, Cheng Zhuo{1}, Xunzhao Yin{1}
{1}Zhejiang University; {2}TU Braunschweig

10:00 - 11:30
Faster Than SPICE: Physics-Guided AI for RF Passive Modeling, EM Simulation, and Parasitic Extraction
Room: San Carlos
Session Chair(s): Quan Chen, SUSTECH, China

Simulation and parasitic extraction are gatekeepers of RF and analog design correctness. This session delivers transformative advances in physics-guided RF passive modeling, topology-aware GNN circuit characterization, foundation models for RF design, EM-TCAD co-simulation, randomized sparse linear solvers, and CNN-based capacitance extraction, together achieving dramatic speedups while preserving physical accuracy.

10:00
449: FoundaRF: A Physics-Informed Foundation Model Framework for Radio-Frequency Passive Components
Houbo He{1}, Chris Han{1}, Hyunsu Chae{2}, Lei Xia{1}, Qiang Zhou{1}, Lawrence Liu{1}, Sensen Li{2}, David Pan{2}, Taiyun Chi{1}
{1}Rice University; {2}University of Texas, Austin

10:15
1229: RF-LEGO: A Compositional Framework with AI-Learned Building Blocks for RFIC Design
Yinyuan Chen {1,2}, Hui Qin {1,2}, Hao Kong {1,2}, Tianyi Ma {3}, Liang Wang {4}, Aoyang Zhang {1,2}, Leibo Liu {1,2}
{1} School of Integrated Circuits, Tsinghua University, Beijing, China; {2} Beijing National Research Center for Information Science and Technology, Beijing, China; {3} Beijing Tianji Interface Intelligent Technology Co., Ltd., Beijing, China; {4} Beijing Microelectronics Technology Institute, Beijing, China

10:30
1652: TARGet: Topology-Aware Fusion-based Radio Frequency Circuit Functional Modeling using Graph Neural Networks
Soroosh Noorzad, Sebastian Bodero, Morteza Fayazi
University of Utah

10:45
940: Fast Transient EM-TCAD Simulation via A Truncated Nonlinear Exponential Integrator Method
Haojie Wu{1}, Xianjie Wu{1}, Hang Zhou{2}, Quan Chen{3}, Hang Zhou{1}
{1}Southern University of Science and Technology; {2}Southern University of Science and Technologyy; {3}School of Microelectronics, Southern University of Science and Technology

11:00
835: RFDS-BLR: A Randomized Feedback Driven Stochastic Block Low Rank Solver for Industrial Scale Circuit Simulation
Yuyang Cai{1}, Dan Niu{1}, Zheng Jiang{1}, Xiaoming Chen{2}, Zhenya Zhou{3}, Zhou Jin{4}
{1}School of Automation, Southeast University, Nanjing, China; {2}Institute of Computing Technology, Chinese Academy of Sciences; {3}Huada Empyrean Software, Beijing, China; {4}Zhejiang University

11:15
1226: Flash-CNNCap: Capacitance Extraction via Image Mapping
Rodriguez Rodriguez Hector, Jiechen Huang, Wenjian Yu, Jiechen Huang
Tsinghua University

10:00 - 11:30
Hardware-Accelerated Verification and Physical Design
Room: San Jose Ballroom
Session Chair(s): Vikash Kumar, ARM

This session's papers redesign how equivalence checking, RTL emulation, GPU simulation, and DRC pattern matching represent their problem, then match that representation to real hardware: cell libraries for SAT solvers, RTL-native compilation, opcodes chosen by NPN equivalence, and geometry-aware hashing on FPGAs.

10:00
1104: Heterogeneous Synergy: Fusing SAT Solvers and GPU Simulation for CEC
Yang Sun{1}, Tianji Liu{2}, Martin D.F. Wong{3}, Evangeline F.Y. Young{1}
{1}The Chinese University of Hong Kong; {2}Institute of Computing Technology, CAS; {3}Hong Kong Baptist University

10:15
441: VeLoR: Verification-Oriented Logic Representation and Parallel Solving for Combinational Equivalence Checking
Zhang Hu, Wentao Jiang, Chengyu Ma, Yinshui Xia, Lunyao Wang, Zhufei Chu
School of Integrated Circuits, Ningbo University

10:30
1941: Compiled Representation Co-Optimization for Instruction-Driven RTL Emulators
Zhengyi Zhang, Sijing Yang, Lingli Wang
Fudan University

10:45
1971: Toucan: An RTL-Native GPU Simulator via MLIR-Based Lowering and Warp-Level Micro-Partitioning
Haoyuan Wang{1}, Yanqing Zhang{2}, Scott Beamer{1}
{1}University of California, Santa Cruz; {2}NVIDIA

11:00
961: ChoruSim: A Scalable Multi-GPU Simulator for Large-Scale SoC Verification Beyond 100M Gates
Zewei Zhong{1}, Yuwei Qu{1}, Xuyang Ge{2}, Shaoqun Li{1}, Kaixiang Zhu{1}, Wai-Shing Luk{1}, Lingli Wang{1}
{1}State Key Laboratory of Integrated Chips and Systems, Fudan University; {2}College of Computer Science and Artificial Intelligence, Fudan University

11:15
1860: DFMAcc: Geometry-aware FPGA Acceleration for DFM Pattern Matching with Dynamic Layout Partition
Zexu Zhang{1}, Haodong Lu{2}, You Hu{2}, Yang Xu{2}, Genggeng Liu{1}, Jianli Chen{2}, Kun Wang{2}
{1}Fuzhou University; {2}Fudan University

11:30 – 13:00
CEDA Luncheon | Breaking the AI Memory Wall - A Paradigm Shift Toward Memory-Centric Computing
Chia-Lin Yang, National Taiwan University
Room: San Jose Ballroom

13:00 – 14:30
Student Research Competition (Oral)
Room: Willow Glen I/II

13:00 - 14:30
Advances in Multi-Scale Power, Thermal, and PDN Optimization
Room: Blossom Hill I/II
Session Chair(s): Masanori Hashimoto, Kyoto University

This session presents multi-scale advances in thermal, power, and PDN analysis and optimization, spanning chiplet, IC, and PCB levels. Topics include fast thermal modeling, differentiable power optimization, topology-aware low-power modeling, ECO-driven power-grid solvers, GNN-guided IR-drop analysis, and physics-informed decap placement.

13:00
318: GL-MTM: A Fast Global-Local Modal Transfer-Matrix Method for Thermal Analysis of Chiplet-Based Systems
Yubiao Liu{1}, Jinyi Hu{1}, Pengju Zhang{1}, Kuoyuan Jia{1}, Quan Chen{2}, Jianhua Zhang{1}, Liang Chen{1}
{1}Shanghai University; {2}School of Microelectronics, Southern University of Science and Technology

13:15
1142: DiffPower: GPU-Accelerated Differentiable Switching Power Analysis and Optimization
Isaac Jacobson{1}, Zheng Zhao{2}, Rashmi Mehrotra{2}, Guanglei Zhou{1}, Vineet Rashingkar{2}, Yiran Chen{1}
{1}Duke University; {2}Synopsys

13:30
575: TAP-MAC: Low-Power Optimization of Multipliers and Multiply-Accumulators via Topology-Aware Power Modeling
Dongsheng Zuo, Yang Luo, Jiadong Zhu, Yuzhe Ma
The Hong Kong University of Science and Technology (Guangzhou)

13:45
1111: SPEED: A Path-Aware Structure-Pruning Sparse Direct Solver for Efficient ECO-Driven Power Grid Analysis
Jianfei Song, Zhenhua Wu, Zhou Jin, Cheng Zhuo
Zhejiang University

14:00
939: GR²W: GNN-Prior Residual Random Walk for Sign-Off-Guaranteed Localized PDN IR Drop Analysis
Xinguo Wang{1}, Dan Niu{1}, Wei W. Xing{2}, Dekang Zhang{1}, Zhenya Zhou{3}, Lijian Ni{1}, Zhou Jin{4}
{1}Southeast University; {2}University of Sheffield; {3}Empyrean Technology Co.,Ltd.; {4}Zhejiang University

14:15
1454: Analytical Concurrent Decoupling Capacitor Placement for PCB Power Distribution Networks via Physics-Informed Cutting Planes
Zong-Ying Cai, Yao-Wen Chang
National Taiwan University

13:00 - 14:30
AI for Security and Security for AI
Room: Almaden
Session Chair(s): Shaahin Angizi, NJIT; Nader Sehatbaksh, UCLA

AI-based techniques are steadily becoming indispensable tools for researchers for security analysis of various systems. On the other hand, researchers are facing mounting challenges in securing AI-based hardware systems. This track presents a collection of ideas for AI-based security analysis as well as attacks on and defenses for AI-based systems.

13:00
1732: FoREiGNN: Formal Gate-level Netlist Reverse Engineering with the help of Graph Neural Networks
Guangwei Zhao, Kaveh Shamsi
The University of Texas at Dallas

13:15
942: RRAM-DP: Device-Calibrated Differential Privacy for In-Memory Edge Learning
Kwunhang Wong{1}, Jichang Yang{2}, Karl M.H. Lai{2}, Hegan Chen{1}, Songqi Wang{2}, Wei Xuan{1}, Ning Lin{2}, Han Wang{2}, Xiaojuan Qi{2}, Zhongrui Wang{3}, Wei Xuan{4}
{1}AI Chip Center for Emerging Smart Systems (ACCESS), InnoHK; {2}The University of Hong Kong; {3}Southern University of Science and Technology; {4}Shenzhen University of Advanced Technology

13:30
1551: RTLGuard: A Lightweight Teacher-Student Defense for Poisoned RTL Code Generation Models
Mahshid Rezakhani, Kimia Azar, Hadi Kamali
University of Central Florida

13:45
279: CacheTrap: Unveiling a Stealthier Gray-Box Trojan against LLMs
Mohaiminul Al Nahian{1}, Abeer Matar A Almalky{2}, Gamana Aragonda{3}, Ranyang Zhou{3}, Sabbir Ahmed{2}, Dmitry Ponomarev{1}, Li Yang{4}, Shaahin Angizi{3}, Adnan Siraj Rakin{2}
{1}Binghamton University; {2}Binghamton University (SUNY); {3}New Jersey Institute of Technology; {4}University of North Carolina at Charlotte

14:00
333: ROBBIN: Rowhammer-Based Backdoor Injection during Inference
Saion Kumar Roy, Yufei Wang, A. Adam Ding, Yunsi Fei
Northeastern University

14:15
1731: Towards GNN-Based Reverse Engineering of Large Gate-Level Netlists
Guangwei Zhao, Shakil Ahmed, Kaveh Shamsi
The University of Texas at Dallas

13:00 - 14:30
Automated at Last: Intelligent Frameworks for Analog and Mixed-Signal Layout Automation
Room: San Carlos
Session Chair(s): Hai Li, Duke University

Analog layout automation remains EDA's most stubborn frontier. This session presents six innovative frameworks tackling placement constraint generation, core-centered vine representations, cross-node migration, patch-driven legacy reuse, LLM-guided DRC-aware code synthesis, and open-source mixed-signal flows. These papers collectively paint a decisive path toward fully automated, silicon-ready analog and mixed-signal physical design.

13:00
1752: APCG: Automatic Placement Constraint Generation for Analog Circuits
Binfeng Fan{1}, Rongliang Fu{2}, Wenxing Li{3}, Huilong Jiang{3}, Minglei Zhou{4}, Zhixuan Zhang{3}, Da Wang{3}, Fei Wang{5}, Xiaochun Ye{3}, Dongrui Fan{3}, Junying Huang{3}
{1}Hangzhou Institute for Advanced Study, UCAS, Hangzhou, China; {2}Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China; {3}SKLP, Institute of Computing Technology, CAS, Beijing, China; {4}School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China; SKLP, Institute of Computing Technology, CAS, Beijing, China; {5}School of Integrated Circuits, Shandong University, Jinan, China

13:15
127: VinePlace: Compact Analog and Mixed-Signal Placement via Core-Centered Vine Representation
Weijian Fan, Zhiyuan Zhao, Jinpeng Wang, Haoyi Zhang, Jincheng Lou, Runsheng Wang, Xiyuan Tang, Yibo Lin
Peking University

13:30
727: AMigration: Analog placement Migration via Adaptive Constraints Relaxation
Yapeng LI{1}, Peng Xu{2}, Shixin Chen{3}, Jindong Tu{4}, Tongkai Wu{4}, Chong Tong{1}, Bei Yu{3}, Tinghuan Chen{4}
{1}The Chinese University of Hongkong, Shenzhen; {2}The Chinese University of Hongkong; {3}The Chinese University of Hong Kong; {4}The Chinese University of Hong Kong, Shenzhen

13:45
1269: PatchWork: A Patch-Driven Framework for Efficient Analog Layout Automation via Legacy Knowledge Reuse
Shixin Chen{1}, Peng Xu{2}, Yapeng LI{3}, Tinghuan Chen{4}, Bei Yu{1}
{1}The Chinese University of Hong Kong; {2}The Chinese University of Hongkong; {3}The Chinese University of Hongkong, Shenzhen; {4}The Chinese University of Hong Kong, Shenzhen

14:00
655: SGLayout: Schema-Guided Closed-Loop Refinement for LLM-Driven Parameterized Layout Generation
Siyang Chen{1}, Yuan Pu{2}, Yuhan Qin{2}, Zhaotan Lin{2}, Siyuan Liang{2}, Zhuolun He{2}, Lining Zhang{1}, Bei Yu{2}
{1}Peking University; {2}The Chinese University of Hong Kong

14:15
1804: A Procedural Approach to Mixed-Signal Layout: Hybrid GDSFactory and OpenROAD Integration with Fine-Grained LEF Abstraction
Kevin Guan{1}, Camilo Velez Cuervo{1}, Fadi Kurdahi{2}
{1}University of California, Irvine; {2}University of California Irvine

---

13:00 - 14:30
Beyond Determinism: Harnessing Approximate and Stochastic Computing in Silicon
Room: Guadalupe
Session Chair(s): Tianyu Jia, Peking University; Jincong Lu, Samsung

From co-designed stochastic FIR filters to approximate logic synthesis and application-aware eDRAM, this session charts the frontier of non-deterministic hardware. These papers reveal how probabilistic, approximate methodologies can unlock superior performance and power gains, challenging traditional deterministic design dogma.

