



ISLEDA 2026

International Symposium of EDA

May 8-10, 2026 | Singapore

Conference Program



TABLE OF CONTENTS

CONFERENCE COMMITTEE	1
SUPPORTERS & SPONSORS	3
CONFERENCE VENUE	4
AGENDA OVERVIEW	5
KEYNOTE SPEAKERS	9
WORKSHOP	19
PANELS	20
SPECIAL SESSIONS	23
TUTORIALS	30
BEST PAPER CANDIDATES	40
TECHNICAL SESSIONS	42
POSTER SESSION	102



CONFERENCE COMMITTEE

Executive Committee

General Chairs

Prof. Hao, Yue, Xidian University, Academician of the Chinese Academy of Sciences

Prof. Wang, Runsheng, Peking University

Steering Committee

Prof. Wei, Shaojun, Tsinghua University

Prof. Zeng, Xuan, Fudan University

Dr. Girard, Patrick, French National Center for Scientific Research

Prof. Deen, Jamal, McMaster University

Dr. Tanurhan, Yankin, Synopsys

Technical Program Committee

Conference Chairs

Dr. Hui, Chiu Wing Colin, GlobalFoundries

Prof. Ho, Tsung-Yi, The Chinese University of Hong Kong

Prof. Qu, Gang, University of Maryland

Technical Program Chairs

Prof. Yu, Bei, The Chinese University of Hong Kong

Prof. Schlichtmann, Ulf, Technical University of Munich

Ms. Leung, Kafai, A*STAR Institute of Microelectronics

Keynote Chair

Prof. Liang, Yun, Peking University

Publication Chair

Prof. Xu, Qiang, The Chinese University of Hong Kong

Panel Chair

Mr. Ong, Jonathan, Applied Materials

Industry Chair

Mr. Lam, Lup Meng, Synopsys

Special Session Chair

Mr. Ghosh, Swaroop, Pennsylvania State University

Publicity Chair

Prof. Basu, Kanad, Rensselaer Polytechnic Institute

Tutorial Chair

Prof. Sasan, Avesta, University of California, Davis

Tutorial Co-Chair

Prof. Lim, Yeow Kheng, National University of Singapore

Outreach Chair-US

Prof. Forte, Domenic, University of Florida

Outreach Chair-Europe

Prof. Halak, Basel, University of Southampton

Outreach Chair-Asia

Prof. Li, Yongfu, Shanghai Jiao Tong University

Outreach Chair-Canada

Prof. Chun, Peter, University of Alberta

Track Committee Chairs

[1] System-Level Modeling and Design Methodology

Sun, Qi, Zhejiang University

Zhao, Jieru, Shanghai Jiao Tong University

[2] Memory Architecture and Near/In Memory Computing

Chen, Xiaoming, Institute of Computing Technology, CAS

Zhang, Li, Technical University of Darmstadt

[3] Analog-Mixed Signal Design Automation

Zhu, Keren, Fudan University

Du, Li, Nanjing University

Centurelli, Francesco, Sapienza University of Rome



[4] High-Level, Behavioral, and Logic Synthesis and Optimization

Qian, Weikang, Shanghai Jiao Tong University

Pilato, Christian, Politecnico di Milano

[5] Analysis and Optimization for Power and Timing

Xie, Zhiyao, The Hong Kong University of Science and Technology

Li, Bing, Technische Universität Ilmenau

[6] Physical Implementation

Ma, Yuzhe, The Hong Kong University of Science and Technology (Guangzhou)

Lin, Yibo, Peking University

[7] Testing, Validation, Simulation, and Verification

Zhang, Hongce, The Hong Kong University of Science and Technology (Guangzhou)

Zheng, Yuanjin, Nanyang Technological University

[8] Design for Manufacturability and Reliability

Chen, Yu-Guang, National Central University

Li, Meng, Peking University

[9] Packaging & Multi-Physics Simulation

Peng, Yarui, Southeast University

Tseng, Tsun-Ming, Technical University of Munich

[10] Technology and Modeling

Ng, Kian Ann, Singapore Institute of Technology

Chen, Quan, Southern University of Science and Technology

[11] Emerging Technologies and Applications

Lee, Vincent, National University of Singapore

Li, Xingquan, Peng Cheng Laboratory

[12] AI & Open Source EDA

Yu, Cunxi, University of Maryland

Tan, Chee Wei, Nanyang Technological University



SUPPORTERS & SPONSORS

Special Supporters



www.pftn-semi.com/

Diamond Sponsors



<https://www.semitronix.com/>

www.primarius-tech.com/

www.empyrean.com.cn/



<http://www.sichip.com.cn/>

www.witmem.com

Platinum Sponsors



www.univista-isg.com

www.hexineda.com

Gold Sponsors



www.giga-da.com

www.amedac.com

Silver Sponsor



www.e2uninova.com



CONFERENCE VENUE

Marina Bay Sands Singapore

Address: Marina Bay Sands is located at 10 Bayfront Avenue, Singapore 018956.



AGENDA OVERVIEW

Registration & Conference Materials Collection Foyer of Peony Ballroom @ Marina Bay Sands Singapore		May 07 10:00-18:00 May 08 08:00-18:00 May 09 08:00-16:30 May 10 08:00-16:30
May 8, 2026 Friday		
08:30-09:00	Opening Ceremony & Announcement of ISED 2027 Keynote Speech	Peony Ballroom
09:00-09:40	Prof. Benini, Luca Università di Bologna Speech Title: Designing Open Platforms for the Embodied AI Era	
09:40-10:20	Prof. Yeo, Yee Chia Deputy Chief Executive (Innovation & Enterprise), A*STAR Speech Title: Accelerating Electronic-Photonic System Design and Integration - Regional Innovation Meets Global Needs	
10:20-10:40	Coffee Break	
10:40-11:20	Mr. Kengeri, Subramani Corporate Vice President and GM, Systems to Materials, Applied Materials Speech Title: System Technology Co-optimization in AI System Solutions	
11:20-12:00	Prof. Yeo, Kiat Seng Singapore University of Technology and Design Speech Title: The Rise of AI and Its Impact on ICs	
12:00-12:40	Lunch Break	
12:40-13:20	Mr. Chan, Don Vice president, Research & Development, Cadence Design Systems, Inc. Speech Title: From Point Tools to Agentic AI: How EDA has impacted the Semiconductor Industry	Orchid 4301 Orchid 4303 Orchid 4302 Orchid 4304
13:30-16:20	Workshop Bridging the Gaps: Virtual Fabrication-to-Design Flow for DTCO/STCO	
13:30-18:00	Special Session on EDA Standards	
13:30-15:30	Panel Session 01 Digital Twins for Verification & Manufacturing: Closing the Loop From Model to Fab	
	Technical Session 01 Heterogeneous & Optical Compute-Infrastructure Innovations	



13:30-15:30	Technical Session 02 Logic and High-level Synthesis	Orchid 4201
	Technical Session 03 Fast Thermal & Reliability Modeling for 3D ICs	Orchid 4202
	Technical Session 04 Hardware Design, Verification & ML-Enabled RTL	Orchid 4203
15:30-16:00	Coffee Break	Peony Ballroom
15:30-16:00	Poster Session	
16:00-18:00	Technical Session 05 Topology/Layout, ADC & LLM-enabled Synthesis	Orchid 4302
	Technical Session 06 Formal Verification & Equivalence Checking	Orchid 4304
	Technical Session 07 Timing Analysis & Optimization	Orchid 4201
	Technical Session 08 Lithography & Mask Optimization	Orchid 4202
	Technical Session 09 Placement & Floorplanning Algorithms	Orchid 4203
	Panel Session 02 Agentic AI + Scalable Compute in EDA: From Point Tools to Flow Orchestration	Orchid 4204
18:30-20:00	Banquet	Peony Ballroom