13:00
1773: Towards Energy-Efficient and Reliable Embedded DRAM via Application-Aware Optimizations
Zihan Zhang{1}, David Kong{2}, Duc Anh Nguyen{1}, Mark Hempstead{1}, Lillian Pentecost{3}, Gage Hills{2}, Marco Donato{1}, Gage Hills{2}
{1}Tufts University; {2}Harvard University; {3}Amherst College

13:15
368: DBS-CIM: A Dynamic Bit-block Sparse CIM Accelerator via Hardware–Algorithm Co-Design for DNN Inference
Xu Chen{1}, Kunming Shao{2}, Tim Kwang-Ting Cheng{2}, Chi-Ying Tsui{2}, Yi Zou{1}
{1}South China University of Technology; {2}The Hong Kong University of Science and Technology

13:30
233: Co-DSC: A Pole-Configurable Explicit-IIR Architecture for Dynamic Stochastic Computing FIR Filters
Lu Chen{1}, Xinlan Zhang{1}, Yexian Lin{1}, Kuncai Zhong{1}, Weikang Qian{2}
{1}Hunan University; {2}Shanghai Jiao Tong University

13:45
1350: A Quantization-Aware Compensation Framework with Frequency-Selective Design for Approximate DCT in VVC
Xu Wang, Ke Chen, Lixia Han, Bi Wu, Chenggang Yan, Chenghua Wang, Weiqiang Liu
Nanjing University of Aeronautics and Astronautics

14:00
180: ProU-SC: An Area-Efficient Random Number Source via Progressive Spatial Uniformity for Stochastic Computing
Jiangyuan Wang{1}, Chang Meng{2}, Jinhua Cui{1}, Kuncai Zhong{1}, Weikang Qian{3}
{1}Hunan University; {2}Eindhoven University of Technology; {3}Shanghai Jiao Tong University

14:15
643: Integrating Approximate Logic Synthesis into Approximate High-Level Synthesis
Jian Shi{1}, Ruicheng Dai{1}, Chang Meng{2}, Yue Yang{1}, Weikang Qian{1}
{1}Shanghai Jiao Tong University; {2}Eindhoven University of Technology

---

13:00 - 14:30
Breaking the Memory Wall: XPU-PIM Orchestration and Digital CIM Architectures
Room: Willow Glen III
Session Chair(s): Liang Zhang, Beihang University; Zheqi He, University of Science and Technology of China

Near-data processing is reshaping AI infrastructure. From 3D NAND to sophisticated LUT macros, this session highlights XPU-PIM scheduling, hierarchical mask encoding, and accuracy loss estimation. Witness the latest digital Compute-in-Memory frameworks driving latency-sensitive LLMs through SLO-aware and phase-split execution models.

13:00
134: NAND-MoE: Micro-Architecture-Aware Weight Planning and Heterogeneous Token-Chunking Pipelining for MoE-Focused Inference on 3D NAND CIM
Zihao Liu{1}, Hanjie Liu{1}, Guangyao Wang{1}, Qiguang Deng{2}, Kaixuan Zhang{2}, Wang Kang{1}
{1}School of Integrated Circuit Science and Engineering, Beihang University, China; {2}Zhicun Research Lab, Beijing, China

13:15
156: DASH: Data Allocation Scheduler for Heterogeneous XPU-PIM Accelerators towards Edge LLM Inference
Yueqi Zhang{1}, Chao Fang{2}, Qi Wu{1}, Ye Lin{1}, Yuan Du{1}, Li Du{1}
{1}Nanjing University; {2}KU Leuven

13:30
1455: REACH: Architecting Residue-Inspired Outlier-Aware CIM Accelerator With Hierarchical Mask Encoding for LLM Inference
Zhonghao Chen, Hongtao Zhong, Teng Yi, Huazhong Yang, Xueqing Li
Tsinghua University

13:45
1501: PALETTE: A Reconfigurable LUT-CIM Macro Featuring On-Chip Parallel LUT Update and Pipelined Accumulation
Hiroto Tagata{1}, Takashi Sato{1}, Hiromitsu Awano{2}
{1}Kyoto University; {2}Nagoya University

14:00
1512: Small-Noise is All You Need: Fast and Precise Estimation of Compute-In-Memory Accuracy Loss
Soojung Bae{1}, Tanner Andrulis{1}, Vivienne Sze{1}, Joel Emer{2}
{1}MIT; {2}MIT/NVIDIA

14:15
1892: SplitPIM: SLO-Aware XPU-PIM Orchestration for Phase-Split LLM Inference
Shuai Zhou, Xinran Li, Jianli Chen, Jun Yu, Kun Wang
Fudan University

13:00 - 14:30
LLMs for RTL Design, Verification, and Debugging
Room: Blossom Hill III
Session Chair(s): Saurabh Kotiyal, AMD

Foundation models are transforming digital hardware design and verification. This session highlights AI-assisted RTL development, including specification understanding, RTL generation, automated debugging, assertion generation, and joint optimization of functional correctness and implementation quality. Together, reasoning-enabled LLMs emerge as effective collaborators throughout the digital hardware design lifecycle, improving productivity, quality, efficiency.

13:00
323: CorrectRTL: Boost RTL Generation Quality Through Automated Self-Correction with Instruction Decomposition
Jing WANG{1}, Shang Liu{2}, Hangan Zhou{2}, Wenji Fang{2}, Yuchao Wu{2}, Yugao Zhu{2}, Zhiyao Xie{2}
{1}The Hong Kong University of Science and Technology; {2}Hong Kong University of Science and Technology

13:15
919: Work Smarter, Not Harder: An LLM-Enabled Agentic Framework for Automated RTL Generation and Debugging
Jie Zhou{1}, Youshu Ji{2}, Shiyue Zhang{1}, Junhao Ye{1}, Ning Wang{3}, Wenjun Han{4}, Xinwei Fang{5}, Nan Guan{3}, Xi Wang{1}, Zhe Jiang{1}, Hugo Jiang{6}
{1}School of Integrated Circuits, Southeast University, China; National Center of Technology Innovation for EDA, China; {2}National Center of Technology Innovation for EDA, China; {3}Department of Computer Science, City University of Hong Kong, Hong Kong; {4}Nanjing Research Institute of Electronics Technology, China; {5}Department of Computer Science, University of York, UK; {6}University of Cambridge

13:30
1870: COEVO: Co-Evolutionary Framework for Joint Functional Correctness and PPA Optimization in LLM-Based RTL Code Generation
Heng Ping, Peiyu Zhang, Shixuan Li, Wei Yang, Anzhe Cheng, Shukai Duan, Xiaole Zhang, Paul Bogdan
University of Southern California

13:45
840: Stay on the Right Track: A Reasoning-Aware Dual-Agent Framework for Automated RTL Debugging
Jie Zhou{1}, Shiyue Zhang{1}, Youshu Ji{2}, Xinyao Jiao{1}, Ning Wang{3}, Wenjun Han{4}, Xinwei Fang{5}, Nan Guan{3}, Xi Wang{1}, Zhe Jiang{1}, Hugo Jiang{6}
{1}School of Integrated Circuits, Southeast University, China; National Center of Technology Innovation for EDA, China; {2}National Center of Technology Innovation for EDA, China; {3}Department of Computer Science, City University of Hong Kong, Hong Kong; {4}Nanjing Research Institute of Electronics Technology, China; {5}Department of Computer Science, University of York, UK; {6}University of Cambridge

14:00
1006: Invertible Specifications for Robust Hardware Design with LLMs
Andrew S. Cassidy, Guillaume Garreau, Jay Sivagnaname, Mike Grassi, Bernard Brezzo, John V. Arthur, Dharmendra S. Modha, Andrew Cassidy
IBM Research

14:15
720: AssertLLM2: A Comprehensive LLM Benchmark for Assertion Generation from Design Specifications
Yuchao Wu{1}, Wenji Fang{1}, Jing WANG{2}, Wenkai Li{1}, Ziyan Guo{1}, Zhiyao Xie{1}
{1}Hong Kong University of Science and Technology; {2}The Hong Kong University of Science and Technology

13:00 - 14:30
Reliability and Testing at Scale
Room: San Jose Ballroom
Session Chair(s): Keren Zhu, Fudan University; Huaixi Lu, Amazon

This session's papers find exactly where vulnerability concentrates, then prioritize protection and testing there: a neural network tracing how a fault spreads through a circuit, radiation physics mapped onto a chip layout, a sampling algorithm choosing which scan to run next, and timed automata blocking only unsafe DRAM commands.

13:00
284: An Efficient Framework for Soft Error Analysis and Selective Hardening via Graph Selective State-Space Model
Borui Li{1}, Yuyan Wang{1}, Changhao Yan{1}, Keren Zhu{1}, Zhaori Bi{1}, Nan Zhang{2}, Zhipeng Liu{2}, Xuan Zeng{1}
{1}Fudan University; {2}Shanghai Semisight Information Technology Co. Ltd.

13:15
518: Layout-Based Soft Error Analysis for Datapaths in Low Earth Orbit
Barry Lyu, Jacob Mack, Cheng Chiu, Dennis Sylvester, Nathan Bleier
University of Michigan

13:30
497: TreeFI: Value-Aware Statistical Fault Injection for Deep Neural Networks
Noam Bires, Marcello Traiola, Angeliki Kritikakou, Elisa Fromont
Univ Rennes, Inria, IRISA, CNRS

13:45
343: SAVAL: Signal-Driven Adaptive Validation for Post-Silicon Testing via Sequential Decision-Making and Mechanism-Informed Counterfactual Exploration in NAND Flash Memory
Sanghyeok Park{1}, Soyoon Park{2}, Simon S. Woo{3}
{1}Sungkyunkwan University, Samsung Electronics; {2}Samsung Electronics; {3}Sungkyunkwan University

14:00
567: SPGFS: A Distributed ATPG Framework with Synchronous Pattern Generation and Fault Simulation
Lipei Zhang{1}, Fuyu Wang{1}, Minghua Shen{1}, Liuzheng Wang{2}, Yu Huang{2}
{1}Sun Yat-sen University; {2}HiSilicon

14:15
1898: Formal Runtime Enforcement for Fault-Tolerant DRAM-based Memory Controllers
Lalit Mohanani, Ayush Anand, Dr. Debiprasanna Sahoo, Dr. Srinivas Pinisetty
Department of Computer Science and Engineering, Indian Institute of Technology Bhubaneswar

14:30 – 15:00
Coffee Break | Exhibits
Room: Ballroom Foyer

15:00 - 16:30
Advances in Statistical, Multi-Corner, and AI-Driven Timing Analysis
Room: Blossom Hill I/II
Session Chair(s): Tsung-Wei Huang, University of Wisconsin at Madison; Hui-Ru Jiang, National Taiwan University

This session presents advances in timing analysis, prediction, optimization, and sign-off across modern designs. Topics include LVF-based statistical timing, corner-adaptive pre-routing prediction, multimodal crosstalk estimation, differentiable multi-corner bus-skew synchronization, heterogeneous 3D IC optimization, and conformal-calibration-based FPGA timing sign-off.

15:00
599: STELA: Statistical Timing Engine with Complete LVF Analysis and Efficient Path Enumeration
Xiaotong Feng{1}, Shichang Ye{1}, SenZhi Wang{1}, Yue Zhang{2}, Peng Wang{1}
{1}School of Integrated Circuit Science and Engineering, Beihang University; {2}Step Function Technology Ltd.

15:15
1167: CAPTiming: Pre-Routing Multi-Corner Timing Prediction via Corner-Adaptive Timing Propagation
Zhi Li{1}, Junming Jiao{1}, Xiangfei Hu{2}, YuHan Dong{1}, Zeyuan Deng{1}, Peng Cao{1}
{1}Southeast University; {2}SouthEast University

15:30
1390: Crosstalk-Aware Timing Estimation via Multimodal Learning
Yuren Zhou{1}, Fan Yang{1}, Yuyang Ye{2}, Qing He{1}
{1}Tongji University; {2}The Chinese University of Hong Kong

15:45
1123: Differentiable Multi-corner Bus Skew Synchronization via Neural Surrogate Model
Jaemin Seo, Myungjun Kook, Jongho Yoon, Seokhyeong Kang
Pohang University of Science and Technology

16:00
781: HeteroOpt3D: Timing Optimization for Heterogeneous Hybrid-bonding-based Face-to-Face 3D ICs
Yuhao Ji{1}, Siying Yu{1}, Yunqi Shi{2}, Fangzhou Liu{1}, Yuxuan Zhao{1}, Jiaxi Jiang{1}, Zizheng Guo{3}, Chao Qian{2}, Yibo Lin{3}, Bei Yu{1}
{1}The Chinese University of Hong Kong; {2}Nanjing University; {3}Peking University

16:15
267: Dr. RTL: Autonomous Agentic RTL Optimization through Tool-Grounded Self-Improvement
Wenji Fang, Yao Lu, Shang Liu, Jing WANG, Ziyan Guo, Junxian He, Fengbin Tu, Zhiyao Xie
Hong Kong University of Science and Technology

15:00 - 16:30
AI for Physical Design and Semiconductor Manufacturing
Room: Blossom Hill III
Session Chair(s): Hongxin Kong, Synopsys, Inc.; Yu-Guang Chen, National Central University

Artificial intelligence is transforming physical design and semiconductor manufacturing by improving optimization, manufacturability, and design quality. This session features AI-driven techniques for placement, routability prediction, parasitic estimation, mask rule checking, computational lithography, and wafer anomaly detection, highlighting AI's growing impact across the backend design and manufacturing flow.

15:00
1643: HELIX: Placement Heuristics Evolution via LLM-Instructed Code Exploration
Andrew B. Kahng, Sayak Kundu, Bodhisatta Pramanik
University of California San Diego

15:15
1086: CLDRoute: Conditional Latent Diffusion for Routability Map Generation in Physical Design
Kiran Thorat{1}, Nicole Meng{2}, Caiwen Ding{3}, Yingjie Lao{2}, Zhijie Jerry Shi{1}
{1}University of Connecticut; {2}Tufts University; {3}University of Minnesota Twin Cities

15:30
1331: ParasGB: A Graph Benchmark Suite for Parasitic Estimation on AMS Circuits
Jiajun Zou{1}, Jiawei Liu{2}, Ao Liu{1}, Junnong Tian{1}, Yibin Zhang{3}, Chengjie Liu{4}, Yuxi Wang{4}, Shan Shen{1}, Wenhua Gu{1}, Jun Yang{5}, Wenjian Yu{3}
{1}Nanjing University of Science and Technology; {2}Chinese University of Hong Kong; {3}Tsinghua University; {4}Nanjing University; {5}Southeast University

15:45
1355: UniMRC: A Differentiable Multimodal Framework for Unified Curvilinear-Manhattan Mask Rule Checking
Wenxuan Dong{1}, Tianyi Li{1}, Kai Ma{1}, Zhen Wang{1}, Hanbin Dong{1}, Nan Wang{1}, Su Zheng{2}, Xinyun Zhang{1}, Jingyi Yu{1}, Hao Geng{1}
{1}ShanghaiTech University; {2}The Chinese University of Hong Kong

16:00
1461: Ariadne's Thread: A Multimodal Benchmark Weaving Point Solutions into Process Chains for AI-Driven Computational Lithography
Shenshuo Yao, Liuke Wang, Yuyang Chen, Yinuo Zhu, Zhen Wang, Jiajie Zhang, Tianyi Li, Yiwen Wu, Xinheng Li, Kai Ma, Nan Wang, Wenxuan Dong, Xinyun Zhang, Jingyi Yu, Hao Geng
ShanghaiTech University

16:15
1314: RWAD: Reliability-Guided Text-Free Zero-Shot Wafer Anomaly Detection
Haowen Lan{1}, Tianrui Zhang{1}, Ziheng Zang{1}, Silin Chen{1}, Kangjian Di{1}, Andy Liu{1}, Guohao Wang{2}, Wenzheng Zhao{2}, Ningmu Zou{1}
{1}Nanjing University; {2}ZetaTech Co.,Ltd

15:00 - 16:30
AI Hardware Architectures and Intelligent Design Automation
Room: Almaden
Session Chair(s): Iraklis Anagnostopoulos, National Technical University of Athens; Naehyuck Chang, Samsung Electronics Device Solutions Research America

This session presents innovations spanning processing-in-memory architectures, tensor-core microarchitecture exploration, compute-in-memory compilation, AI-assisted hardware generation, and machine learning techniques for RTL optimization and physical design. Together, these works demonstrate how advances across the hardware design stack can accelerate next-generation AI systems.

15:00
463: Tri-Stack-PIM: A Heterogeneous Multi-Die 3D-Integrated Processing-in-Memory Architecture for Efficient LLM Inference
Shuao Jia{1}, Zichao Ling{1}, Chen Bai{2}, Kang Zhao{1}, Jianwang Zhai{1}
{1}Beijing University of Posts and Telecommunications; {2}Fudan University

15:15
544: TensorForge: Synthesis-Backed Design Space Exploration of Multi-Precision Tensor Core Microarchitectures
Wei Liu{1}, Mingyuan Ma{1}, Yi Yu{2}, Xiaoyou Song{1}, Qianliang Liu{1}, Yuhan Wang{1}, Yushan Guan{1}, Jingzhou Li{1}, Wenxuan Xie{1}, Yumin Hou{2}, Hu He{1}, Mingyuan Ma{1}
{1}Tsinghua University; {2}Beijing University of Technology

15:30
1482: PolyCIM: Improving Data Reuse in Digital CIM Accelerators with Polyhedral-Based Compilation
Yingjie Qi, Cenlin Duan, Yiou Wang, Yikun Wang, Xiaolin He, Weisheng Zhao, Jianlei Yang
Beihang University

15:45
1638: DiffAxE: Diffusion-driven Hardware Accelerator Generation and Design Space Exploration
Arkapravo Ghosh{1}, Abhishek Moitra{2}, Abhiroop Bhattacharjee{2}, Ruokai Yin{2}, Priyadarshini Panda{1}
{1}University of Southern California; {2}Yale University

16:00
748: Cedar: Learning to Optimize RTL via Accumulating Verified Rewrite Rules
Niansong Zhang{1}, Chenhui Deng{2}, Johannes Maximilian Kuehn{2}, Chia-Tung Ho{2}, Cunxi Yu{3}, Zhiru Zhang{1}, Brucek Khailany{2}
{1}Cornell University; {2}NVIDIA; {3}University of Maryland, College Park

16:15
782: Semi-Supervised Congestion-Aware Global Placement Optimization
Xiaodong Zhu{1}, Wenxin Yu{1}, Zhilei Wang{1}, Youzhi Zheng{1}, Yishan Zhang{1}, Jun Hu{1}, Jingwei Lu{2}
{1}Southwest University of Science and Technology; {2}TikTok

15:00 - 16:30
Beyond Logic Optimization
Room: San Carlos
Session Chair(s): Weihua Xiao, New York University

This session focuses on innovative approaches that extend the boundaries of traditional logic optimization. Topics include Boolean matching, Boolean decomposition, arithmetic optimization, approximate logic synthesis, and low-power optimization. Together, these contributions introduce novel optimization methodologies that improve circuit quality, efficiency, and scalability for next-generation digital designs.

15:00
1572: Toward Elevating Boolean Matching to the Word Level
Kuo-Wei Ho, Jiun-Hao Chen, Hsin-Ying Tsai, Jie-Hong Jiang
National Taiwan University

15:15
1728: CODC Computation for Arithmetic Circuits using Algebraic Geometry
Bhavani Priya Sampath Kumar{1}, Ritaja Das{1}, Bailey Martin{1}, Priyank Kalla{2}, Florian Enescu{3}
{1}University of Utah; {2}The University of Utah; {3}Georgia State University

15:30
485: Multi-Output Boolean Decomposition via Compatible Vector Clustering
Ruofei Tang{1}, Yiwen Wang{2}, Xing Li{2}, Lei Chen{2}, Mingxuan Yuan{2}, Martin Wong{1}, Jianliang Xu{1}
{1}Hong Kong Baptist University; {2}Huawei Noah's Ark Lab

15:45
631: E-ALS: Approximation-Potential-Aware E-Graph Rewriting for Approximate Logic Synthesis
Bin Sun{1}, Jianan Mu{2}, Jiaxi Zhang{3}, Rengang Zhang{1}, Zhiteng Chao{2}, Jing Ye{2}, Huawei Li{1}
{1}State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences; CASTEST, Beijing; {2}State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China; {3}Department of Computer Science, Peking University

16:00
1947: Multiplier Optimization by Exploring Logical Correlations Among Partial Products
Ao Liu{1}, Siting Liu{2}, Hui Wang{1}, Cheng Chen{1}, Jie Han{3}, Honglan Jiang{1}
{1}Shanghai Jiao Tong University; {2}ShanghaiTech University; {3}University of Alberta

16:15
743: GLOW: Glitch Optimization with Retiming for Low Power
Jiantao Liu{1}, Carmine Rizzi{2}, Pravriti Jaipuriyar{3}, Ulrike Schorr{3}, Sascha Richter{3}, Ankush Sood{3}, Lana Josipović{2}, Ankush Sood{4}
{1}ETH Zürich; {2}ETH Zurich; {3}Cadence Design Systems; {4}Cadence design systems

---

15:00 - 16:30
Beyond Text: Hardware Architectures for Spatial, Temporal, and Vision Generation
Room: Willow Glen III
Session Chair(s): Nathaniel Bleier, University of Michigan; Yi Sheng, University of South Florida

Generative AI extends far beyond language. This session delves into custom accelerators for 4D Gaussian splatting, asynchronous sparse RNNs, and diffusion transformers. See how decoupled algorithms and frequency-domain pipelines are accelerating LiDAR, event-based vision, and high-fidelity video generation on-device.

15:00
160: Splatoon: An Integrated System for Point-Cloud LiDAR-Based Gait Recognition with Algorithm-Hardware Co-Design
Haoyuan Li{1}, Shaoxiong Zhang{2}, Zhenzhe Chen{2}, Weirong Dong{2}, Quan Cheng{2}, Feng Liang{3}, Takashi Sato{2}
{1}Xi'an Jiaotong University/Kyoto University; {2}Kyoto University; {3}Xi'an Jiaotong University

15:15
223: MAF-4D: A Motion-Adaptive Foveated Accelerator for 4D Gaussian Splatting with Balanced Scheduling
Ruohang Xu, Zecheng Zhou, Guoxiang Li, Xinyu Qu, Runnan Xu, Yufei Ma, Xinyu Qu
Peking University

15:30
814: APT: Accelerating Diffusion Transformers via Attention Probability-Guided Pruning and Quantization
Sungyeob Yoo, Seeyeon Kim, Joonyong Park, Seunghee Han, Joo-Young Kim
KAIST

15:45
1084: Moxie: A High-Throughput, Low-Latency FPGA Accelerator of Asynchronous Sparse Convolutional RNNs for Event-Based Vision
Andrea Galimberti{1}, Stefano Cecchettini{2}, Riccardo Santambrogio{1}, Matteo Matteucci{1}, Davide Zoni{1}
{1}Politecnico di Milano; {2}Barcelona Supercomputing Center

16:00
1115: FrequencyFormer: A Co-Designed Sensor-to-Processor Pipeline for Frequency-Domain Vision Transformer Inference
Chengwei Zhou{1}, Ovishake Sen{2}, Xuming Chen{1}, Rishith Paramasivam{2}, Shaahin Angizi{3}, Swarup Bhunia{2}, Baibhab Chatterjee{2}, Gourav Datta{1}
{1}Case Western Reserve University; {2}University of Florida; {3}New Jersey Institute of Technology

16:15
1124: SAND: A Sparse-Aware and Diagonal-Decoupled Algorithm-Hardware Co-Acceleration Framework for Floating-Point Video Generation Models
Haotian Chen, Qunkang Meng, Zhan Li, Xingyu Shi, Jie Hu, Zhihan Zhang, Hongyu Wu, Zhiyi Chen, Hao Wang, Sheng Chang
Wuhan University

15:00 - 16:30
Computing in Memory
Room: Guadalupe
Session Chair(s): Hai Li, Duke University

This sessions covers aspects for computing in memory technologies

---

15:00
786: NOVA-Flash: A Low-Power Nc-Si Flash CIM Macro Enabling In-Situ Input-Weight Fusion and Domain-Partitioned Signed MAC
Xianping Liu{1}, Jian Huang{2}, Zihan Zheng{2}, Longbin Chen{2}, Xinrui Zhang{2}, Ruibin Zhou{2}, Haobin Li{2}, Jian Cheng{3}, Zhiyi Yu{2}
{1}Sun Yat-sen University,Pen Cheng Laboratory; {2}Sun Yat-sen University; {3}Peng Cheng Laboratory

15:15
704: PolySim: Deterministic Polynomial Surrogates for Cross-Modal Retrieval on CiM
Xinzhao Li{1}, Charles Power{1}, Pengyu Ren{2}, Jongun Won{2}, Likai Pei{2}, Yuting Hu{3}, Jinjun Xiong{3}, Alptekin Vardar{4}, Ningyuan Cao{2}, X. Sharon Hu{2}, Thomas Kämpfe{5}, Kai Ni{2}, Ruiyang Qin{1}
{1}Villanova University; {2}University of Notre Dame; {3}University at Buffalo; {4}Fraunhofer IPMS; {5}TU Braunschweig

15:30
349: PRISM: A Programming-Free On-Device Multi-task Adaptation Framework for ReRAM-based Computing-in-Memory Accelerator
Yihang ZUO{1}, Wanhao YU{2}, Asmer Hamid Ali{1}, Li Yang{2}, Deliang FAN{1}
{1}Arizona State University; {2}University of North Carolina at Charlotte

15:45
986: HFSCIM: Hybrid FeFET-SRAM CIM Architecture for Continuous-Time Inference and Online Reinforcement Learning
Hegan Chen{1}, Zihao Xuan{1}, Yewen Li{1}, Ning Lin{2}, Jia Chen{1}, Yan Zeng{2}, Jichang Yang{2}, Songqi Wang{2}, Han Wang{2}, Zhongrui Wang{3}, XUAN ZIHAO{4}
{1}AI Chip Center for Emerging Smart Systems (ACCESS), InnoHK; {2}the University of Hong Kong; {3}Southern University of Science and Technology; {4}The Hong Kong University of Science and Technology，AI Chip Center for Emerging Smart Systems

16:00
167: TEA-Map: Thermally Efficient Mapping Exploration for 3D CIM Accelerators with Bayesian Optimization
Lihua An, Jiayi Li, Pingqiang Zhou
Shanghaitech University

15:00 - 16:30
Crafting Cells at Advanced Nodes: SAT, LLMs, and Timing-Aware Layout Automation
Room: San Jose Ballroom
Session Chair(s): Chung-Kuan Cheng, University of California, San Diego; Mengxi Chen, Siemens EDA

As nodes push to 2nm GAA and beyond, standard cells make or break a design's PPA. This session unites exact SAT layout and transistor-network synthesis, LLM-automated design-rule integration, coupling-aware timing optimization, multi-bit flip-flop banking, and mixed-cell-height legalization—showing how automation conquers the angstrom-era cell.

15:00
286: Optimal Transistor-Level Layout of Single- and Multi-Row Cells for 2nm and Beyond
Stefan Hougardy{1}, Benjamin Klotz{1}, Malte Schürks{1}, Armin Settels{1}, Tobias Werner{2}
{1}University of Bonn; {2}IBM

15:15
1052: SAT-Based Free-Topology Transistor Network Synthesis under Stack-Height, Multi-Stage and Multi-Output Options
Byeonggon Kang{1}, Sehyeon Kim{2}, Seokhyeong Kang{2}, Alan Mishchenko{3}, Masahiro Fujita{4}, Susmita Sur-Kolay{1}, Bill Lin{1}, Chung Kuan Cheng{1}
{1}University of California, San Diego; {2}Pohang University of Science and Technology; {3}University of California, Berkeley; {4}The University of Tokyo

15:30
923: AutoDRI: Bridging the Semantic Gap for Automated Design Rule Integration in CP-SAT-Based Cell Synthesis under Multi-Patterning
Zihao Chen{1}, Chung-Kuan Cheng{2}, Yuhao Ren{1}, Yucheng Wang{3}, Zhiang Wang{1}
{1}State Key Laboratory of Integrated Chips and Systems, Fudan University, Shanghai, China; {2}Computer Science and Engineering Department, UC San Diego; {3}University of California, San Diego

15:45
591: ScoutCell: Sensitivity-Guided Coupling Optimization for Timing-Driven Standard Cell Synthesis
Kairong Guo, Haoyi Zhang, Haichuan Liu, Chunyuan Zhao, Yibo Lin
Peking University

16:00
1034: Multibit Flip-Flop Banking via Compact Placement and Iterative Matching
Yewei Wang{1}, Fangyao Chen{2}, Shiju Lin{1}
{1}The Hong Kong University of Science and Technology (Guangzhou); {2}Huazhong University of Science and Technology

16:15
1205: Fine-Grained Irregular Region-Based Legalization for Non-Integer Mixed-Cell-Height Designs
Yan-Yu Chen, Da-Wei Huang, Shao-Yun Fang
National Taiwan University of Science and Technology

15:00 - 16:30
Design Automation for Early Fault-Tolerant Quantum Computing
Room: Willow Glen I/II
Session Chair(s): Anupam Chattopadhyay, Nanyang Technological University

As quantum computing transitions from the NISQ era to the early fault-tolerant (Early FTQC) regime, the field faces an urgent need for cross-layer design automation (QDA) methodologies that bridge algorithms, compilers, and hardware. In this emerging phase, where logical qubits will remain scarce, noisy, and resource-intensive, progress depends critically on automating key design tasks that have long been central to classical EDA: resource estimation, physical modeling, placement and routing, timing analysis, and calibration optimization. This special session brings together researchers spanning quantum architecture, compiler optimization, circuit synthesis, and device modeling to discuss how design-automation principles can accelerate the realization of Early FTQC. The session will conclude with an outlook on open-source infrastructures and benchmarking methodologies to foster reproducibility and community collaboration. Attendees will gain a holistic understanding of how design-automation concepts can drive the next leap from prototype-scale NISQ devices toward robust, scalable Early FTQC systems.