May 9, 2026 | Saturday

Keynote Speech

08:30-09:10	Prof. Parhi, Keshab University of Minnesota Speech Title: Energy-Efficient Domain-Specific AI Models and Agents	Peony Ballroom
09:10-09:50	Prof. Chakrabarty, Krishnendu Fulton Professor of Microelectronics, Arizona State University Speech Title: Toward Self-Healing Interconnects in Advanced Packaging: Defect Analysis, Built-in Self-Test, and Repair	
09:50-10:30	Prof. Alioto, Massimo National University of Singapore Speech Title: Advances in Energy-Efficient Always-On AI Accelerator Architectures - From Battery-Powered to Purely-Harvested Systems	
10:30-11:00	Coffee Break	
11:00-11:40	Prof. Takahashi, Atsushi Institute of Science Tokyo Speech Title: Routing Algorithms - from the basics to recent advances	
11:40-12:20	Prof. Wunderlich, Hans-Joachim Professor Emeritus of the University Stuttgart Speech Title: From Hidden Faults to Observable Failures	
12:20-13:30	Lunch Break	
13:30-16:30	Technical Session 10 Compact Modeling & Parameter Extraction	Orchid 4301
	SCIS Special Session on Advanced IC Design and Technology in AI Era	Orchid 4302
	Technical Session 11: AI-Driven Front-End & Circuit Optimization	Orchid 4303
	Panel Session 03 Chiplets at Scale: Can Standards, Packaging, and Qualification Enable a True Multi Vendor Ecosystem?	Orchid 4304
	Technical Session 12 Routing & Layout Engines	Orchid 4201
	Technical Session 13 Reliability, Device-Circuit Co-Optimization & Emerging Tech	Orchid 4202
	Technical Session 14 In-Memory Computing: Architectures for AI, Security, and Beyond	Orchid 4203
16:30-	Excursion	



May 10, 2026 | Sunday

Keynote Speech		
09:00-09:40	Prof. Gupta, Puneet University of California, Los Angeles Speech Title: Scale-out Chiplet Based Systems: Design, Architecture and Pathfinding	Peony Ballroom
09:40-10:00	Morning Coffee Break	
10:00-12:20	Technical Session 15 Device Characterization, Cryogenics & Cross-Layer Design	Orchid 4301
	Technical Session 16 Automated Circuit Design & Optimization	Orchid 4302
	Technical Session 17 Prediction & Netlist Analytics	Orchid 4303
	Technical Session 18 Test, Emulation, Layout & Lifecycle Tools	Orchid 4304
	Technical Session 19 Power, IR Drop & Waveform Prediction (ML-Accelerated)	Orchid 4201
	Technical Session 20 Layout, Pattern Matching & Front-End Cell/Mask Flow	Orchid 4202
	Technical Session 21 Architecture, Placement, Routing & Security	Orchid 4203
12:00-13:30	Lunch Break	Peony Ballroom
13:30-16:30	Tutorial 01 Parallel Sparse Direct Solver for SPICE	Orchid 4301
	Tutorial 02 ECOS Studio: Building an Open Ecosystem for Silicon Design and Education	Orchid 4302
	Tutorial 03 AI for EDA Feedback Loops: Leveraging Manufacturing Data to Evolve PDKs, Models, and Design Rules	Orchid 4303
	Tutorial 04 CMOS Technology and Its Application to Design Multi-Standard RF/mm-Wave Low-Noise Amplifiers	Orchid 4304
	Tutorial 05 Design Methodologies for Intelligent & Perceptive Hardware Security - From Circuits to Machine Learning Algorithms	Orchid 4201
	Tutorial 06 Design of Energy-Constrained AI Accelerators for Edge Computing	Orchid 4202
	Tutorial 07 Sub-THz Circuits and Systems for Radar Transceivers	Orchid 4203
14:00-18:30	Summer of Talents	Peony Ballroom



KEYNOTE SPEAKERS

 09:00-09:40, May 8

 Peony Ballroom



Prof. Benini, Luca

Università di Bologna

Bio.: Luca Benini holds the chair of digital Circuits and systems at ETHZ and is Full Professor at the Università di Bologna. He received a PhD from Stanford University. His research interests are in energy-efficient parallel computing systems, smart sensing micro-systems and machine learning hardware. He is a Fellow of the IEEE, of the ACM, a member of the Academia Europaea and of the Italian Academy of Engineering and Technology. He is the recipient of several awards including the 2023 IEEE CS E.J. McCluskey Award, and the 2024 IEEE CS Open Source Hardware contribution Award.

Title: Designing Open Platforms for the Embodied AI Era

Abstract: AI is accelerating into the generative era with increasing capabilities being "embodied" everywhere, from earbuds to cars to humanoid robots. Embodied AI needs to tackle major challenges in energy efficiency, safety, security, and real-time predictability, while curtailing computational complexity. In this talk I will focus on chip and system design for embodied AI, moving from ultra-low power AI-enhanced MCUs for smart wearables and nano-robots to large 3D-integrated systems-in-package for autonomous vehicles, humanoids and satellites. I will emphasize the strategic importance of AI-enhanced co-design flows and of an open-platform (hardware and software) approach to ensure fast innovation cycles as well as long term sustainability, safety and security.

 09:40-10:20, May 8

 Peony Ballroom



Prof. Yeo, Yee Chia

Deputy Chief Executive (Innovation & Enterprise), A*STAR

Bio.: Prof. Yeo Yee Chia is currently the Deputy Chief Executive (Innovation & Enterprise), A*STAR. He provides strategic leadership and oversight of the Innovation & Enterprise Group and National Platforms. He orchestrates A*STAR's industry approach and priorities to ensure that A*STAR's translational activities achieve significant national impact and outcomes.

Prior to this, he was Assistant Chief Executive (Innovation & Enterprise), where he strengthened A*STAR's industry partnerships and commercialisation efforts. Prof. Yeo also held various leadership roles in A*STAR, including Executive Director (IP Management) and Executive Director of the Science & Engineering Research Council (SERC). He leads substantial research efforts at A*STAR's Institute of Microelectronics, working on frontier microelectronic devices and key challenges in semiconductor technology. He plays a significant role in



Singapore's Future of Microelectronics initiative to steer next-generation microelectronics strategy.

Prof. Yeo is a renowned scholar in academia and a technology leader in the semiconductor industry, with expertise in transistor architecture, device modelling and simulation, materials, tools and processing technologies.

He spent 10 years at Taiwan Semiconductor Manufacturing Company (TSMC). He was Director of Research and Development (R&D) at TSMC where he led organisations spanning research, pathfinding, and development. He co-developed TSMC's industry-leading 7 nm, 5 nm, and 3 nm technologies, which has impacted the world in areas such as high performance computing, artificial intelligence, and consumer electronics.

In academia, he is Professor of Electrical and Computer Engineering, National University of Singapore (NUS), where he was an award-winning, high impact researcher with an outstanding track record of achievements. He published over 700 research papers and is an inventor of 289 US patents. 44 PhD students graduated under his tutelage.

He received his PhD and MS degrees in Electrical Engineering and Computer Sciences from the University of California, Berkeley, and MEng and BEng degrees in Electrical Engineering from NUS.

Title: Accelerating Electronic-Photonic System Design and Integration - Regional Innovation Meets Global Needs

Abstract: The semiconductor industry is in a phase of unprecedented growth, driven by the rapid rise of AI and data-centric applications. Complexity of next-generation technologies continues to escalate. As high-performance systems increasingly integrate electronics and photonics, system design and integration now demand multi-domain expertise, advanced infrastructure and close collaboration across the value chain. Even as semiconductor firms invest significantly in R&D, the pace and complexity of innovation are pushing against the limits of what they can address alone. This talk explores how heterogeneous integration across advanced packaging, silicon photonics, and co-packaged optics is reshaping system innovation, and how System Technology Co-Optimisation connects these vectors at the system level, optimising performance, energy efficiency and integration. It also shows how companies can leverage A*STAR's world-class research expertise, production-grade infrastructure and collaborative R&D model to accelerate system design from prototype to production, and stay competitive in an era where speed, resilience and partnership will define the next wave of global semiconductor leadership.

10:40-11:20, May 8

Peony Ballroom



Mr. Kengeri, Subramani

Corporate Vice President and GM, Systems to Materials, Applied Materials

Bio.: Subramani (Subi) Kengeri is the Corporate Vice President and General Manager at Applied Materials, leading e Systems-to-Materials group since joining in 2020. He has led world-wide teams in technical and executive roles at Globalfoundries, TSMC, and Texas Instruments. He holds over 50 U.S. patents, and has delivered more than 100 invited talks and interviews.

Title: System Technology Co-optimization in AI System Solutions

Abstract: The Landscape of AI systems solutions is evolving and the role of System Technology Co-optimization (STCO) is becoming increasingly important, as we navigate beyond the traditional boundaries



of Moore's Law. To push past both physical and economic constraints of technology scaling, we need to identify and harness structural shifts that offer non-linear improvements in efficiency, all while maintaining a comprehensive, integrated systems perspective. In this new era defined by AI and humanoid advancements, it's no longer viable for technology, hardware, and software to operate in silos. The integration across the entire stack—from materials and transistors up to system architecture—is now essential.

EDA, combined with STCO, forms a foundational framework to enable this level of integration. By applying these principles, we can deliver the kind of massive throughput and energy efficiency that will drive the next intelligence revolution. As we move forward, embracing this holistic approach will be critical to unlocking the full potential of next-generation computing.