1989: Invited Paper: Structure-Aware Placement and Routing of Multi-Controlled Toffoli on Bivariate Bicyclic Code Architectures
Anik Basu Bhaumik, Suman Dutta, Siyi Wang, Anupam Chattopadhyay
Nanyang Technological University

1993: Invited Paper: Shot-Budget-Aware Scheduling of Post-Selected Error Detection in Quantum Optimization Circuits
Kuan-Cheng Chen{1}, Marcelin Gallezot{1}, Samuel Yen-Chi Chen{2}, Samuel Yen-Chi Chen{2}, Kuan-Cheng Chen{3}
{1}JIJ Inc.; {2}Wells Fargo; {3}JIJ

1986: Invited Paper: Thermal-Aware Design Automation for Early Fault-Tolerant Quantum Computing: A Multi-Scale Modeling Framework
Zhemin Zhang{1}, Jixi He{1}, Zhiding Liang{1}, Zhemin Zhang{1}, Zhiding Liang{2}
{1}The Chinese University of Hong Kong; {2}University of Notre Dame

1987: T-Count Reduction in Quantum Boolean Circuits by Allowing Relative Phase
David Clarino, Zanhe Qi, Shigeru Yamashita, Shigeru Yamashita
Ritsumeikan University

16:30 – 16:45
Transition

16:45 - 17:45
AI-Driven PPA Modeling, EDA Tuning, and Physical Design Optimization
Room: Blossom Hill I/II
Session Chair(s): Masanori Hashimoto, Kyoto University; Andy Yu-Guang Chen, National Tsing Hua University

This session presents machine-learning-driven methods for improving PPA across design stages. Topics include Transformer-based multi-corner PPA modeling for microprocessors, diversified Bayesian optimization for EDA parameter tuning, retrieval-augmented post-layout optimization, and netlist-tomography-based soft-guide generation across P&R stages.

16:45
536: TransPPA: A Transformer-Based Approach for Multi-Corner PPA Modeling of Microprocessors
Tong Liu, Yan Tan, Zijun Jiang, Yangdi Lyu
The Hong Kong University of Science and Technology (Guangzhou)

17:00
627: Macro-Parameter Identification and Diversified Bayesian Optimization for EDA Parameter Tuning
Zhengfei Chen{1}, Hang Liu{1}, Shizhang Wang{1}, Ruidong Li{2}, Chengjun Guo{1}, Hao Geng{3}, Qi Sun{1}, Cheng Zhuo{1}
{1}Zhejiang University; {2}Shandong Yunhai Guochuang Innovative Technology Co., Ltd.; {3}ShanghaiTech University

17:15
15: RAL-Opt: Overcoming Parametric Limitations in Post-Layout Optimization via Retrieval-Augmented Learning
Jiajie Xu{1}, Songlin Li{1}, Leilei Jin{2}, Zhaohan Wang{1}, Xiangfei Hu{3}, Hao Yan{1}, Longxing Shi{1}
{1}Southeast University; {2}The Chinese University of Hong Kong; {3}SouthEast University

17:30
1696: Netlist Tomography-Based Soft Guide Generation with Refinement Across P&R Stages for Improved PPA
Andrew B. Kahng, Sayak Kundu, Yiting Liu, Dooseok Yoon
University of California San Diego

16:45 - 17:45
Chasing Speed: Accelerating AI and Beyond
Room: Blossom Hill III
Session Chair(s): Pierre-Emmanuel Gaillardon, University of Utah; Vivek Tiwari, Synopsys

This session features FPGA and heterogeneous computing solutions for accelerating modern AI and data-intensive applications. Topics include CPU-FPGA collaborative computing, sparse and quantized LLM inference, adaptive accelerator runtimes, NPU programming, and domain-specific acceleration techniques that push the boundaries of performance and efficiency for next-generation intelligent workloads.

16:45
1272: FMiner: CPU-FPGA Heterogeneous Framework for Frequent Itemset Mining and Beyond
Esmerald Aliaj{1}, Se-Min Lim{2}, Sang-Woo Jun{3}, Se-Min Lim{2}
{1}University of California, Irvine; {2}Kookmin University; {3}University of California Irvine

17:00
363: FAST-MoE: FPGA Acceleration of Sparsity Tailored MoE-LLM with Adaptive N:4 Sparsity-and-Address Scheduling
Young Chan Kim, Nam Joon Kim, Wonjun Hwang, Hyun Kim
Seoul National University of Science and Technology

17:15
458: MOSAIC: A Runtime-Adaptive Sparse Attention Accelerator for Long-Context LLM Inference on FPGA
Zhiyi Chen, Guoyu Li, Zheng Guan, Haodong Lu, Kun Wang
Fudan University

17:30
530: TileFuse: A Fused Mixed-Precision Kernel Library for Efficient Quantized LLM Inference on AMD NPUs
Wesley Pang, Gregory Hyegang Jun, Feiyang Liu, Deming Chen
University of Illinois Urbana-Champaign

16:45 - 17:45
Constraint-Aware Optimization for FPGA and PCB Design Automation
Room: San Jose Ballroom
Session Chair(s): Tianshu Hou, The Chinese University of Hong Kong; Hailiang Hu, AMD

This session highlights constraint-aware physical design for FPGA and PCB systems, covering timing-driven systolic array placement and irregular board-boundary-aware PCB migration. The papers demonstrate graph-based modeling and geometry-aware optimization for improving implementation quality, adaptability, and design automation efficiency.

16:45
237: GRASP: A Graph-Based Approach to Timing-Driven Systolic Array Placement on FPGAs
Donghao Fang{1}, Hailiang Hu{1}, Wuxi Li{2}, Giacomo Coraluppi{1}, Runzhi Wang{1}, Jiang Hu{1}
{1}Texas A&M University; {2}AMD

17:00
724: Ignition: Irregular Board Boundary Aware PCB Migration
Yapeng LI, Beichen Li, Pengjia Li, Tinghuan Chen
The Chinese University of Hongkong, Shenzhen

16:45 - 17:45
From Perception to Action: Efficient and Reliable Physical AI at the Edge
Room: San Carlos
Session Chair(s): Subham Sett, Cadence Design Systems; Zhiyu Zeng, Cadence Design Systems

Embodied AI is outgrowing the hardware that must run it in real time. This session confronts the gap across perception, decision, and action: automating mixed-precision accelerator design, accelerating diffusion action experts, and monitoring perception backbones for silent corruption. Together, they show how co-design delivers efficient, trustworthy Physical AI.

16:45
604: PRISM: Few-Shot Runtime Integrity Monitoring of Learned Perception via Corruption Subspace Families
Ismail Mohamed, Mohamed Ibrahim
The University of Texas at Dallas

17:00
656: S^{2}AE: Exploiting Action Similarity and Layer Significance to Accelerate Diffusion Transformer Action Experts
Longke Yan, Xin Zhao, Zhuoyang Yan, Jiancong Li, Fengbin Tu
The Hong Kong University of Science and Technology

17:15
686: APEX-RBD: Mixed-Precision Exploration Framework for Hardware-Efficient Robot Dynamics Accelerator Design
Xingyu LIU{1}, Hanwei Fan{2}, Chaofang Ma{2}, Jiawei Liang{2}, Guangyu Hu{2}, Jiang Xu{3}, Wei Zhang{1}
{1}Hong Kong University of Science and Technology; {2}The Hong Kong University of Science and Technology; {3}Hong Kong University of Science and Technology (Guangzhou)

16:45 - 17:45
Ising Machines, Photonic, Hyperdimensional Memory
Room: Guadalupe
Session Chair(s): Hai Li, Duke University

Emerging computing paradigms promise faster search and more efficient intelligence, but their gains depend on navigating real hardware constraints. This session explores FPGA–ASIC orchestration and structure-aware SAT mapping for capacity-limited Ising chips, noise-resilient on-chip fine-tuning in photonic vision accelerators, and structured hyperdimensional memory for spatio-temporal reasoning. Together, the papers show how algorithm–hardware co-design makes these approaches more practical.

16:45
1029: An FPGA-ASIC Co-Design Framework for Capacity-Constrained Physics-based Ising Chips
Ruihong Yin, Yue Zheng, Chaohui Li, Ahmet Efe, Abhimanyu Kumar, Ziqing Zeng, Ulya R. Karpuzcu, Sachin S. Sapatnekar, Chris H. Kim, Ahmet Efe
University of Minnesota

17:00
1795: On the Limits of Ising Machines for Circuit-Derived SAT
Ahmet Efe{1}, Abhimanyu Kumar{1}, Nafisa S. Prova{1}, Hüsrev Cılasun{1}, Ziqing Zeng{1}, Tahmida Islam{1}, Ruihong Yin{2}, Chaohui Li{1}, Peter Kreye{1}, Chris H. Kim{1}, Sachin S. Sapatnekar{1}, Ulya R. Karpuzcu{1}
{1}University of Minnesota; {2}University of Minnesota Twin Cities

17:15
1012: Opto-ViT-v2: Noise-Resilient On-Chip Fine-Tuning for Photonic Near-Sensor Vision Transformer Accelerators
Xuming Chen{1}, Deniz Najafi{2}, Mehrdad Morsali{2}, Chengwei Zhou{1}, Zahra Ghanaatian{3}, Mahdi Nikdast{3}, Shaahin Angizi{2}, Gourav Datta{1}
{1}Case Western Reserve University; {2}New Jersey Institute of Technology; {3}Colorado State University

17:30
601: Vector Field Transforms: Structured Hyperdimensional Memory for Spatio-Temporal Reasoning
William Chung{1}, Andrew Ding{1}, Hansen Jin Lillemark{2}, Mohsen Imani{1}
{1}University of California Irvine; {2}University of California San Diego

16:45 - 17:45
Meeting the Deadline: Timing Guarantees for AI on Heterogeneous Platforms
Room: Almaden
Session Chair(s): Caiwen Ding, University of Minnesota; Hongce Zhang, Hong Kong University of Science and Technology (Guangzhou)

This session addresses real-time guarantees for AI on heterogeneous platforms: pre-assembly worst-case latency prediction for edge systems, precise GPU L1 data cache contention analysis for tight WCET estimation, hard and soft real-time task support in batched LLM serving, and priority-aware object detection scheduling across multi-accelerator SoCs, enabling dependable autonomous systems.

16:45
228: PARTs: Pre-Assembly Algorithm-to-Hardware Real-Time Orchestration for Edge Systems
Yinchen Ni, Jintao Chen, An Zou
Shanghai Jiao Tong University

17:00
894: Precise Contention Analysis for GPU L1 Data Cache
Shuai Zhao{1}, Yaowei Liang{1}, Chen Jie{1}, Yinjie Fang{2}, Mingyue Cui{1}, Lei Yang{1}, Wanli Chang{2}
{1}Sun Yat-sen University; {2}Hunan University

17:15
1372: CoRT: Supporting Hard and Soft Real-Time Tasks in Batched LLM Serving
Yeongwoo Ha{1}, Taemin Kim{1}, Changhun Han{1}, Chan Heo{1}, Jaeheon Kwak{2}, Hyosu Kim{3}, Kilho Lee{4}, Hoon Sung Chwa{5}, Sangeun Oh{1}
{1}Korea University, Seoul, Korea, Republic of; {2}Ajou University, Suwon, Korea, Republic of; {3}School of Computer Science and Engineering, Chung-Ang University; {4}Soongsil University; {5}Daegu Gyeongbuk Institute of Science & Technology, DGIST, South Korea

17:30
1541: DRIFT: Dynamic Intra-Frame Parallelization for Fast Continuous Object Detection on Multi-Accelerator SoCs
Justin Davis, Ismet Dagli, Mehmet Belviranli
Colorado School of Mines

17:45 – 19:15
Job Fair
Room: Ballroom Foyer

8:00 – 8:30
Registration
Room: Ballroom Foyer

8:30 – 9:30
Keynote: Design for AI, AI for Design
Paul Cunningham, System Verification Group, Cadence
Room: San Jose Ballroom

9:30 – 10:00
Coffee Break | Exhibits
Room: Ballroom Foyer

10:00 – 17:45
The 8th Workshop on Top Picks in Hardware and Embedded Security
Room: San Carlos

10:00 - 11:30
Emerging Design Frontiers: Quantum Circuits, Photonics, and 3D Systems
Room: Guadalupe
Session Chair(s): Liang Zhang, Beihang University

Quantum circuits, 3D memory, photonic placement, chiplets, and advanced-package power delivery.

10:00
851: quSAS: Structure-Aware Quantum Circuit Simulation with 1.58-bit Representation and Sparsity-based Computation
Hyoju Seo{1}, Hyunjoon Shin{1}, Seokhyeon Lee{1}, Myeongjin Kwak{2}, Yongtae Kim{1}
{1}Kyungpook National University; {2}Kyungpook national university

10:15
997: Quantifying Carbon Footprint of Next Directions in 3D NAND Flash and DRAM Fabrication
Danielle Grey-Stewart{1}, Mariam Elgamal{1}, David Kong{1}, Zhenyang Chen{2}, Zhengyin Piao{2}, Yuan Yao{2}, Gage Hills{1}
{1}Harvard University; {2}Yale University

10:30
1110: SQD-Enabled Circuit Compression for Resource-Efficient Quantum Chemistry
Kangyu Zheng{1}, Yidong Zhou{2}, Jinglei Cheng{3}, Zhemin Zhang{1}, Shaohua Li{1}, Zhiding Liang{1}, Kangyu Zheng{1}, Zhiding Liang{4}
{1}The Chinese University of Hong Kong; {2}Rutgers University; {3}EigenAI; {4}University of Notre Dame

10:45
1221: Routability-aware Force-directed Placement for PICs with Orientation Optimization and Design-Rule Awareness
Yen-Cheng Chang{1}, Minh Anh Phan{1}, Zhidan Zheng{2}, Shao-Yun Fang{1}, Tsun-Ming Tseng{3}, Ulf Schlichtmann{3}
{1}National Taiwan University of Science and Technology; {2}TUMCREATE Ltd.; {3}Technical University of Munich

11:00
1369: HMCM: Hierarchical Multi-Chiplet Mapping for LLM Inference on Heterogeneous Computing-in-Memory Systems
Lixia Han, Zike Xu, Zengguang Zou, Zihang Chen, Hui Chen, Ke Chen, Yijun Cui, Weiqiang Liu
Nanjing University of Aeronautics and Astronautics

11:15
1153: Vertical PDN Routing and Embedded IVR Placement Co-Optimization for Advanced Packaging
Zhen Zhuang{1}, Siyuan Liang{1}, Yuan Du{2}, Li Du{2}, Bei Yu{1}
{1}The Chinese University of Hong Kong; {2}Nanjing University

10:00 - 11:30
Machine Learning for Circuit, Device, and Technology Modeling
Room: Blossom Hill III
Session Chair(s): Albert Zeng, Cadence Design System; Ivan Kissiov, Siemens

Machine learning is transforming analog design and semiconductor modeling through physics-aware optimization and efficient surrogate learning. This session features advances in transistor sizing, circuit representation learning, TCAD acceleration, graph neural networks, and multimodal optimization, demonstrating how AI accelerates design-space exploration while preserving physical accuracy and predictive fidelity.

10:00
538: ATLAS: Adaptive TDA-guided Landscape-Aware Transistor Sizing
Youngmin Oh{1}, Jihwan Won{2}, Yuntae Park{1}, Bosun Hwang{1}, Suwan Kim{3}, Suwan Kim{3}
{1}InfiniTree; {2}Kwangwoon University; {3}Kyung Hee University

10:15
326: Alignment Unlocks Complementarity: A Framework for Multiview Circuit Representation Learning
Zhengyuan Shi{1}, Jingxin Wang{2}, Wentao Jiang{3}, Chengyu Ma{3}, Ziyang Zheng{1}, Zhufei Chu{3}, Weikang Qian{2}, Qiang Xu{1}
{1}The Chinese University of Hong Kong; {2}Shanghai Jiao Tong University; {3}Ningbo University

10:30
1388: Achieving Precise Multi-Scale Semiconductor Modeling via a Physics-Cognizant Neural Network Framework
Tianrun Yang{1}, Hanshi Yang{1}, Yadong Wei{2}, Jizhou Luo{1}, Jianqun Yang{2}, Xingji Li{2}, Ying Ma{1}
{1}Faculty of Computing,Harbin Institute of Technology; {2}Technology Innovation Center of Materials and Devices at Extreme Environment,Harbin Institute of Technology

10:45
1759: Mesh-Native Physics-Informed Graph Surrogates for TCAD-in-the-Loop Design Space Exploration
Leonid Popryho{1}, Ayoub Sadeghi{1}, Inna Partin-Vaisband{2}
{1}University of Illinois Chicago; {2}University of Illinois, Chicago

11:00
1873: GNN-NPN: Efficient NPN Classification for Large Functions using Graph Neural Network
Chengdi Cao, Hanyu Wang, Wan-Hsuan Lin, Jason Cong
University of California, Los Angeles

11:15
55: Multimodal Prior-guided Zero-shot Design Constraint Tuning under Domain Shifts
Jiayi Tu{1}, Baofu Han{2}, Yi Zhang{1}, Yuxuan Cai{1}, Meng Zhang{1}, Tinghuan Chen{3}
{1}Southeast University; {2}Southeast University & National University of Singapore; {3}The Chinese University of Hong Kong, Shenzhen

10:00 - 11:30
Multiphysics Electromigration and Reliability Analysis for Advanced Interconnects
Room: Blossom Hill I/II
Session Chair(s): Siddhartha Nath, Google Deepmind

Advanced interconnects at scaled nodes require multiphysics electromigration and reliability analysis spanning stress evolution, electro-thermomigration coupling, and full-chip signoff. This session presents analytical and learning-based methods for electromigration stress analysis, along with design-rule-aware post-route repair. It also explores high-sigma yield in 3D NAND and SRAM through risk-aware sampling and failure-center localization. Together, the papers address how to predict, identify, and mitigate failures across chip designs.

10:00
319: A Unified Eigenfunction-Based Solver for Electromigration Stress Evolution Across Nucleation and Post-Voiding Phases
Yonglong Cheng, Yici Cai, Qiang Zhou
Department of Computer Science and Technology, Tsinghua University, Beijing, China

10:15
1297: Fast Electromigration Stress Evolution Analysis for General Interconnect Trees Using Matrix-Exponential Method
Wenjie Zhu{1}, Zixuan Meng{1}, Xunsheng Tan{1}, Weiguang Sheng{1}, Pengpeng Ren{1}, Zhigang Ji{1}, Min Tang{1}, Yun Liang{2}, Hai-Bao Chen{1}
{1}Shanghai Jiao Tong University; {2}Peking University

10:30
825: Full-Chip Multiphysics-Aware Electromigration Analysis Methodology Using a Dynamic-Transformer Graph Attention Network
Binyu Yin{1}, Linlin Cai{1}, Yunlong Zhang{2}, Yaojie Mo{2}, Weizhi Li{3}, Wangyong Chen{1}
{1}Sun Yat-sen University; {2}Sun Yet-sen University; {3}Sun Yat-Sen University

10:45
408: A Design Rule-Aware Post-Route ECO Framework for Fixing Electromigration Violations
Bo-Ying Wang{1}, Yong-Fong Chang{2}, Hsun-Wei Pao{2}, Yung-Chih Chen{3}, Tien-Hsin Peng{2}, Yu-Ting Hung{2}, Tsai-Chieh Cheng{1}, Yi-Ting Li{1}, Wuqian Tang{1}, Chun-Yao Wang{1}, Shih-Chieh Chang{1}
{1}National Tsing Hua University; {2}Mediatek Inc., Taiwan; {3}National Taiwan University of Science and Technology

11:00
1571: HOLMES: In-Context Failure-Center Localization for High-Dimensional Yield Estimation
Wei Xing{1}, Xixi Zhou{2}, Kaiqi Huang{2}, Jiaye Pan{2}, Hong Qiu{2}, Xin Wang{2}, Shan Shen{3}
{1}University of Sheffield; {2}Shenzhen University; {3}Nanjing University of Science and Technology

11:15
764: RA^2L-NAND: Risk-Aware Active Learning for High-Sigma Exploration in 3D NAND
Hao Chen{1}, Junlang Huang{2}, Zibo Sha{1}, Zhenhua Wu{1}, Shuo Zhang{1}, Zhou Jin{1}, Cheng Zhuo{1}
{1}Zhejiang University; {2}Carnegie Mellon University

10:00 - 11:30
Powering the Third Dimension: Co-Optimizing PDNs, Signal Integrity, and Thermal Performance
Room: San Jose Ballroom
Session Chair(s): Jiang Hu, Texas A&M University; Sandra Shaji, Samsung Semiconductor

This session focuses on physical design optimization for advanced packaging and 3D integration. Topics include power delivery network design, signal integrity, backside PDN, voltage regulator planning, and routing optimization, emphasizing differentiable optimization and AI-driven techniques for jointly improving power, thermal, timing, and manufacturing objectives.

10:00
951: Signal Integrity and Yield Aware Advanced Packaging by Differentiable Post-Layout Gridless Optimization
Hao-Ju Chang{1}, Yi Kuang Shih{1}, Hui-Tzu TING{2}, Wen-Hao Liu{3}, Hung-Ming Chen{2}, Chien-Nan Jimmy Liu{1}, Wen-Hao Liu{3}
{1}National Yang Ming Chiao Tung University; {2}Institute of Electronics, National Yang Ming Chiao Tung University; {3}Nvidia

10:15
275: CNE-PDN: Joint Optimization of PDN Geometry and TSV Layout via Component-Level Neural Parasitic Extraction for 2.5D IC
Chengkai Zhu, Jinhong Huang, Boxiang Song, Xihan Yang, Siwen Wang, Yiyang Zhao, Keren Zhu, Changhao Yan
Fudan University

10:30
197: Routing-Driven Differentiable Co-Optimization of Power and Thermal Integrity for Large Scale Glass Substrate
Min Gyu Park{1}, Zheng Yang{2}, Sung Kyu Lim{1}
{1}University of Southern California; {2}Georgia Institute of Technology

10:45
153: CSCO: A Backside-PDN-Aware Clock–Signal Co-Optimization Framework for Improved PPA
Zixiao Wang{1}, Leilei Jin{1}, Zhen Zhuang{1}, Rongmei Chen{2}, Bei Yu{1}
{1}The Chinese University of Hong Kong; {2}Peking University

11:00
361: V^2Plan: Voltage Regulator and Vertical Interconnect Planning for Power and Thermal Integrity in 3D Integration
Siyuan Miao{1}, Lingkang Zhu{2}, Xiangqiao Meng{2}, Yumeng Liu{3}, Min Gao{4}, Rumin Zhang{5}, Chen Wu{6}, Lei He{1}
{1}University of California Los Angeles; {2}Department of Electrical and Electronic Engineering The Hong Kong Polytechnic University, Hong Kong Eastern Institute of Technology, Ningbo, China; {3}BTD Technology Inc.; {4}Google, Mountain View, CA, USA; {5}Engineering Research Center of Chiplet Design and Manufacturing of Zhejiang Province, \Ningbo Institute of Digital Twin; {6}Ningbo Institute of Digital Twin, Eastern Institute of Technology, and the Engineering Research Center of Chiplet Design and Manufacturing of Zhejiang Province, Ningbo, China.

11:15
491: Sizing and Placement of Power-Gating Headers in Backside PDN Architectures
Seohyun Kim, Shilong Zhang, Youngjin Chae, Youngsoo Shin
KAIST

10:00 - 11:30
Squeezing the Void: Hardware-Algorithm Co-Design for Attention and Sparsity
Room: Willow Glen III
Session Chair(s): Ing-Chao Lin, National Yang Ming Chiao Tung University; Yufei Ma, Peking University

Sparsity is the key to scaling context windows. This session tackles dual-sparse structures, bit-serial processing, and head-specific attention mapping. Learn how hardware-algorithm co-design dynamically prunes networks and leverages distributed architectures, transforming theoretical sparsity into tangible throughput gains for massive-scale inference.

10:00
22: Celty: SpMSpV GPU Kernel and SIMT Co-Design for Efficient Dual-Sparse LLM Inference
Ruokai Yin{1}, Priyadarshini Panda{2}
{1}Yale University; {2}University of Southern California

10:15
301: CuNeS: Accelerating Dynamic Neuron-Sparse LLM Inference with Hybrid GPU Cores
Zihan Wang, Yuanshuo Wang, Shangzhi Shen, Cheng Tang, Jiajun Ji, Lei Gong, Teng Wang, Wenqi Lou, Chao Wang, Xuehai Zhou
University of Science and Technology of China

10:30
434: Mosaic: Fast Sparse Matrix Factorization with Parallel Tiling-Merging Computation
Jaeung Ryu, Seongmin Ki, Jaeha Park, Sangkyu Jeon, Sungju Ryu
Sogang University

10:45
852: L2C2: Enhancing Large-scale Chiplets for Agent Long-Context LLM Inference with Sparse and Distributed Attention
Kangbo Bai, Zhonghua Mu, Hongze Zhou, Tianyu Jia
Peking University

11:00
884: HASA: Head-specific Attention Sparsity Acceleration for High-resolution Next-scale Visual Generation
Keyi Li, Ziao Xiao, Zhiyuan Chen, Yufei Ma, Ziao Xiao
Peking University

11:15
1763: BRIM: Workload-Balanced Dual-Sided Bit-Serial Sparse Inference Accelerator
Varun Manjunath{1}, Ruokai Yin{2}, Donghyun Lee{1}, Arkapravo Ghosh{1}, Priyadarshini Panda{1}
{1}University of Southern California; {2}Yale University

11:30 – 13:00
Lunch
Room: San Jose Ballroom

13:00 - 14:30
Circuits and Accelerators for Secure and Private Computing
Room: Almaden
Session Chair(s): Zain Ul Abideen, University of Idaho

Ingenious hardware design is crucial for efficient and secure realization of various mathematical algorithms that ensure data confidentiality, integrity and privacy. This track presents new circuit and accelerator schemes that promise to improve current state-of-the-art for various algorithms, ranging from fundamental multipliers to more advanced fully homomorphic encryption.

13:00
430: HyFI: A Bitwidth-Flexible Integer Multiplier Using Hybrid Decomposition
Xiaoyi Ye{1}, Cheng Wang{1}, Mingyu Gao{2}
{1}Xi'an Jiaotong University; {2}Tsinghua University

13:15
248: FASTAR: FRI Accelerator for Scalable Transparent ARguments of Knowledge
Tengkai Gong, Xiaolin Xu, Xiaolin Xiao
Northeastern University

13:30
178: FLIGHT: Fine-Grained High-Throughput Hardware Accelerator for Fully Homomorphic Encryption
Kaiyuan Zhang{1}, Keshab K. Parhi{2}, Yingjie Lao{1}
{1}Tufts University; {2}University of Minnesota

13:45
327: Cirrus: Efficient FHE Acceleration via Restructured Parallel Half-Bootstrapping on Switch-Agile Arrays
Lan Yang{1}, Changxu Liu{1}, Baoyu Fan{2}, Fan Yang{1}
{1}Fudan University; {2}IEIT Systems Co Ltd

14:00
200: Apollo: Differentiated Bootstrapping Insertion for General Purpose FHE
Jinkai Zhang, Yinghao Yang, Yuehao Xu, Hongyan Li, Xiaowei Li, Hang Lu
State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences

14:15
540: SuperNet: A Scalable Unified Permutation Network for High-Parallelism FHE Accelerators on Multi-Die FPGA
Kai Li{1}, Lutan Zhao{2}, Qingyun Niu{1}, Zhengbang Yang{1}, Lu Gao{1}, Rui Hou{2}
{1}State Key Laboratory of Cyberspace Security Defense, Institute of Information Engineering, CAS and School of Cyber Security, University of Chinese Academy of Sciences; {2}State Key Laboratory of Cyberspace Security Defense, Institute of Information Engineering, CAS

13:00 - 14:30
Efficient AI Models and Learning Algorithms for EDA
Room: Blossom Hill III
Session Chair(s): Rongjian Liang, Nvidia Research; Zhuo Feng, Stevens Institute of Technology

As AI becomes central to electronic design automation, improving model efficiency and scalability is essential for practical deployment. This session highlights advances in efficient LLM inference, memory management, tensor representations, hardware-aware optimization, intelligent control, and AI infrastructure, enabling more computationally efficient, scalable, and effective AI-assisted design automation.