11:20-12:00, May 8

Peony Ballroom



Prof. Yeo, Kiat Seng

Singapore University of Technology and Design

Bio.: Professor Kiat Seng YEO, (PhD, PPA, PBS, FSAEng, FSNAS, FCAE, FIEEE, FAAET, FAIIA, FAAIA) received his BEng and PhD degrees in EE from Nanyang Technological University (NTU), Singapore. He has 37 years of experience in industry, academia, start-ups and consultancy. Currently, he is Advisor for Global Partnerships (President's Office) and Director for Innovation and Enterprise (China) at the Singapore University of Technology and Design (SUTD) and Distinguished Professor at Tianjin University. Prof. Yeo is a widely known authority in low-power RF/mm-wave IC design and a recognized expert in CMOS technology. He was the Founding Chairman of the University Research Board, Founding Chairman of the Board of Graduate Studies, Member of the Academic Council, Associate Provost for Research, Founding Associate Provost for Graduate Studies, and Founding Associate Provost for International Relations at SUTD. Before his appointment at SUTD, he was Associate Chair (Research), Head of the Division of Circuits and Systems, and Founding Director of VIRTUS of the School of Electrical and Electronic Engineering at NTU. He has secured over S\$70M of research grants as PI, published 14 books (4 Amazon Best Sellers), 7 book chapters, 700+ journal and conference papers, and holds 55 patents (27 US patents). Professor Yeo holds/held key positions in many international conferences as Advisor, General Chair and Technical Chair. He is one of the 7 "double academicians" of the Singapore Academy of Engineering and the Singapore National Academy of Science. He is also a Foreign Fellow of the Canadian Academy of Engineering and Fellows of the ASEAN Academy of Engineering and Technology, the Asia-Pacific Artificial Intelligence Association, the International Artificial Intelligence Industry Alliance, and the IEEE. He received 2 National Day Awards from the President of Singapore in 2009 and 2020 and awarded the Nanyang Alumni Achievement Award by NTU in 2009. He is the principal author of Integrated Circuit Design Research Ranking for Worldwide Universities 2008 and World University Research Rankings (WURR) 2020. Professor Yeo was recognized among the Top 2% Scientists Worldwide by Stanford University from 2020 to 2025, World's AI Top Scientist by the International Artificial Intelligence Industry Alliance in 2023, and Top Scholar by ScholarGPS from 2023 to 2025.

Title: **The Rise of AI and Its Impact on ICs**

Abstract: The rise of artificial intelligence (AI) has been more prominent in the last few years. It has and will



continue to shape and accelerate the growth of both the electronic design automation (EDA) and the semiconductor industry markets. For example, the EDA market is expected to increase from US\$16 billion in 2026 to over US\$22 billion by 2030. During the same period, the semiconductor industry market is projected to grow from US\$820 billion to over US\$1 trillion.

Traditionally, the growth of both EDA and semiconductor industry is dominated by the 4 C's: Computers, Communications, Consumers, and Cars. But its growth will now come from the 4 I's: Intelligence, Integration, Innovation, and Interdiscipline. The combination of these I's and C's will result in more ICs. Therefore, in addition to new energy, drones and smart cars, ICs will also be widely used in the human body, robots, future communications, Bio.medical, aerospace and many other fields. In the foreseeable future, there will be more microchips in the human body than in any electronic devices. At that time, humans will become more robotic, and robots will become more human.

As AI and IC continue to evolve, it is important to know the forces that have driven it along its historical trajectory, and to discover how much further it could go. What is the next big thing? How is it going to affect us? This keynote will attempt to answer these questions.

12:40-13:20, May 8

Peony Ballroom



Mr. Chan, Don

Vice president, Research & Development, Cadence Design Systems, Inc.

Bio.: Don Chan is the vice president of R&D at Cadence Design Systems. He received his MS degree in electrical engineering from Santa Clara University in 1990. From 1986-1989, he was developing SPICE models. From 1989-1992, he was as ASIC Design Engineer at Fujitsu Microelectronics. From 1993-2018, Don held various leadership positions in both applications engineering and R&D at Synopsys. From 2018-present, Don is the VP of R&D collaborating with the world's #1 foundry in developing the most advanced nodes. In addition, Don is also the product owner for the frontend products at Cadence which includes Conformal, Modus and Stratus. Don has given many keynote speeches in many conferences in Asia. Don is also currently a Lecturer at Santa Clara University.

Title: From Point Tools to Agentic AI: How EDA has impacted the Semiconductor Industry

Abstract: Electronic Design Automation (EDA) is the enabler for today's semiconductor growth. Moore's Law has been driving the semiconductor industry for the past several decades. Correspondently, EDA has increased the productivity of engineers by more than 10,000X over the decades. In this talk, we discussed how EDA together with design technology co-optimization has advanced beyond Moore's law and transistor scaling. With transistor scaling becoming increasingly challenging, the future of designs will approach 3D in the next few years. Driven by high performance computing (HPC) and AI applications, systems will be comprised of heterogenous chipelets with different process nodes. In the era of AI, design methodology will become AI-assisted design methodology and scripts will be driven by agentic AI and agents. The future of EDA research lies in agentic workflows.



KEYNOTE SPEAKERS



08:30-09:10, May 9



Peony Ballroom



Prof. Parhi, Keshab

University of Minnesota

Bio.: Keshab K. Parhi received the B.Tech. degree from the Indian Institute of Technology (IIT), Kharagpur, in 1982, the M.S.E.E. degree from the University of Pennsylvania, Philadelphia, in 1984, and the Ph.D. degree from the University of California, Berkeley, in 1988. He has been with the University of Minnesota, Minneapolis, since 1988, where he is currently the Erwin A. Kelen Chair and a Distinguished McKnight University Professor in the Department of Electrical and Computer Engineering. He has published over 750 papers, is the inventor of 36 patents, and has authored the textbook VLSI Digital Signal Processing Systems: Design and Implementation (Wiley, 1999). His current research addresses VLSI architecture design of artificial intelligence and machine learning systems, signal processing systems, hardware security, quantum information science, and data-driven neuroengineering and neuroscience with applications to neuro-psychiatric disorders.

Dr. Parhi is the recipient of numerous awards including the 2003 EEE Kiyo Tomiyasu Technical Field Award; and the 2017 Mac Van Valkenburg award, and the 2012 Charles A. Desoer Technical Achievement award, from the IEEE Circuits and Systems Society. He served as the Editor-in-Chief of the IEEE Trans. Circuits and Systems, Part-I during 2004 and 2005, and currently serves as the Editor-in-Chief of the IEEE Circuits and Systems Magazine. He is a Life Fellow of IEEE, a Fellow of the American Association for the Advancement of Science (AAAS), the Association for Computing Machinery (ACM), the American Institute of Medical and Bio.logical Engineering (AIMBE), and the National Academy of Inventors (NAI).

Title: Energy-Efficient Domain-Specific AI Models and Agents

Abstract: The field of artificial intelligence (AI) has taken a tight hold on broad aspects of society, industry, business, and governance in ways that dictate the prosperity and might of the world's economies. The AI market size is projected to grow from \$189 billion in 2023 to \$4.8 trillion by 2033. Currently, AI is dominated by large language models (LLMs) that exhibit linguistic and visual intelligence. However, training these models requires a massive amount of data scraped from the web as well as large amounts of energy (50-60 GWh to train GPT-4). Despite these costs, these models often hallucinate, a characteristic that prevents them from being deployed in critical application domains. In contrast, the human brain consumes only 20W of power. What is needed is the next level of AI evolution in which lightweight domain-specific multimodal models with higher levels of intelligence can reason, plan, and make decisions in dynamic environments with real-time data and prior knowledge, while learning continuously and evolving in ways that enhance future decision-making capability. This will define the next wave of AI, progressing from today's large models, trained with vast amounts of data, to nimble energy-efficient domain-specific agents that can reason and think in a world full of uncertainty. To support such agents, hardware will need to be reimagined to allow energy efficiencies $\geq 1000\times$ over the state of the art. Such a vision of future AI systems is presented in this talk. The talk is based on collaborative work with colleagues as presented in arXiv 2510.22052.