13:00
1154: MeltKV: Memory-Aligned KV Cache Compression via Variable-Bit Packed Vector Quantization for Long-Context LLMs
Jiyun Han{1}, Seongjoon Cho{1}, Seungkyu Choi{2}
{1}Kyung Hee University; {2}Yonsei University

13:15
974: Cluster-of-Thought: Semantic-Structured KV-Cache Management for Long-Context Reasoning Models
Yangbo Wei{1}, Qifan Wang{1}, Junhong Qian{2}, Shaoqiang Lu{1}, Zhenyu Su{1}, Yiheng Jiao{3}, Chen Wu{4}, Lei He{5}
{1}Shanghai Jiao Tong University, Shanghai, Shanghai, China; Eastern Institute of Technology, Ningbo, China; {2}Shanghai Jiao Tong University, Shanghai, Shanghai, China;; {3}Fudan University, Shanghai, China; {4}Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, China; {5}Eastern Institute of Technology, Ningbo, China

13:30
879: FlowTT: Exploiting Computation Flow Reuse in Irregular Tensor-Train Embedding
Jongmin Seok, Chae Eun Rhee
Hanyang University

13:45
372: Tri-Balancing Act: Noise-Guided Mixed-Precision Search for Accuracy–Energy–Resilience Co-Optimization
Ruge Xu{1}, Yi Ren{2}, Yushan Pan{3}, Weikang Qian{1}, Xinfei Guo{1}
{1}Shanghai Jiao Tong University; {2}Shanghai JiaoTong University; {3}Xi'an Jiaotong-Liverpool University

14:00
524: MetaField: Metaplastic Dynamic Neural Fields with Confidence-Driven MPC Arbitration for Efficient Edge Control
Patrick Sigler, Hammad Omar, Mohamed Ibrahim
The University of Texas at Dallas

14:15
1624: HiddenRAG: Minimizing Retrieval-Augmented Generation Latency via In-Storage Processing
Zhongzhi Yu{1}, Yujie Zhao{1}, Sixu Li{1}, Yujie Wang{2}, Chaojian Li{1}, Amrit Khera{1}, Shruti Belgali{1}, Yang (Katie) Zhao{2}, Yingyan (Celine) Lin{3}, Zhongzhi Yu{4}, Chaojian Li{5}
{1}Georgia Institute of Technology; {2}University of Minnesota, Twin Cities; {3}Georgia Tech; {4}Nvidia; {5}Hong Kong University of Science and Technology

13:00 - 14:30
Emerging System-Level Design Methodologies
Room: Blossom Hill I/II
Session Chair(s): Meng Li, Peking University

The increasing complexity of modern computing platforms demands new approaches to architecture design, evaluation, and optimization. This session showcases research on system-level methodologies that improve the efficiency, scalability, and specialization of emerging computing systems, highlighting the growing role of cross-layer design and intelligent automation in advancing future computer architectures.

13:00
82: OpenIsingHW: An Emulator Enabling the System-level Performance Analysis of Digital Ising Accelerators
Jiacong Sun{1}, Sofie De Weer{1}, Yingna Huang{2}, Toon Bettens{1}, Jerald Yoo{3}, Wim Dehaene{1}, Giuseppe Maria Sarda{1}, Marian Verhelst{1}
{1}KU Leuven; {2}National University of Singapore; {3}Seoul National University

13:15
1315: PackStress: Physics-Aware Learning for Full-chip Thermal Stress Modeling in 3D Packaging
Tao Xu, Zhenhua Wu, Zhou Jin, Cheng Zhuo
Zhejiang University

13:30
166: Chipstitch: A Scalable and Cost-Efficient Platform for Accelerated Silicon Prototyping
Jeongeun Kim, Sabrina Yarzada, Paul Chen, Yuanhan Li, Wenbo Zhu, Christopher Torng
University of Southern California

13:45
505: Aquas: Enhancing Domain Specialization through Holistic Hardware-Software Co-Optimization based on MLIR
Yuyang Zou, Youwei Xiao, Chenyun Yin, Yansong Xu, Yuhao Luo, Yitian Sun, Ruifan Xu, Renze Chen, Yun Liang
Peking University

14:00
1097: VESICA: Fused Vector Extension for RISC-V Embedded GPUs
Dao Xian Ding, Michael Srouji, Brandon Kam, Nader Sehatbakhsh, Blaise Tine
UCLA

14:15
820: Mapping Without Graphs: Learning Coherence Traffic for Task Placement
Guochu Xiong, Tianrui Ma, Weichen Liu
Nanyang Technological University

13:00 - 14:30
Event-Driven Intelligence: Neuromorphic Architectures, SNN Compilation, and Energy-Efficient Edge Learning
Room: Guadalupe
Session Chair(s): Ismail Bustany, AMD

This session presents co-designed neuromorphic AI advances across the stack: INT4 local online SNN training, formally bounded spike-propagation compilation, precision-reconfigurable SOT-MRAM compute-in-memory, sparsity-adaptive CPU execution, noise-robust distributed spike encoding for sensing, and hybrid SNN-HDC acceleration for few-shot continual edge learning with improved efficiency.

13:00
296: Lonic: Algorithm-Hardware Co-Design for Energy-Efficient Fully Local Online SNN Training with INT4 Precision
Peilin Chen, Xiaoxuan Yang
University of Virginia

13:15
1739: Formal Worst-Case Spike-Propagation Latency Bounds for Neuromorphic SNN Compilation
Vikash Kumar, Vikash Kumar
Arm

13:30
1923: An Energy-Efficient SOT-MRAM Compute-in-Memory Engine With Spike Encoding and Dynamic Precision Modulation
Deyang Yu, Chenchen Liu, Chuanjie Zhang, Xiao Fang, Weijian Qi, Yutong Tang, Weisheng Zhao
Beihang University

13:45
1523: SASA: Sparsity-Adaptive Framework for Spiking Neural Network Acceleration on CPU
Kang You{1}, Binglun Cai{1}, Ziling Wei{1}, Lee Jun Yan{1}, Chen Nie{1}, Ran Wang{1}, Jiaxuan Zhang{1}, He Zhu{1}, Yiyao Yang{1}, Zekai Xu{1}, Cheng Zou{1}, Shuo Yang{1}, Yu Zou{2}, Yu Feng{1}, Haojin Zhu{1}, Zhezhi He{3}
{1}Shanghai Jiao Tong University; {2}Institute of Information Engineering, Chinese Academy of Sciences; {3}School of Computer Science, Shanghai Jiao Tong University; Shanghai AI Laboratory

14:00
1790: Spikes as Needed: Energy-Efficient and Noise-Robust Distributed Neuromorphic Encoding
Yuchen Yuan{1}, Alberta Dadeboe{2}, Ramashish Gaurav{2}, Youzuo Lin{3}, Yang (Cindy) Yi{2}, Lei Yang{1}
{1}George Mason University; {2}Virginia Tech; {3}University of North Carolina at Chapel Hill

14:15
725: SynapseHD: A Hybrid SNN-HDC Accelerator for On-Chip Few-Shot Continual Learning via Algorithm-Hardware Co-Design
Lingfeng Zhou{1}, Guangyao Lin{2}, Huiyao Wang{2}, Bohan Wang{2}, Zhiyi Yu{2}, Shanlin Xiao{2}
{1}Sun Yat-Sen University; {2}Sun Yat-sen University

13:00 - 14:30
GenAI for Chip Design Across the Lifecycle: Design, Optimization, Security, and Reliability
Room: Willow Glen I/II
Session Chair(s): Christian Pilato, Politecnico di Milano

AI-driven methodologies are rapidly transforming the design of next-generation integrated circuits, enabling new levels of automation, adaptability, and intelligence across the chip design lifecycle. As this paradigm accelerates, it exposes complex challenges in chip design spanning design, optimization, security, and reliability that AI-driven methodologies aim to address. Addressing these challenges is essential to unlock the full potential of scalable and robust design flows for emerging domains such as AI acceleration, large-scale inference, and high-performance computing. In this proposed special session, we bring together leading experts to explore cutting-edge advances in self-evolving optimization agents, automated physical design frameworks, AI-guided hardware security testing, and cross-stage reasoning for silicon quality improvement. The session aims to synthesize insights from both academia and industry, highlight opportunities for end-to-end AI integration across the design stack, and chart a forward-looking roadmap for building intelligent, reliable, and high-performance chip design systems at scale.

1977: Invited Paper: ChipMind II — A Self-Improving LLM-Driven Agentic Framework for Chip Design
Fan Wu{1}, Jonti Talukdar{2}, Peter Domanski{3}, Bahar Farahani{4}, Agastya Seth{3}, Sri Sai Rakesh Nakkilla{3}, Pragya Sharma{3}, Chenghao Fu{4}, Sanmitra Banerjee{2}, Krishnendu Chakrabarty{3}, Farshad Firouzi{1}, Farshad Firouzi{3}, Farshad Firo{1}
{1}Johns Hopkins University; {2}NVIDIA; {3}Arizona State University; {4}IEEE Member

1979: Invited Paper: AutoPD: Question-Driven MCTS from Paper Artifacts to PPA-Adaptive Physical-Design Algorithms
Chia-Tung Ho{1}, Haoyu Yang{1}, Rongjian Liang{2}, Chenhui Deng{1}, Cunxi Yu{2}, Brucek Khailany{1}
{1}NVIDIA; {2}Nvidia

1980: AI for Silicon, Silicon for AI: Towards Autonomous Hardware Security at the Edge
Nikhilesh Singh{1}, Mohamadreza Rostami{2}, Mozhgan Navardi{3}, Tinoosh Mohsenin{3}, Ahmad-Reza Sadeghi{4}
{1}TU Darmstadt; {2}Technical University of Darmstadt; {3}Department of Computer and Electrical Engineering - Johns Hopkins University; {4}Technical University Darmstadt

1996: Can Agents Design Better Chips with a Higher Level Abstraction?
Zijian Ding, Yang Zou, Yizhou Sun, Jason Cong
UCLA

13:00 - 14:30
Next-Gen Scale: MoE Serving, Speculative Decoding, and Distributed Memory Hierarchies
Room: Willow Glen III
Session Chair(s): Zihan Wang, University of Science and Technology of China

How do we serve trillion-parameter models? This session addresses the systems-level engineering behind Mixture-of-Experts and distributed LLMs. Explore hybrid sparsity mapping, runtime-reconfigurable speculative decoding, and HBM power reduction strategies that make extreme-scale, multi-node inference practically deployable and energy-efficient.

13:00
1098: Analyzing the impact of emerging technology on AI Datacenter serving distributed LLM inference using ECHELON
Abhimanyu Bambhaniya{1}, Ritik Raj{2}, Geonhwa Jeong{3}, Souvik Kundu{4}, Sudarshan Srinivasan{5}, Midhilesh Elavazhagan{5}, Madhu Kumar{5}, Suvinay Subramanian{6}, Tushar Krishna{2}, Souvik Kundu{5}, Suvinay Subramanian{6}
{1}Georgia Tech; {2}Georgia Institute of Technology; {3}Meta; {4}Intel Labs; {5}Intel; {6}Google

13:15
1164: NMP-hi: An Evaluation and Optimization Framework on Near-Memory-Processing Hierarchies for Large Language Models
Jun Wang, Zeyu Guo, Zhenwei Wang, Zening Niu, Shengzhe Yan, Zhaori Cong, Jinshan Yue
Institute of Microelectronics of the Chinese Academy of Sciences, University of Chinese Academy of Sciences

13:30
1376: SPECTRA: Adaptive Execution of Speculative Decoding on a Runtime-Reconfigurable Tiled Architecture
Gabriele Tombesi, William Baisi, Je Yang, Elisavet Lydia Alvanaki, Kevin Lee, Michael Lippe, Biruk Seyoum, Luca P. Carloni, Je Yang, Michael Lippe
Columbia University

13:45
1566: SPIMOE: Exploiting Hybrid Sparsity for Reasoning MoE Inference on Heterogeneous PIM Architectures
Rubing Yang, Cenlin Duan, Yingjie Qi, Xiaolin He, Xiao Ma, Jianlei Yang
Beihang University

14:00
1866: DARA: A Data-Aware Refreshing Architecture for HBM Power Reduction in LLM Accelerators
Jiajie Fan{1}, Lebei Cui{1}, Anmin Liu{2}, Xinyao Wang{1}, Zhutianya Gao{1}, Chen Zhang{1}
{1}Shanghai Jiao Tong University; {2}Peking University

14:15
1871: Delta-MoE: Rethinking Transfer Granularity for CPU–GPU MoE Serving
Jiangnan Yu{1}, Kunming Shao{1}, Shiwei Liu{2}, Chixiao Chen{3}, Yuan Xie{1}
{1}The Hong Kong University of Science and Technology; {2}Tongji University; {3}Fudan University

14:30 – 15:00
Cofee Break | Exhibits
Room: Ballroom Foyer

15:00 - 16:30
Foundation Models, Benchmarks, and Trustworthy AI for Hardware Design
Room: Blossom Hill III
Session Chair(s): Yu Li, Zhejiang University

The success of AI-assisted hardware design depends on robust benchmarks, high-quality datasets, specialized foundation models, and rigorous evaluation methodologies. This session presents research strengthening trustworthy AI for EDA through multimodal datasets, adversarial robustness analysis, hardware-specific language models, and AI-assisted design-space exploration, establishing the foundation for next-generation intelligent hardware design tools.

15:00
630: HWE-Bench: Benchmarking LLM Agents on Real-World Hardware Bug Repair Tasks
Fan Cui{1}, Hongyuan Hou{2}, Zizhang Luo{3}, Chenyun Yin{3}, Yun Liang{3}
{1}School of Integrated Circuits, Peking University; {2}Peking university; {3}Peking University

15:15
1180: VeriMM: A Large-Scale Multimodal Verilog Dataset with High-quality Testbenches for Hardware Code Generation
Siyu Yun, Qingsong Peng, Yu Li, Cheng Zhuo
Zhejiang University

15:30
1163: VeriObf: An Adversarial Obfuscation Framework for Evaluating LLM Robustness in RTL Code Understanding
Qingsong Peng{1}, Liqi Lin{1}, Yu Li{2}, Qi Sun{2}, Tianming Ni{3}, Cheng Zhuo{2}, Liqi Lin{4}
{1}Zhejiang university; {2}Zhejiang University; {3}Anhui Polytechnic University; {4}zhejiang university

15:45
695: AutoDevice: A Specialized LLM for Semiconductor Device Generation with Code-Chain Reasoning
Yuqi Jiang, Yucheng Cui, Chuyue Liu, Yupeng Hu, Xunzhao Yin, Qi Sun, Cheng Zhuo
Zhejiang University

16:00
807: MicroEvo: Knowledge-Guided LLM Sampling for Efficient Microarchitecture Design Space Exploration
Jia Xiong{1}, Runkai Li{1}, Chenxu Niu{2}, Guangyuan Gao{1}, Changwen Xing{1}, Yifan Zhang{1}, Xinlai Wan{1}, Jieran Cui{1}, Chen BAI{3}, Yusheng Hua{4}, Ying Wang{5}, Ming Ling{6}, Xi Wang{7}, Tao Xie{8}, Chen BAI{9}
{1}Southeast University; {2}NVIDIA Corporation; {3}The Hong Kong University of Science and Technology; {4}Nanjing University of Posts and Telecommunications; {5}Institute of Computing Technology, Chinese Academy of Sciences; {6}National ASIC System Engineering Technology Research Center, Southeast University; {7}National ASIC Center, School of Integrated Circuit, Southeast University; {8}Peking University/Fudan University; {9}Fudan University

16:15
1339: HLS-Seek: QoR-Aware Code Generation for High-Level Synthesis via Proxy Comparative Reward Reinforcement Learning
Qingyun Zou{1}, Feng Yu{1}, Hongshi Tan{1}, Yao Chen{2}, Bingsheng He{1}, Weng-Fai Wong{1}
{1}National University of Singapore; {2}Huazhong university of science and technology

15:00 - 16:30
From Ad Hoc to Standardized: Early Explorations of 3D IC Design Flows Across Integration Paradigms
Room: Willow Glen III
Session Chair(s): Leilei Jin, Chinese University of Hong Kong

Three-dimensional integration and advanced packaging, including monolithic 3D ICs, 2.5D chiplet-based integration, and emerging 5.5D integration with glass interposers, have demonstrated strong potential to sustain system scaling beyond the limits of monolithic design. However, in contrast to the well-established RTL-to-GDS implementation flows for conventional 2D IC design, the lack of standardized implementation flows across these integration paradigms remains a fundamental bottleneck. The absence of unified implementation flows raises a high barrier to both the development and evaluation of new methodologies for 3D IC design. Researchers often rely on ad hoc assumptions, incomplete toolchains, or simplified design enablements, making it difficult to develop solutions that generalize across different integration paradigms and implementation scenarios. Moreover, inconsistent evaluation settings hinder the interpretation of results across different works, limiting the ability to understand design trade-offs and to accumulate progress toward a coherent methodology. This special session presents four complementary efforts that collectively address this challenge by advancing implementation methodologies across diverse 3D integration paradigms: open-source implementation platforms for advanced packaging and 3D ICs; a commercial-grade full-chip methodology for face-to-face 3D ICs; AI-driven multi-physics co-optimization for glass-interposer-based 5.5D systems; and multi-domain design optimization of thermally limited 3D ICs for AI and HPC systems. Together, these works highlight a clear shift from paradigm-specific solutions toward flow-level and cross-domain co-optimization, motivating the development of unified, scalable, and interpretable implementation frameworks for next-generation 3D ICs and advanced packaging. By bringing together representative efforts under shared contexts, the session provides a coherent perspective on the evolving 3D IC design space, helps clarify key challenges and research directions, and facilitates more consistent interpretation of results across integration paradigms.

1985: CU3D: A Comprehensive Timing-driven Design Flow for Face-to-Face-Bonded 3D ICs (Invited)
Haoyang Xu, Leilei Jin, Tsung-Yi Ho, Leilei Jin
The Chinese University of Hong Kong

2010: OpenEYE: A Differentiable Framework for Signal Integrity Optimization in 2.5D IC Interconnects
Youngjin Kim{1}, Min Gyu Park{1}, Majid Ahadi Dolatsara{2}, Rick Rabinovich{2}, Sung Kyu Lim{1}, Youngjin Kim{1}
{1}University of Southern California; {2}Keysight Technologies

2000: Toward Unified Open-source Implementation Platforms for 3D Integration Paradigms
Osama Hammad{1}, Junyeong Jang{2}, Arthur Koucher{1}, Jingzheng Li{3}, Matt Liberty{1}, Cho Moon{1}, Ioannis Savidis{4}, Deepashree Sengupta{1}, Yu Shao{5}, Zhiang Wang{3}, Zehao Yang{6}, Zhiyu Zheng{3}
{1}Precision Innovations; {2}UC San Diego; {3}Fudan University; {4}Drexel University; {5}ShanghaiTech University; {6}Shanghai Jiao Tong University

2002: Invited Paper: Opportunities for Multi-Domain Design Optimization of Thermally-Limited 3DICs for AI & HPC Systems
John Wilson, Walker Turner
NVIDIA

15:00 - 16:30
From Layout Decomposition to Mask Optimization: Advances in Multiple Patterning Lithography
Room: Blossom Hill I/II
Session Chair(s): Zhou Jin, Zhejiang University

From layout decomposition to mask optimization, multiple patterning lithography demands scalable integration from layout engines to computational lithography signoff. This session presents GNN-guided DSA-MP refinement, closed-loop pattern clustering, unsupervised universal MPL decomposition, weakly convex inverse lithography, and frequency-domain and parametric multi-mask source–mask optimization for advanced-node manufacturing.

15:00
828: AICL: Alignment-Driven Iterative Closed-Loop Optimization for Layout Pattern Clustering
Shuo Liu{1}, Yilong Li{1}, Renjie Wang{1}, Rongliang Fu{2}, Chao Wang{1}
{1}Southeast University; {2}The Chinese University of Hong Kong

15:15
1900: UniLD: A Universal Layout Decomposer for Multiple Patterning Lithography via Unsupervised Graph Neural Network
Junqi Wang, Yuxuan Dong, Yang Xu, Haodong Lu, Jianli Chen, Kun Wang
Fudan University

15:30
460: GNN-Guided Local Refinement for Efficient Layout Decomposition with DSA-MP
Yu Feng{1}, Liyi Zhou{1}, Haoyu Wang{1}, Yiming Xu{1}, Hao Geng{2}, Shaozhi Deng{3}, Ningsheng Xu{1}, Shisheng Xiong{1}
{1}Fudan University; {2}ShanghaiTech University; {3}Sun Yat-sen University

15:45
542: Moreau-ILT: Efficient Inverse Lithography Exploiting Weak Convexity
Chaofan Wang, Ziyang Yu, Su Zheng, Zixiao Wang, Yifan Shi, Bei Yu
The Chinese University of Hong Kong

16:00
616: FSMO: Frequency-Domain Source Mask Optimization
Zihao Yue, Xijin Zhao, Chun Zhang
Tsinghua University

16:15
679: Multi-Mask Parametric Source Mask Optimization
Mengjia Dai, Ziyang Yu, Guojin Chen, Yuhao Ji, Bei Yu
The Chinese University of Hong Kong

TENX SEMI

Chip Design Reinvented

TenX Semi is pioneering the path toward true, superhuman chip design quality and productivity

Our tools and agents pair AI with deterministic domain-specific methods across the entire design lifecycle to shorten development timelines while improving design quality

We're hiring.

Our team brings together expertise across semiconductors, chip design, and frontier AI research. We're always looking for exceptional people. Join us in reinventing chip design.

tenxsemi.com

15:00 - 16:30
Generative AI Meets Open-Source EDA: Accelerating Design Productivity and Talent Development
Room: Willow Glen I/II
Session Chair(s): Stelios Diamantidis, Cognichip Inc.

Generative AI is emerging as a major force in electronic design automation, with the potential to reshape not only how hardware is designed and verified, but also how these skills are taught and transferred to future generations of engineers. In digital design, large language models and related techniques are being explored for tasks such as requirements interpretation, RTL generation, code completion, testbench creation, debugging support, documentation, and design-space exploration. Although many of these capabilities are still evolving, their direction is already clear: generative AI is becoming an increasingly relevant layer in modern design workflows that can augment productivity, accelerate iteration, and make advanced design activities more accessible to less experienced users. At the same time, the growth of open-source EDA ecosystems is creating a strong foundation for experimentation, accessibility, and reproducibility. Open-source toolchains, libraries, educational infrastructures, and cloud-accessible environments enable a broader community to engage with realistic chip-design workflows without the barriers traditionally associated with commercial tools. This convergence offers a practical path toward democratizing advanced design capabilities: AI can lower the cognitive barrier to entry, while open infrastructures can reduce the operational barrier. This special session focuses on that convergence. Its central premise is that generative AI and open-source EDA should be considered together rather than as separate trends. Generative AI becomes more impactful when embedded in runnable, transparent, and reusable workflows; open-source EDA becomes more powerful when paired with intelligent assistants that help users navigate complexity, generate artifacts, explain tool outputs, and connect design intent to implementation and verification. The combination affects how quickly designers can prototype ideas, how effectively students can learn advanced concepts, how instructors can assemble modern curricula, and how institutions can expand training and talent development in semiconductor design. The session will examine concrete value across the digital design stack while emphasizing that correctness, semantic faithfulness, synthesizability, verification coverage, robustness, traceability, and explainability remain essential. Overall, it aims to provide a forward-looking and balanced perspective on how the EDA community can shape this transformation responsibly and productively.

15:00
Welcome

15:05
2001: Invited Paper: ODE4EC: An Open EDA Ecosystem for Semiconductor Workforce Development
Christian Pilato{1}, Luis Miguel Pinho{2}, Rene Scholz{3}, Frank K. Gürkaynak{4}, Luca Benini{5}, Stefan Wallentowitz{6}
{1}Politecnico di Milano; {2}INESC TEC; {3}IHP; {4}ETH Zurich; {5}ETH Zurich and University of Bologna; {6}Hochschule München University of Applied Science

15:25
2006: Invited Paper: Teaching Chip Design in the GenAI Era
Akashdeep Saha{1}, Prithwish Basu Roy{2}, Prince Steven Annor{1}, Sayanti Jana{3}, Matthew DeLorenzo{3}, Jeyavijayan Rajendran{3}, Ozgur Sinanoglu{1}, Johann Knechtel{1}, Muhammad Shafique{1}, Ramesh Karri{2}, Akashdeep Saha{4}
{1}NYUAD; {2}New York University Tandon School of Engineering; {3}Texas A&M University; {4}Indian Institute of Technology Kharagpur

15:45
Invited Talk: Why Open-Source EDA Matters for the Future of AI-Driven Chip Design
Ehsan Kamalinejad
Cognichip Co-Founder & CTO

16:05
Invited Talk: From RTL Generation to Verification: Advancing AI for Hardware Design
Prashanth Vijayaraghavan
IBM

15:00 - 16:30
Learning the Heat: AI for Thermal, Timing, and Reliability in Advanced 2.5D/3D Integration
Room: San Jose Ballroom
Session Chair(s): Heechun Park, UNIST; Zhen Zhuang, The Chinese University of Hong Kong

This session presents recent advances in AI-assisted modeling for advanced packaging, including thermal prediction, timing estimation, thermal stress analysis, and warpage modeling. The papers demonstrate how surrogate learning and physics-informed machine learning enable fast, accurate design exploration and reliability analysis for heterogeneous 2.5D/3D systems.

15:00
1032: Structure-Aligned Thermal Surrogate Prediction for Heterogeneous 2.5D/3D Chiplet Packages via Point Cloud
Dekang Zhang{1}, Dan N{2}, Wei Xing{3}, Yichao Cao{4}, Zongyi Li{5}, Zhenya Zhou{6}, Changyin Sun{7}, Zhou Jin{8}
{1}Southeast University; {2}School of Automation, Southeast University, Nanjing, China; {3}University of Sheffield; {4}Zhongnan University; {5}JCET Group Co., Ltd.; {6}Empyrean Technology Co., Ltd.; {7}Anhui University; {8}Zhejiang University

15:15
1367: TEMPO: Topology-Agnostic Thermal Surrogate Learning for Advanced 2.5D/3D Packages with Scarce Training Data
Bingrui Zhang{1}, Shuai Pang{2}, Wei Xing{3}, Yuquan Sun{2}, Jimin Wen{1}, Yuanqing Cheng{2}
{1}Xpeedic Co., Ltd.,; {2}Beihang University; {3}The University of Sheffield

15:30
577: QuantizeHeat: Generalizable Thermal Prediction for 2.5D Chiplets via Thermal-Aware Discrete Latent Space Mapping
Ze Hao{1}, Yu Zhang{1}, Haoxuan Song{2}, Tianyu Liang{3}, Yijun Chen{1}, Kang Zhao{1}, Jianwang Zhai{1}
{1}Beijing University of Posts and Telecommunications; {2}University of the Chinese Academy of Sciences; {3}Harbin Institute of Technology, Shenzhen

15:45
584: GTStress: A Graph-Augmented Neural Framework for Thermal Stress Simulation of Parameterized TSV Arrays with Superposed Physical Priors and Query-Graph Fusion
Zhixuan Dong{1}, Chenlong He{1}, Changhao Yan{1}, Zhaori Bi{1}, Keren Zhu{1}, Bowen Jia{2}, Xuan Zeng{1}
{1}Fudan University; {2}Wuhan University of Technology

16:00
1602: WAFER: Warpage Analysis via Fourier-Enhanced Residual Calibration for Large-Scale 2.5D IC Designs
Pin-Han Chen{1}, Chanhee Yang{2}, Sungchul Joo{3}, Hyungsoo Moon{3}, Che-Ming Chang{1}, Taek-Soo Kim{2}, Sung Kyu Lim{1}
{1}University of Southern California; {2}Korea Advanced Institute of Science and Technology; {3}Absolics Inc.

16:15
1848: Distill3D: Cross-Stage Knowledge Distillation for Early-Stage 3D IC Timing Prediction
Mingjun Wang{1}, Lancheng Zou{1}, Guanqi Li{2}, Feng Gu{1}, Yuntao Lu{1}, Tianmeng Yang{3}, Boyu Han{4}, Jianan Mu{5}, Huawei Li{5}, Sung Kyu Lim{6}, Bei Yu{1}
{1}The Chinese University of Hong Kong; {2}University of Chinese Academy of Sciences; {3}Peking University; {4}Stanford; {5}Institute of Computing Technology, Chinese Academy of Sciences; {6}University of Southern California

15:00 - 16:30
Quantum Circuits Across Platforms: Compilation, Mapping, and Emulation
Room: Guadalupe
Session Chair(s): Ron Duncan, Synopsys

Simulators, Compilers and Design approaches for Quantum circuits.