KEYNOTE SPEAKERS

09:10-09:50, May 9

Peony Ballroom



Prof. Chakrabarty, Krishnendu

Fulton Professor of Microelectronics, Arizona State University

Bio.: Prof. Krishnendu Chakrabarty is the Fulton Professor of Microelectronics in the School of Electrical, Computer and Energy Engineering at Arizona State University (ASU). He is also the CTO of the SWAP Hub for the Department of War Microelectronics Commons (<https://microelectronics.asu.edu/southwest-advanced-prototyping-hub/>), and Director of the ASU Center for Semiconductor Microelectronics (ACME, <http://acme.asu.edu>). Before moving to ASU, he was the John Cocke Distinguished Professor and Chair of Electrical and Computer Engineering at Duke University. Prof. Chakrabarty is a recipient of the Humboldt Research Award from the Alexander von Humboldt Foundation, Germany, the IEEE Transactions on CAD Donald O. Pederson Best Paper Award, the IEEE Transactions on VLSI Systems Prize Paper Award, the ACM Transactions on Design Automation of Electronic Systems Best Paper Award, and over a dozen best paper awards at major conferences. He is also a recipient of the IEEE Computer Society Technical Achievement Award, the IEEE Circuits and Systems (CAS) Society Charles A. Desoer Technical Achievement Award, the IEEE CAS Society Vitold Belevitch Award, the Semiconductor Research Corporation (SRC) Technical Excellence Award, the SRC Aristotle Award, the SRC Innovation Award, the IEEE-HKN Asad M. Madni Outstanding Technical Achievement and Excellence Award, and the IEEE Test Technology Technical Council Bob Madge Innovation Award. He is a Research Ambassador of the University of Bremen (Germany) and he was a Hans Fischer Senior Fellow at the Institute for Advanced Study, Technical University of Munich, Germany during 2016-2019. He is a 2018 recipient of the Japan Society for the Promotion of Science (JSPS) Invitational Fellowship in the "Short Term S: Nobel Prize Level" category. He is also a recipient of the Distinguished Alumnus Award from the Indian Institute of Technology, Kharagpur. Prof. Chakrabarty is a Fellow of ACM, IEEE, and AAAS, and a Golden Core Member of the IEEE Computer Society. He is a Fellow of the National Academy of Inventors.

Title: Toward Self-Healing Interconnects in Advanced Packaging: Defect Analysis, Built-in Self-Test, and Repair

Abstract: This talk will focus on how we can recover from defects in fanout wafer-level packaging (FOWLP) and hybrid bonding. Defects in FOWLP can arise from the coefficient of thermal expansion mismatch, warpage, die shift, and post-molding protrusion, causing misalignment and imperfect bonding during redistribution layer (RDL) buildup. As a result, high-density back-end-of-line (BEOL) interconnects, RDLs, and through-mold vias are susceptible to warpage-induced stress, strain, and deformation. Hybrid bonding interfaces are vulnerable to nanoscale voids, contamination, and alignment-induced opens or shorts. The speaker will present an analysis of these defects, fault models, and impact of these defects on logic functionality and timing. The presentation will showcase recent breakthroughs in built-in self-test and repair for FOWLP and hybrid bonding, translating



cutting-edge academic research into practical solutions.

KEYNOTE SPEAKERS

09:50-10:30, May 9

Peony Ballroom



Prof. Alioto, Massimo

National University of Singapore

Bio.: Massimo Alioto is Provost's Chair Professor at the ECE Department of the National University of Singapore, where he leads the Green IC group, the Integrated Circuits and Embedded Systems area, and the FD-fAbrICS center on intelligent&connected systems. Previously, he held positions at the University of Siena, Intel Labs - CRL (2013), University of Michigan - Ann Arbor (2011-2012), University of California - Berkeley (2009-2011), EPFL - Lausanne.

He is (co)author of 400 publications on journals and conference proceedings, and four books with Springer (with two more coming). His primary research interests include ultra-low power and self-powered systems, green computing, circuits for machine intelligence, hardware security, and emerging technologies.

He was the Editor in Chief of the IEEE Transactions on VLSI Systems and Deputy Editor in Chief of the IEEE Journal on Emerging and Selected Topics in Circuits and Systems. He was the Chair of the Distinguished Lecturer Program for the IEEE CAS Society, and was a Distinguished Lecturer for the SSC and CAS Society. Previously, Prof. Alioto was the Chair of the "VLSI Systems and Applications" Technical Committee of the IEEE Circuits and Systems Society (2010-2012). He served as Guest Editor of numerous journal special issues (JSSC, TCAS-I, JETCAS...), Technical Program Chair of several IEEE conferences (ISCAS, SOCC, PRIME, ICECS), and TPC member (ISSCC, ASSCC). His research group contribution has been recognized through various best paper awards (e.g., ISSCC), and in the ten technological highlights of the TSMC annual report, among the others. Prof. Alioto is an IEEE Fellow.

Title: Advances in Energy-Efficient Always-On AI Accelerator Architectures - From Battery-Powered to Purely-Harvested Systems

Abstract: Artificial intelligence (AI) has pervasively augmented the capabilities of today's silicon systems with an unprecedented level of context awareness and physical signal insights. For this reason, AI accelerator always-on operation is becoming a fundamental and non-negotiable requirement, which conflicts with the high computational intensity of common models and neural networks. This requires fundamental advances in energy efficiency of always-on AI accelerator architectures to achieve a much more favorable tradeoff between power (or energy/inference) and quality of insights (e.g., accuracy) from algorithm to architecture and circuit. As further challenge, AI accelerators are being progressively pushed into silicon systems with extremely limited power budget down to battery-powered and even purely-harvested systems, which make the embedment of always-on AI even more challenging.

In this keynote, recent directions to drastically improve the energy efficiency of AI always-on accelerators are presented and exemplified with a wide range of silicon demonstrations with state-of-the-art low consumption.



Examples cover a wide range of AI model classes, use cases and sensing modalities such as CNNs, transformers and vision systems. The underlying design principles are illustrated for a very wide range of power budgets down to μ Ws, enabling always-on operation even without any form of system energy storage for extremely small form factors and low cost. As common thread, a clear pathway to make AI truly ubiquitous in next-generation silicon systems is unraveled, calling upon our entire research and chip design community to join forces in making it feasible at scale.

🕒 11:00-11:40, May 9

📍 Peony Ballroom



Prof. Takahashi, Atsushi

Institute of Science Tokyo

Bio.: Atsushi TAKAHASHI received his B.E., M.E., and D.E. degrees in electrical and electronic engineering from Tokyo Institute of Technology, Tokyo, Japan, in 1989, 1991, and 1996, respectively. He had been with the Tokyo Tech as a Research Associate from 1991 to 1997, and as an Associate Professor from 1997 to 2009 and from 2012 to 2015, and as a Professor from 2015. He had been with the Osaka University as an Associate Professor from 2009 to 2012. He is currently a Professor with Department of Information and Communications Engineering, School of Engineering, Tokyo Institute of Technology. He served as the TPC Chair of ASP-DAC 2018, the General Chair of ASP-DAC 2023, and he is currently a vice chair of Steering Committee of ASP-DAC. He also served as the chair/co-chair/TPC member of top conferences such as ASP-DAC, ISPD, ICCAD, DATE and VLSI-DAT. He was a Board of Governors of IEEE CASS from 2015 to 2019, the chair of All Japan Joint Chapter of IEEE CEDA from 2015 to 2016, the chair of Japan Joint Chapter of IEEE CASS from 2020 to 2021, and the President of Engineering Sciences Society of IEICE in 2021. His research interests include in VLSI layout design and combinational algorithms. He is a fellow of IEICE, and a senior member of IEEE and IPSJ, and a member of ACM.

Title: Routing Algorithms - from the basics to recent advances

Abstract: Routing has been extensively studied as one of key issues in physical design area of the electric design automation from its emergence, and various routing algorithms, tools and systems have been developed and used in practice. However, research and development of routing is still required since new constraints and objectives forced on routing arise continuously from new technologies and innovative design flow. In this talk, various routing algorithms from classical one to advanced one will be discussed.



KEYNOTE SPEAKERS

11:40-12:20, May 9

Peony Ballroom



Prof. Wunderlich, Hans-Joachim

Professor Emeritus of the University Stuttgart

Bio.: Hans-Joachim Wunderlich is a professor emeritus of the University of Stuttgart and Life Fellow of IEEE. He received his diploma degree in mathematics from the University of Freiburg, Germany, in 1981 and his Dr. rer. nat. (Ph.D.) from the University of Karlsruhe in 1986. Since 1991, he has been a full professor. From 2002 to 2018, he served as the director of the Institute of Computer Architecture and Computer Engineering at the University of Stuttgart, Germany. He has been an associate editor for various international journals and involved in the organization of a variety of IEEE conferences on the design, testing, and fault tolerance of electronic systems. He has published 17 books and book chapters and more than 330 reviewed scientific papers in journals and conferences. His research interests include testing, reliability, fault tolerance, and design automation of microelectronic systems.

Title: From Hidden Faults to Observable Failures

Abstract: During the last few years, silent data corruption (SDC) received increasing attention, and nearly all the players operating AI-Giga-Factories (Artificial Intelligence) reported a certain amount of semiconductor chips computing erroneous results. This phenomenon is actually not as new as it appears, but it is publicly noticed, since nowadays a single customer may run several 100 thousands of chips, and failures are not any more a one-time exception rather than a pattern.

Nearly two decades ago, automotive companies integrating millions of devices in their produced cars in total, made the same observation and took counteractions. With the increased scaling and variability, the classic view of manufacturing test, fault coverage and test escapes does not hold any more as test is now an indeterministic process.

Even if a test procedure will lead to a complete test coverage for nominal circuits instances, variations may create an instance where the test is incomplete. One reason is the exponentially increasing number of possibly critical paths, and not all of them can be tested. The other reason is in the fact that variations may invalidate a test even for severe gross faults. The problem is aggravated by adding voltage and temperature to the process variations leading to an intractable number of distributions.

Since not only defective but also weak devices have to be identified during test, the challenges become even harder. Weak devices may not fail the structural and functional manufacturing test but lead to early life failures severely impacting reliability. Hence, extra-functional properties have to be observed including leakage, power-consumption, temperature and behavior under varying conditions likes voltages and temperatures. These hidden faults mostly lead to rare timing events, and structural scan based or self-testable test techniques are still necessary but not anymore sufficient. They have to be complemented by system level test methods which in turn may require extremely long initialization sequences.



In the realm of automotive electronics these limitations of manufacturing tests are widely known, and international standards define requirements for concurrent and periodic online test or even fault tolerance. Today, general processor design, GPUs and applications in High Performance Computing (HPC) and AI-Giga-Factories are following this trend. Silicon Lifecycle Management (SLM) covers all the phases from the design, prototyping, bring-up, operational phase and wear-out, and especially the last three phases rely on comprehensive collecting all necessary observables by telemetry. The term telemetry denotes not only the collection of data from sensors, monitors, BIST registers, as well as test and fault tolerance actions but also the evaluation, classification and prediction of the system state more and more by means of embedded AI.

First step is always the analysis of the root causes of failures and diagnosis of faulty systems, and the powerful capabilities of today's systems allow the implementation of AI-based analysis algorithms on chip, in edge-computing or in the cloud. Combining high quality manufacturing test and telemetry leads to the required level of reliable computing.



09:00-09:40, May 10



Peony Ballroom



Prof. Gupta, Puneet

Professor of University of California, Los Angeles

Bio.: Puneet Gupta received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Delhi, New Delhi, India, in 2000, and the Ph.D. degree from the University of California at San Diego, San Diego, CA, USA, in 2007. He is currently a Faculty Member with the Electrical and Computer Engineering Department, University of California at Los Angeles. He Co-Founded Blaze DFM Inc., Sunnyvale, CA, USA, in 2004 and served as its Product Architect until 2007. He has authored over 200 papers, 18 U.S. patents, a book and two book chapters in the areas of design-technology co-optimization as well as variability/reliability aware architectures. Dr. Gupta is an IEEE Fellow and was a recipient of the NSF CAREER Award, the ACM/SIGDA Outstanding New Faculty Award, SRC Inventor Recognition Award, and the IBM Faculty Award. He has led the multi-university IMPACT+ Center which focused on future semiconductor technologies. He currently leads the System Benchmarking theme within the SRC CHIMES JUMP 2.0 center.

Title: Scale-out Chiplet Based Systems: Design, Architecture and Pathfinding

Abstract: As conventional technology scaling becomes harder, advanced packaging provides a viable pathway to building larger systems at lower cost. Waferscale chiplet-based systems, as much as 100X larger than largest modern SoCs pose new opportunities and challenges in their architecture and design. We describe a waferscale GPU concept and discuss our experience designing a 2000 chiplet waferscale processor system, pointing out key challenges and solutions. Next, we describe our ongoing work on developing a cross-stack pathfinding framework for large distributed 2.5D/3D systems, identifying areas where technology development would help design metrics substantially, especially for the important class of distributed machine learning training/inference applications.



Workshop

Workshop: Bridging the Gaps: Virtual Fabrication-to-Design Flow for DTCO/STCO



13:30-16:20, May 8



Orchid 4301

13:30-14:00 | Invited Talk

Shen, Chen

Shanghai PFTN Semiconductor Co., Ltd.

Title: Virtual Fabrication Platform for DTCO/STCO

14:05-14:35 | Invited Talk

Liu, Bin

IKAS INDUSTRIES

Title: Embodied Intelligence Reshaping Semiconductor Equipment: A New Paradigm for AI Agent-Driven R&D and Manufacturing Platforms

14:40-15:10 | Invited Talk

Li, Jiang-Wei

Huaxin Jushu (Hangzhou) Microelectronics Co., Ltd.

Title: Enabling DTCO Exploration with Automatic Std-cell Layout Synthesis and Optimization

15:15-15:45 | Invited Talk

Wang, Zhi-Ang

Fudan University

Title: Open-Source System – Technology Co-Optimization Infrastructure for Advanced Integration

15:50-16:20 | Invited Talk

Hu, Jin-hai

IME Singapore

Title: Toward Agile and Intelligent Analog Circuit Design Automation



PANELS

Panel Session 01: Digital Twins for Verification & Manufacturing: Closing the Loop From Model to Fab



13:30–15:30, May 8



Orchid 4302

Why this topic?

Design communities increasingly argue for digital twins of the development process itself—so AI can optimize resource queues, failure triage, and iteration cadence across teams and tools.

IEEE communities spotlight the hard problems: interpretability, validation/verification (VVUQ), safety, adversarial robustness, and scalable human-in-the-loop decisions.

Panelists could explore practical paths to link design-centric twins with yield/metrology data—using International Roadmap for Devices and Systems (IRDS) guidance on lithography, yield enhancement, and sustainability—to reduce respins, tighten process windows, and connect verification results to manufacturing controls.

We'll map where twins belong in 2.5D/3D and chiplet ecosystems (thermal/mechanical co-simulation, parametric drift tracking) and what organizational changes (data contracts, governance) are required to deliver measurable ROI in 12–18 months.

Abstract

Digital twins are increasingly proposed as a way to connect verification evidence, fault behavior, manufacturing data, and process learning into a closed optimization loop. For semiconductor teams, that could mean earlier detection of latent reliability risks, faster root-cause analysis, tighter process control, and better alignment between design assumptions and manufacturing reality, which reduces respins and improves yield. This panel examines whether digital twins can deliver measurable ROI and what governance is needed for adoption.

Key Debate Questions (For vs Against)

Are digital twins a practical solution for advanced packaging and chiplet ecosystems, or an overhyped concept with limited scalability?

Does creating effective digital twins require deeper access to process, metrology, and operational data than what the semiconductor ecosystem is realistically willing to share?

Can digital twins truly reduce verification cycles and improve manufacturability, or will they add cost and complexity without clear benefits?

Should the industry focus first on narrow, high-value twins for failure triage, process control, or packaging, or can broader end-to-end twins across verification and manufacturing be made operational in the near term?

Moderator

Yeo, Kiat Seng, Singapore University of Technology and Design

Panelists

Wunderlich, Hans-Joachim, University of Stuttgart

Schlichtmann, Ulf, Technical University of Munich

Hellebrand, Sybille, Paderborn University

He, Lei, Eastern Institute of Technology, Ningbo

Lianto, Prayudi, Advanced Packaging Development Center, Applied Materials



PANELS

Panel Session 02: Agentic AI + Scalable Compute in EDA: From Point Tools to Flow Orchestration



16:00-18:00, May 8



Orchid 4204

Why This Topic?

Agentic AI and foundation models are rapidly evolving from point assistants into orchestrators of broader design and verification workflows. As these systems move beyond prompt-based assistance toward multi-agent task decomposition, optimization, and closure, they are also reshaping the compute profile of EDA—driving new demands on cloud scalability, heterogeneous infrastructure, and cost-aware scheduling. The next wave of innovation may come not only from better models, but from tighter co-optimization of AI, EDA algorithms, and compute platforms.

Abstract:

Agentic AI is evolving from point assistance into a broader orchestration layer across design, verification, and implementation workflows. At the same time, the growing use of foundation models and the development of multi-agent systems is reshaping the compute profile of EDA, raising new questions about infrastructure scalability, domain specialization, and workflow integration. With perspectives from EDA, cloud infrastructure, academia, and advanced manufacturing, this panel will debate whether the next leap in design productivity will come from better models, tighter coupling to EDA engines, or a new generation of compute-aware design flows.