15:00
205: MERA: A Mid-Circuit-Measurement Error-Aware Compilation Framework for Quantum Circuits
Ming Zhong{1}, Zhemin Zhang{1}, Xiangyu Ren{2}, Chenghong Zhu{3}, Siyuan Niu{4}, Zhiding Liang{1}
{1}The Chinese University of Hong Kong; {2}University of Edinburgh; {3}The Hong Kong University of Science and Technology; {4}University of Central Florida

15:15
207: RZ-Flow: Readout-Zone and Feedforward-Aware Compilation for Dynamic Zoned Neutral-Atom Quantum Computers
Yifeng Peng{1}, Xinyi Li{1}, Yidong Zhou{2}, Siyuan Niu{3}, Hao Wang{1}, Bo Yuan{4}, Ying Wang{1}
{1}Stevens Institute of Technology; {2}Rutgers University; {3}University of Central Florida; {4}Rutgers Univerrsity

15:30
388: A Fine-Grained Optimization Approach for Cluster State Synthesis on Photonic Quantum Devices
Ching-Yao Huang, Wai-Kei Mak
National Tsing Hua University

15:45
345: Does Scaling Up Quantum Circuits Always Help? You May Need FALCON: A Fast Latency-Aware Hypergraph Compiler for Large-Scale Variational Quantum Circuits
Yifeng Peng{1}, Xinyi Li{1}, Yidong Zhou{2}, Samuel Yen-Chi Chen{3}, Huanhsin Tseng{4}, Kuan-Cheng Chen{5}, Siyuan Niu{6}, Zhiding Liang{7}, Ying Wang{1}
{1}Stevens Institute of Technology; {2}Rutgers University; {3}Wells Fargo; {4}Brookhaven National Laboratory; {5}JIJ Inc. / Imperial College London; {6}University of Central Florida; {7}Rensselaer Polytechnic Institute

16:00
888: HSF-S: Speed-Optimized Compilation and Acceleration for Hybrid Schrödinger–Feynman Quantum Circuit Emulation
Sechan Park, Kyeongwon Lee, Mundo Jeong, Chaebin Jung, Woojoo Lee
Chung-Ang University

16:15
1048: LaMDaMS: Locality-aware Mapping and Dependency-aware Measurement Scheduling for Modular qLDPC Architecture
Nazmul Hasan, Muhammad Rashedul Haq Rashed
The University of Texas at Arlington

15:00 - 16:30
Squeezing Intelligence In: Model Compression and Bespoke Hardware for the Extreme Edge
Room: Almaden
Session Chair(s): Xiaoxuan Yang, University of Virginia; Zhiyao Xie, Hong Kong University of Science and Technology

This session presents six approaches to efficient AI on severely constrained platforms: quantization for DSP-less MCUs and CPU-only LLM inference, hierarchy-aware pruning for Vision Transformers, checkpointing for intermittent hyperdimensional computing, bespoke RISC-V core generation for flexible electronics, and a streaming tone-mapping accelerator, showing co-design unlocks extreme-edge efficiency gains.

15:00
270: MCUPACK: Efficient Packing with Data Reuse and Reduced Flush Overhead in DSP-less MCUs
Dahoon Park, Yeeun Hong, Wonseok Heo, Jaeha Kung
Korea University

15:15
1667: PolyQ: Codesigning End-to-End Quantization Framework for Scalable Edge CPU LLM Inference
Hyunwoo Oh{1}, Suyeon Jang{1}, Hanning Chen{1}, KyungIn Nam{2}, Sanggeon Yun{1}, Ryozo Masukawa{1}, Mohsen Imani{2}
{1}University of California, Irvine; {2}University of California Irvine

15:30
367: Hierarchy-Aware Soft Pruning for N:M Sparsification in Vision Transformers
Long Zhu{1}, Jingbo Jiang{2}, Jingyang Zhu{3}, Xizi Chen{1}, Chi-Ying Tsui{2}, Xizi Chen{4}
{1}College of Informatics, Huazhong Agricultural University; {2}The Hong Kong University of Science and Technology; {3}NVIDIA Corporation; {4}Huazhong Agricultural University

15:45
1236: Progress Gambit: Rethinking Checkpointing for Intermittent Hyperdimensional Computing
Ya-Tung Tsai{1}, Wen Sheng Lim{2}, Yu-Cheng Chen{1}, Chia-Heng Tu{1}, Yuan-Hao Chang{2}, Yu-Cheng Chen{1}
{1}National Cheng Kung University; {2}National Taiwan University

16:00
1477: TailoR-V: Automated Bespoke Flexible RISC-V Core Generation for Extreme-Edge IoT applications
Polykarpos Vergos{1}, Georgios Zervakis{2}, MEHDI B TAHOORI{1}
{1}Karlsruhe Institute of Technology; {2}National Technical University of Athens

16:15
1591: CEGF-TM: A Cost-Effective Streaming Tone-Mapping Accelerator Using Guided Frame Statistics for Thermal Imaging
Cheol-Ho Choi{1}, Hyunmin Choi{1}, Taehyun Kim{1}, Seongtaek Hong{1}, Taehyung Kim{1}, Jeongkyu Hong{2}
{1}Hanwha Systems Co., Ltd.; {2}University of Seoul

16:30 – 16:45
Transition

16:45 - 17:45
Co-Design for Emerging Platforms: Neuromorphic Sensing and Programmable Biochips
Room: Blossom Hill III
Session Chair(s): Xiaoxuan Yang, University of Virginia

This session highlights CAD methods for emerging domain-specific platforms: NTT-based neuromorphic auditory sensing that accelerates event-driven feature extraction, and waste-aware placement/assignment for fully programmable valve-array biochips. Together, the papers show how algorithm-hardware and operation-aware co-optimization improve latency, energy, resource utilization, and reliability.

16:45
946: N-Cochlea: Algorithm Hardware Co-Design Framework for NTT-Based Auditory Feature Extraction
Yiwei Si{1}, Danni Wang{1}, Shan He{2}, Donghui Guo{2}, Sizhao Li{1}
{1}Harbin Engineering University; {2}Xiamen University

17:00
658: WA-COP: Waste-Aware Co-Optimization of Module Placement and Fluid-to-Cell Assignment for Fully Programmable Valve Array Biochips
Jingyi Wang{1}, Yuhan Zhu{1}, Ziyang Chen{2}, Zhisheng Chen{1}, Genggeng Liu{1}, Xing Huang{2}
{1}Fuzhou University; {2}Northwestern Polytechnical University

16:45 - 17:45
Emerging Memory and Storage Architectures
Room: Blossom Hill I/II
Session Chair(s): Nan Wu, George Washington University; Meng Li, Peking University

Modern AI applications place unprecedented demands on memory and storage systems, motivating new architectural solutions across the memory hierarchy. This session explores architectural techniques for improving the performance, reliability, and efficiency of modern memory and storage systems, showcasing the critical role of memory and storage in data-intensive and AI-driven workloads.

16:45
794: A Unified DRAM Address Mapping with Dynamic-KV Co-Optimization for Heterogeneous NPU-PIM LLM Inference
Xuhong Qiang{1}, Chenchen Liu{1}, Hai Huang{1}, Kaixin Tan{1}, Chao Xie{1}, Feng Qiu{1}, Bei Liu{1}, Xiang Chen{2}, Weisheng Zhao{1}
{1}Beihang University; {2}Peking University

17:00
1560: PM-Scrub: Enhancing CXL-MRAM Reliability with an Adaptive Scrubbing Scheme
Zhuang Ma, Yang Ou, Lizhou Wu, Hanzhi Xun, Simiao Zhao, Yanjing Wang, Jiaqing Xu, Zhengbin Pang, Wei Chen, Mingche Lai, Simiao Zhao
National University of Defense Technology

17:15

1912: EVA: Efficient Read-Disturb Management Using Error-Aware Voltage Adjustment for High-Capacity 3D Flash-Based SSDs

Beomjun Kim{1}, Haun Jeon{1}, Seonguk Jang{1}, Ilpyo Hong{1}, Jaeyong Lee{2}, Myoungjun Chun{3}, Myungsuk Kim{1}
{1}Kyungpook National University; {2}SK hynix; {3}Soongsil University

17:30

514: HIPER: Logarithmic Prefix Matching for KV Cache Offloading on ZNS SSDs

Chia-Han Lu{1}, Chieh-Lin Tsai{1}, Yu-Ming Huang{2}, Yu-Chen Lu{2}, Tei-Wei Kuo{1}, Yuan-Hao Chang{1}, Chia-Han Lu{1}
{1}National Taiwan University; {2}Macronix International Co., Ltd.

16:45 - 18:15
Hierarchical IR-Drop for Chiplets and 3D-ICs: Scalable PGA, Correlation, and Interoperability
Room: Willow Glen I/II
Session Chair(s): Abhijeet Chakraborty, Synopsys

AI-era systems increasingly rely on chiplets, advanced packaging, and 3D stacking, where power delivery, transient droop, and reliability must be managed across die boundaries and design hierarchies. A key bottleneck in scaling power integrity for such systems is the lack of a consistent abstraction and data exchange model that connects early power modeling to signoff-level IR-drop and reliability analysis. While standardized multi-level power modeling abstractions (e.g., IEEE 2416) enable architectural exploration, signoff IR-drop and electromigration data remain locked in tool-specific representations (e.g., XPGV, ROM), making correlation, reuse, and hierarchical integration difficult. This disconnect increasingly leads to late-stage correlation gaps, conservative guard-bands, and limited reuse across design hierarchies. This session presents a unified technical pipeline spanning abstraction-level modeling, hierarchical power-grid analysis (PGA), and multi-die power delivery. It begins by examining how power-model fidelity evolves across abstraction levels and how mismatches propagate into IR-drop sensitivity, timing margins, and reliability outcomes. It then introduces scalable hierarchical PGA methods that preserve signoff fidelity from IP to SoC. Extending into 3D-IC systems, the session addresses vertical PDN challenges including TSV/microbump interfaces, inter-die coupling, and thermally amplified droop. The session is structured to explicitly connect early modeling choices to downstream physical effects. Each talk presents concrete methodologies, measurable correlation insights, and practical guidance for integrating modeling, analysis, and validation across the design stack. Together, the session highlights where fidelity is lost across abstraction levels and how it can be recovered through improved modeling, aggregation, and validation. Unlike prior efforts that treat modeling, IR-drop, and packaging independently, this session focuses on their coupling across abstraction levels, hierarchy, and multi-die integration. A cross-vendor panel will debate whether the field should converge toward a hierarchical IR-drop data exchange layer or continue with tool-specific ecosystems augmented by adapters, considering fidelity, scalability, and IP protection. The timing is driven by industry shifts toward chiplets, 3D-IC integration, and early-stage power-driven design. Without a consistent methodology, teams face late-stage surprises, over-conservative guard-bands, and limited reuse across hierarchies. This session addresses that gap with both practical methods and a forward-looking path toward interoperable design flows.

16:45
Welcome
Nagu Dhanwada
Google

16:50
Power Modeling Fidelity Across Abstraction Levels
Yuqin Wang, Rhett Davis
NC State

17:05
Power Delivery for 3D-IC Systems: PDN Co-Design, Vertical Coupling, and Correlation
Paul Franzon
NC State

17:20
Hierarchical Power-Grid Analysis at Scale: Reusable IR-Drop Metrics From IP to SoC
Hailin Jiang
Cadence

17:35
3D Integration Trade-offs for Power Delivery
Sung Kyu Lim
USC

17:50
Panel Discussion
Moderator: Nagu Dhanwada (Google)
Panelists: Rhett Davis (NC State), Paul Franzon (NC State), Hailin Jiang (Cadence), Sung Kyu Lim (USC), Pratyush Kamal (Siemens EDA->Ergodex AI), Abhijeet Chakraborty (Synopsys -> UCLA).

Paper(s):
1992: Compiler-Driven Full-System Design Space Exploration of a Configurable CNN Accelerator
Yuqin Wang, W. Rhett Davis
North Carolina State University

16:45 - 17:45
Photonics & Optoelectronic Systems Design
Room: Guadalupe
Session Chair(s): Tathagata Srimani, Carnegie Mellon University

This session focuses on the design, evaluation, and automation of photonic and optoelectronic systems, from on-chip optical networks to LLM-assisted photonic circuit design. Papers span optical network-on-chip optimization, frameworks for emerging vision devices, and agentic design tooling.

16:45
1347: SenseFlow：A Unified Framework for Evaluating Emerging Optoelectronic Devices in Intelligent Vision Systems
Jinghan Xu, Ligong Zhang, Jiaqi Li, Shuhan Wang, Yizhan Liu, Haotong Zhu, Xionghao Qin, Zheng Zhou, Peng Huang, Xiaoyan Liu
Peking University

17:00
348: Signal-Integrity-Aware Design for Wavelength-Routed Optical NoC via Co-Optimization of Spectral Power and Transient Energy
Liaoyuan Cheng, Jianpeng Cai, Mengchu Li, Tsun-Ming Tseng, Ulf Schlichtmann
Technical University of Munich

17:15
36: PICopilot: An LLM-based Agentic Framework for Assisting Photonic Integrated Circuit Design via Script Generation
Xiaohan Jiang{1}, Zeyu Li{1}, Wei Zhang{1}, Jiang Xu{2}
{1}The Hong Kong University of Science and Technology; {2}The Hong Kong University of Science and Technology (Guangzhou)

16:45 - 17:45
Smarter by Design: Learning and Optimization for Chiplet and 3D Systems
Room: San Jose Ballroom
Session Chair(s): Feng Ling, Xpeedic; Rongmei Chen, Peking University

This session showcases emerging optimization methodologies for chiplet-based and 3D systems. The papers introduce reinforcement learning, differentiable optimization, thermal-aware planning, and memory optimization techniques that address routing, placement, power, timing, and performance challenges in next-generation heterogeneous integration.

16:45
1055: PWaT3D-Planner: Bridging Power and Timing Optimization with Thermal-Mechanical Awareness
Cristhian Roman{1}, Donghao Fang{1}, Haodong Chang{1}, Cunxi Yu{2}, Yiran Chen{3}, Jiang Hu{1}
{1}Texas A&M University; {2}University of Maryland, College Park; {3}Duke University

17:00
1064: TurnRL: Reinforcement-Learning Driven Turn-Restricted Adaptive Routing for Chiplet-Based System
Kangqi Zhang, Jiazhen Wang, Ronald Dreslinski, Nathan Bleier
University of Michigan

17:15
881: Differentiable Thermal-Aware Placement for Monolithic 3D ICs with Stack-Aware Thermal Modeling
Gyumin Kim, Heechun Park
UNIST

17:30
789: 3D-ReDPB: Reordering-Enabled Diagonal-Pattern-Based Compression for Scientific SpMV on 3D DRAM PNM
Xiang Li{1}, Tianyan Zhong{2}, Wenbin Jia{2}, Xinyuan Lin{2}, Zongle Huang{2}, Chen Tang{2}, Wenxun Wang{2}, Wenyu Sun{3}, Sheng Zhang{1}, Yongpan Liu{2}
{1}Tsinghua Shenzhen International Graduate School; {2}Tsinghua University; {3}Qiyuan Lab

18:15 – 20:15
Gala Dinner
Room: San Jose Ballroom

Technical Program – Nov 12

8:30 – 18:00
2026 QCAS Quantum Computing Applications and Systems Workshop
Room: Willow Glen I/II

8:30 – 12:10
Memory-Centric Computing for LLM Inference
Room: Blossom Hill I/II

8:00 – 16:30
SSH-SoC: Safety and Security in Heterogeneous Open System-on-Chip Platforms
Room: Blossom Hill III

8:00 – 17:00
SUSHI'26: Sustainable Hardware Security
Room: Guadalupe

9:30 – 10:00
Coffee Break

11:30 – 13:00
Lunch

14:30 – 15:00
Coffee Break