Will cloud-scale compute become essential for next-generation AI-driven EDA workflows, or will hybrid and on-prem strategies remain the more practical path for most design teams?

Key Debate Questions(For vs Against):

1. Will agentic AI deliver scalable end-to-end EDA productivity gains, or will success depend on domain specialization, compute efficiency, and flow integration?
2. Does agentic AI fundamentally redefine how design teams explore, optimize, and close designs, or is it mainly an incremental acceleration layer on top of existing EDA flows?
3. Will cloud-scale compute become essential for next-generation AI-driven EDA workflows, or will hybrid and on-prem strategies remain the more practical path for most design teams?
4. Should the future of AI-driven EDA be built on large general-purpose foundation models, or on smaller domain-specific models tightly coupled to solvers, simulators and signoff engines?
5. Should EDA vendors expose orchestration frameworks that optimize both tools and compute resources, or will customers demand open infrastructure and model portability?

Moderator

Yeo, Kiat Seng, Singapore University of Technology and Design

Panelists

Kengeri, Subramani, Corporate Vice President and GM, Systems to Materials, Applied Materials

Blaimberger, Frank, VP, Global Innovation Lead for Digitized Compliance and Connected Systems

Chan, Don, Vice president, Research & Development, Cadence Design Systems, Inc.

Brophy, Dennis, Siemens EDA

Ng, James, Synopsys Inc

Bhaskar, Akhil, Core Services Leader, Asia Pacific & Japan at AWS





PANELS

Panel Session 03: Chiplets at Scale: Can Standards, Packaging, and Qualification Enable a True Multi Vendor Ecosystem?



13:30–15:30, May 9



Orchid 4304

Why this topic?

Chiplet architectures are moving from concept to deployment, driven by demand for higher performance, faster productization, and heterogeneous integration in AI, HPC, automotive, and advanced edge systems. Interface standards such as UCle aim to reduce integration friction, but real ecosystem adoption depends on far more than connectivity. Packaging technologies, testability, qualification, manufacturability, and lifecycle management remain critical barriers.

Abstract

Chiplet-based architectures promise modularity, faster innovation, and new paths to heterogeneous integration, but scaling beyond tightly controlled custom designs remains difficult. While interface standards such as UCle aim to reduce integration friction, real multi-vendor deployment also depends on packaging technologies, testability, qualification, manufacturability, and lifecycle management. With perspectives spanning open platform architecture, microelectronics test, design-technology co-optimization, and certification, this panel will debate whether chiplets are ready to become a reusable ecosystem, or whether the industry still lacks the workflows, assurance layers, and business alignment needed for true interoperability at scale.

Key Debate Questions (For vs Against)

Can interface standards such as UCle realistically enable a plug-and-play multi-vendor chiplet ecosystem, or will interoperability still depend on much deeper qualification, certification, and integration work?

Do chiplets require a fundamentally new co-design workflow across architecture, design, packaging, and qualification, or can today's semiconductor development processes evolve incrementally to support them?

Will advanced packaging approaches such as hybrid bonding accelerate mainstream chiplet adoption, or will yield, test, and reliability concerns limit them to a narrower set of high-value applications?

Can chiplets become truly reusable open platforms, or do packaging, power delivery, thermal behavior, and process variation make most integrations inherently application-specific?

Will test, validation, and reliability assurance become the real bottleneck for chiplet adoption, even if interface standards continue to mature?

Moderator

Yeo, Kiat Seng, Singapore University of Technology and Design

Panelists

Benini, Luca, Università di Bologna

Chakrabarty, Krishnendu, Arizona State University

Gupta, Puneet, University of California, Los Angeles

Lin, Yibo, Peking University

Wang, Zhiang, Fudan University





SPECIAL SESSIONS

Special Session on EDA Standards



13:30-18:00, May 8



Orchid 4303

Session Chair



Chen, Jienan

University of Electronic Science and Technology of China

13:30-14:10 | Invited Talk



Brophy, Dennis

Director of Ecosystems, Siemens EDA

Bio.: Dennis Brophy is Vice President of Standards for the IEEE Council on Electronic Design Automation, and serves as Chair of the IEEE Design Automation Standards Committee. He is also Director of Ecosystems at Siemens EDA, leading global ecosystem and interoperability initiatives for digital verification technologies. Dennis

has over 46 years of experience in the electronic design automation industry and holds a B.S. in Electrical Engineering and Computer Engineering from the University of California, Davis.

Title: EDA at Inflection Points: The Critical Role of Open Standards from Languages to Agentic AI

Abstract: Over past five decades, electronic design automation has advanced through a series of inflection points driven by rising complexity and the need for interoperability. From the early emergence of SPICE as an open academic technology, through the standardization of design and verification languages such as VHDL, Verilog, SystemVerilog, and SystemC, open standards have repeatedly enabled EDA innovation to scale globally across tools, companies, geographies and invite academic exploration. This talk reflects on how IEEE standards helped transform EDA from tool-centric silos into a globally connected engineering ecosystem, and draws parallels to today's AI era, where agent-based systems and emerging interoperability technologies such as Model Context Protocol are rapidly shaping new workflows. It also discusses emerging challenges around copyright, data usage, and global regulation, and introduces a new IEEE CEDA standards initiative aimed at providing clarity and guidance as the EDA community works to responsibly adopt AI technologies.



14:10-14:50 | Invited Talk

Blaimberger, Frank L.



Vice President, Global Innovation Lead for Digitized Compliance and Connected Systems from TUV SUD Germany

Bio.: After his Master of Science Degree in Electronics Engineering and Ergonomics at University of Munich (TUM). Before his joining of TUV SUD in June 2019, he worked in several companies such as Siemens, Deutsche Telekom and Fujitsu. His bandwidth of

production & manufacturing related experiences as well as his management skills were reached in various and different positions over 24 years - e.g. head of global manufacturing, heading the shopfloor operation and asset technology at 13 factories / sites for data center equipment and semiconductor production.

Frank has implemented technologies, methods and strategies to enable supply chain transformation - implementing Industry 4.0 and Internet of Things (IoT) operations.

He further collaborate and interact with committees, customers and research organizations to settle and deploy smart factory related references, technologies and standards.

At TUV SUD, he is heading the global lead for digitized compliance innovations focusing on operations of digitized assets, industrial software, CyberSecurity, artificial intelligence and digitized compliance management for connected systems.

Orchestrating of skills and technology is Frank's key to success towards digitized transformation and operation. The Vice President interacts and collaborate with Digital Twin consortiums and AI appliance and infrastructure industry in his daily work at TÜV SÜD.

His goal is to make connected and autonomous operation safe, secure and resilient.

Title: Trustworthiness in the Future of Automation-AI Based Decision Capabilities for Safety and Security

Abstract: AI-based decision-making is rapidly becoming a core capability of automated systems in safety- and security-critical domains. As systems grow in complexity and autonomy, trustworthiness is no longer an attribute of individual components but an emergent property of architectures, processes, and ecosystems across the semiconductor lifecycle. This session explores how trustworthiness can be achieved as a foundation for resilient and fail-operational automation.

The talk addresses how the integration of functional safety, cybersecurity, and Safety of the Intended Functionality (SOTIF) enables efficient and robust system behavior. Emphasis is placed on architectural principles, certified components, and continuous digital lifetime monitoring as key enablers of trust across globally distributed supply chains. Recent developments in standards are highlighted, including the forthcoming IEC 61508-2-1, which extends functional safety concepts to semiconductor technologies and clarifies expectations for safety-critical silicon.

The session further discusses predictive maintenance and early detection of degrading semiconductor faults, aligned with ISO/TR 9839:2023, as well as the qualification and controlled use of Agentic AI during design, verification, and production ramp-up. High-integrity System-on-Chip architectures with onboard safety mechanisms are examined in the context of managing dependent and common-cause failures. The role of digital twins in enabling fail-operational behavior and operational resilience is illustrated, along with the combination of Safety Elements out of Context (SEooC) and continuous context monitoring supported by AI and large language models.

This session provides practical insights for engineers, architects, and decision-makers shaping the future



of trusted automation.

14:50-15:30 | Invited Talk



Zheng, Stanley

Chief Standard, EDA² (Electronic Design Automation Ecosystem Development Accelerator)

Bio.: Prof. Stanley Zheng graduated from the Department of Computer Science at Fudan University and holds an MBA from Macau University of Science and Technology.

He has served as Senior Engineer at Inventec, Engineering Manager at Phoenix Technologies, Technical Support Manager/Strategy Director at Intel, and Director of Semiconductor Industry Development at Huawei. His work experience spans OS development, BIOS/EFI deployment, chip technical support, marketing, ecosystem promotion, IP strategy, 7nm chip product development, and industry development etc. He is the only person in Intel China ever served as a global chip leader. He currently serves as the Director of External Cooperation Committee and Chief Standard at EDA².

Title: Data-driven EDA Standard System

Abstract: Introducing the original intention behind establishing diversified & competitive EDA standard system, its definition, vision, and the efforts to achieve the goal.



SPECIAL SESSIONS

SCIS Special Session on Advanced IC Design and Technology in AI Era



13:30-14:50, May 9



Orchid 4302

Session Chair



Li, Meng

Peking University

Invited Talks

13:30-13:50 | Invited Talk



Yang, Jun

National EDA Center

Bio.: Yang Jun, Professor at Southeast University, currently serves as Executive Director of the National Integrated Circuit Design Automation Technology Innovation Center. As an academic leader in high-efficiency integrated circuit design methodology and EDA, he has long been engaged in design methodology research and now focuses on intelligent EDA, aiming to empower IC design with AI. He has led and participated in multiple national projects, including major national science and technology projects, the National 863 Program, and the National Natural Science Foundation of China. In the past five years, he has published 20 SCI papers as first or corresponding author, holds 76 granted invention patents (including 10 U.S. patents), and has achieved 13 technology transfer projects. He has received the Second Prize of the National Science and Technology Progress Award, the First Prize of the Jiangsu Provincial Science and Technology Progress Award, and the First Prize for Technological Invention from the Chinese Institute of Electronics.

Title: Evaluation of LLM in IC Design

Abstract: Integrated circuit design demands various capabilities of large language models, including multimodal, reasoning, knowledge, and tool-calling capabilities. However, due to the relative scarcity of IC design data, models specialized for IC are currently lacking, and general-purpose models do not provide adequate support for IC design. This report focuses on evaluating the capabilities of various general-purpose models for IC design, providing the audience with a relatively objective impression through comparison, and pointing out future development directions.



13:50-14:10 | Invited Talk



Chang, Chip-Hong

Nanyang Technological University

Bio.: Chip-Hong Chang (Fellow, IEEE) received the BEng (Hons.) degree from the National University of Singapore, in 1989, and the MEng and PhD degrees from Nanyang Technological University (NTU) of Singapore, in 1993 and 1998, respectively. He is a professor of the School of Electrical and Electronic Engineering (EEE), NTU. He held joint appointments with the university as assistant chair of Alumni from 2008 to 2014, deputy director of the Center for High Performance Embedded Systems from 2000 to 2011, and program director of the Center for Integrated Circuits and Systems from 2003 to 2009. He was conferred the Venus International Foundation 2022 Science and Technology Award (VISTA 2022) of Excellence in Hardware Security. He has coedited 6 books, published 13 book chapters, more than 120 international journal papers (>90 are in IEEE), and more than 200 refereed international conference papers (mostly in IEEE), and delivered more than 70 keynotes, tutorials, and invited seminars. His current research interests include hardware security, AI security, biometric security, trustworthy sensing, and hardware accelerators for post-quantum cryptography and edge computational intelligence. He currently serves as the senior editor of IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS) and associate editor of the IEEE Transactions on Very Large Scale Integration (VLSI) Systems. He also served as the senior area editor of the IEEE Transactions on Information Forensic and Security, associate editor of the IEEE Transactions on Information Forensic and Security, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, IEEE Access, and IEEE Transactions on Circuits and Systems-I for more than one term. He guest edited more than 10 journal special issues and served in the organizing and technical program committee of more than 70 international conferences (mostly IEEE). He is an IET fellow, AAIA fellow, and 2018–2019 distinguished lecturer of IEEE Circuits and Systems Society.

Title: Silicon Last Stand: Security in the Artificial Intelligence of Everything (AIoE) Era

Abstract: The emergence of the Artificial Intelligence of Everything (AIoE) has triggered a fundamental 'Trust Paradox'. As autonomous intelligence migrates to the extreme edge, our systems are becoming more sophisticated, yet increasingly vulnerable to physical and supply-chain weaponization. In this landscape, the chip serves as the Silicon Last Stand—the final, verifiable frontier of trust.

This talk explores the shift from reactive, software-defined security to proactive, hardware-enforced resilience. We examine the evolution of cyber-physical threats, from smart infrastructure manipulation to the 'cognitive sabotage' of AI models, where the 'Crown Jewels' are no longer just cryptographic keys, but the integrity of the AI's interpretation of reality.

By navigating the 'Open Supply Chain Paradox', this session introduces a Double Helix framework: hardware as the secure sanctuary for AI, and AI as the intelligence to build trustworthy silicon. We advocate for a Zero Trust Architecture rooted in physical primitives, proving that hardware security is no longer a marginal expense, but the foundational investment required to ensure that in our rush to connect AI to the world, we do not leave it unprotected.



14:10-14:30 | Invited Talk



Xu, Qiang

The Chinese University of Hong Kong

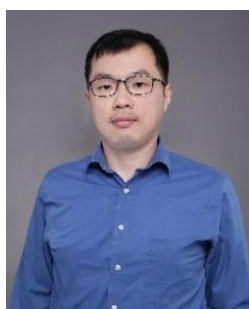
Bio.: Qiang Xu is a Professor of Computer Science & Engineering at The Chinese University of Hong Kong. He previously served as the Chief Scientist at the National Technology Innovation Center for EDA, China. His research interests span AI and Design Automation. With over 200 publications and more than 18,000 citations, his work has garnered multiple best paper awards at prestigious conferences, including an ICCAD Ten Year Retrospective Most Influential Paper Award and an AAAI Most Influential Paper. He has supervised 20 Ph.D. dissertations, many of which have received Outstanding Ph.D. Dissertation Awards or Nominations.

Title: Architecting for Jagged Intelligence

Abstract: The semiconductor industry has mastered deterministic reliability, yet modern AI introduces a sharply contrasting paradigm: "Jagged Intelligence." Powered by Large Language Models, autonomous AI agents exhibit advanced but fundamentally probabilistic capabilities—excelling at complex reasoning while frequently failing at basic logic. This unpredictability creates a severe reliability bottleneck for mission-critical applications.

To move beyond fragile prompt engineering, we must adopt a "Governance-First" paradigm. The presentation introduces a neuro-symbolic kernel designed to enforce strict reliability policies. By decoupling the probabilistic reasoning engine from a deterministic, logic-based governor, this architecture establishes a verifiable firewall that transforms unpredictable AI outputs into reliable, structured actions, laying the foundation for trustworthy agentic computing.

14:30-14:50 | Invited Talk



Wu, Heng

Peking University

Bio.: Heng Wu is an Associate Professor in the School of Integrated Circuits at Peking University. He received the Ph.D. degree in Electrical and Computer Engineering from Purdue University. After graduation, he joined IBM T. J. Watson Research Center in Albany N.Y. USA as a research staff member. After that, He joined Peking University in early 2023. His research focuses mainly on high speed and low power CMOS devices and circuits, transistors architecture and design technology co-optimization (DTCO). Prof. Wu has published more than 100 papers and received 5 best paper awards (including VLSI symposium). He has filed more than 500 patents (with 250 patents granted). He is an IBM Master Inventor and recipient of IEEE Paul Rappaport Award. He has served on the technical committees of major conferences such as: IEEE EDTM, IEEE ISQED, IEEE ICSICT and SSDM. He is a member of IEEE



Semiconductor Manufacturing Committee.

Title: Extending Moore's Law from Wafer Backside: from Interconnect to Device Integrations

Abstract: Moore's Law is challenged by front-side integration bottlenecks like density bottlenecks, wiring congestion and parasitic loss. Wafer backside integration emerges as a key solution to extend it. This talk focuses on the backside integration of interconnects and active transistors. We will elaborate the scaling of GAA by utilizing backside interconnects including BSPDN, BS clock and BS signal, then introduces high density stacking of transistors from backside in the aspect from device concept to design enablement. The integration scheme of backside interconnect and devices, and preliminary verification of dual-sided logic circuits will also be presented. Finally, we will prospect the technical challenges and prospects of backside integration for advanced logic nodes.



TUTORIALS

Tutorial 01: Parallel Sparse Direct Solver for SPICE

🕒 13:30-16:30, May 10

📍 Orchid 4301

Tutorial Description: High-performance sparse linear solvers play a critical role in enabling fast and precise transistor-level circuit simulation and verification. Driven by rapid advances in semiconductor technology, modern integrated circuits have reached unprecedented levels of complexity, often incorporating hundreds of millions of devices. As a result, sparse linear solvers have become increasingly resource-intensive in terms of both runtime and memory consumption during simulation. Additionally, circuit matrices typically feature extremely sparse structures with highly irregular nonzero patterns, further complicating efforts to achieve efficient acceleration. This tutorial focuses on advanced sparse direct solver techniques specially optimized for transistor-level circuit simulation. The tutorial will comprehensively present state-of-the-art algorithms, optimization methods, and parallelization schemes, to tackle the key challenges in designing contemporary sparse direct solvers for circuit simulation.

Here is an expected outline of this tutorial (1.75-2 hours).

1. Background of circuit simulation and challenges of designing sparse direct solvers.
2. Basics of sparse direct solvers.
3. Flow of sparse direct solver for SPICE.
4. Parallelization methodologies of factorization
5. Parallelization methodologies of substitution.
6. Sparsity-aware optimization techniques.
7. Usage example of sparse direct solver in SPICE.
8. Test results.

Organizer/Speaker

Chen, Xiaoming

Institute of Computing Technology, Chinese Academy of Sciences

Bio.: Xiaoming Chen is a Full Professor of Institute of Computing Technology, Chinese Academy of Sciences. He received the BS and PhD degrees in Electronic Engineering from Tsinghua University in 2009 and 2014, respectively. His research interests include design automation for integrated circuits and computer architectures. He has published 1 Springer book and over 150 papers in DAC, ICCAD, DATE, MICRO, HPCA, ASPLOS, IEEE TCAD, IEEE TC, IEEE TPDS, etc. His researches have been adopted by commercial EDA software. He has consistently been ranked among the World's Top 2% Scientists since 2022. He was a recipient of the 2021 NSFC Excellent Young Scientists Fund, 2016 European Design and Automation Association (EDAA) Outstanding Dissertation Award, 2018 DAMO Academy Young Fellow Award, 2023 Hot Paper Award of SCIENCE CHINA Information Sciences, and ASP-DAC 2022 Best Paper Award.



TUTORIALS

Tutorial 02: ECOS Studio: Building an Open Ecosystem for Silicon Design and Education

🕒 13:30-16:30, May 10

📍 Orchid 4302

Tutorial Description: This tutorial introduces an emerging open silicon design ecosystem that aims to significantly lower the barriers to custom chip development for researchers, engineers, and students. While open hardware initiatives have expanded rapidly in recent years, practical adoption is often limited by fragmented open-source EDA tools, limited access to manufacturable process technologies, and the complexity of the full RTL-to-chip design flow. Addressing these challenges requires not only individual tools or IP blocks, but an integrated ecosystem that connects design automation, reusable IP, fabrication access, and education. The tutorial presents a holistic approach to open chip design centered on ECOS Studio, an integrated RTL-to-chip framework that combines open-source EDA tools, an open IP library, and an open-source manufacturable process design kit into a unified workflow. In addition, the tutorial will demonstrate how the toolchain can be used in practice to implement a chip starting from RTL, illustrating the key steps of the design flow and the practical considerations involved in bringing a design to silicon. The goal is to make silicon development more accessible and agile, enabling users to move from RTL to fabricated chips with substantially reduced effort.

Beyond the technical infrastructure, the tutorial also discusses ecosystem-level initiatives that promote large-scale participation in silicon design, including education-oriented chip design programs and SoC development. These efforts illustrate how open ecosystems can accelerate innovation, expand access to silicon prototyping, and foster a broader community around open hardware design.

Organizer/Chair(s)

Xie, Biwei, Institute of Computing Technology, Chinese Academy of Sciences

Liu, Tianji, Institute of Computing Technology, Chinese Academy of Sciences

Speakers

Zhao, Xueyan

Institute of Computing Technology, Chinese Academy of Sciences

Bio.: Xueyan Zhao is currently a fifth-year Ph.D. student at the Institute of Computing Technology, Chinese Academy of Sciences (ICT, CAS). His research interests include open-source EDA toolchains, physical design, as well as placement and timing optimization. He is a member of the ECOS Studio project, focusing on the development of the open-source EDA toolchain. He has achieved distinguished success in prestigious international EDA competitions, including 1st Place in the 2021 SIGDA CADathlon at ICCAD, 1st Place in the 2022 ICCAD CAD Contest, and 2nd Place in the 2023 ICCAD CAD Contest.

Title: A 55nm Digital Design Methodology Enabled by Open-Source PDK and EDA Tools

Liu, Tianji

Institute of Computing Technology, Chinese Academy of Sciences

Bio.: Tianji Liu is an Assistant Professor at the Institute of Computing Technology, Chinese Academy of Sciences. He received his Ph.D. degree from The Chinese University of Hong Kong. His research interests are parallel and efficient algorithms for logic synthesis and verification. He is a recipient of the Best Paper Award at ASP-DAC 2024, as well as Best Paper Award candidates at DAC 2023 and ICCAD 2025.

Title: Enabling Chip Design Education and SoC Development with Open-Source EDA



TUTORIALS

Tutorial 03: AI for EDA Feedback Loops: Leveraging Manufacturing Data to Evolve PDKs, Models, and Design Rules



13:30-16:30, May 10



Orchid 4303

Tutorial Description: Modern Electronics Design Automation (EDA) flows rely heavily on Process Design Kits (PDKs), compact models, and design rules that abstract complex silicon behavior into usable forms for simulation and verification. However, these abstractions are often treated as static, even as real manufacturing processes exhibit variability, drift, and previously unseen failure modes. This tutorial introduces a paradigm shift where AI and machine learning are used to close the loop between fabricated silicon and design environments. By leveraging large-scale manufacturing data—such as electrical test results, in-line metrology, defect maps, and process time-series—participants will learn how data-driven models can uncover hidden patterns, quantify variability, and systematically refine the parameters and assumptions embedded within PDKs and design rules. The tutorial will present practical AI workflows that translate manufacturing insights into actionable updates for EDA, including variability-aware compact model calibration and data-driven design rule refinement. Through case studies inspired by real semiconductor datasets, attendees will see how techniques such as regression, classification, and anomaly detection can be repurposed to inform SPICE model updates, identify layout hotspots, and enable continuous learning in design flows. By the end of the session, participants will gain a clear understanding of how AI can transform EDA from a static, assumption-driven process into a dynamic, feedback-enabled ecosystem grounded in silicon reality.

Organizer/Speaker

RAGHAVAN, NAGARAJAN

Singapore University of Technology and Design

Bio.: NAGARAJAN RAGHAVAN is an Associate Professor at the Singapore University of Technology and Design (SUTD) in the Engineering Product Development (EPD) pillar. Prior to this, he was a postdoctoral fellow at the Massachusetts Institute of Technology (MIT) in Cambridge and at IMEC in Belgium, in joint association with the Katholieke Universiteit Leuven (KUL). He obtained his Ph.D. (Microelectronics, 2012) at the Division of Microelectronics, Nanyang Technological University (NTU), Singapore.

His work focuses on reliability modeling and characterization of advanced logic, III-V and NVM devices and AI-enabled design for manufacturing and reliability of electronic devices. He is the recipient of the SUTD Outstanding Researcher Award for 2022, IEEE Electron Device Society (EDS) Early Career Award for 2016, Asia-Pacific recipient for the IEEE Electron Device Society PhD Student Fellowship in 2011 and the IEEE Reliability Society Graduate Scholarship Award in 2008. To date, he has authored / co-authored more than 320 international peer-reviewed publications and five invited book chapters. He is an Associate Editor of the IEEE Access and Microelectronic Engineering journals and serves in the Editorial Advisory Board of APL Machine Learning. He currently manages a research group with 5 PhD students and 4 research scientists.





TUTORIALS

Tutorial 04: CMOS Technology and Its Application to Design Multi-Standard RF/mm-Wave Low-Noise Amplifiers



13:30-16:30, May 10



Orchid 4304

Tutorial Description: Big challenges lie ahead for chip designers when they choose to develop ICs using silicon technology for low power and high data rate applications. This is because silicon technology suffers from undesirable energy dissipation due to its lossy substrate and high resistive wiring loss at RF/mm-Wave frequencies. Nonetheless, silicon remains the most suitable material satisfying the demands of a rapidly growing semiconductor market through low fabrication cost and ease of achieving system-on-chip or system-in-package integration.

The transistor promises to revolutionize electronics; indeed, it has become an integral and essential part of our life. A key aspect of chip design is the characterization and modelling of passive devices. For example, physical design parameter optimization of on-chip inductors is performed to achieve similar inductance values as an experimental control to investigate how its core diameter, conductor spacing, and width affect their peak quality factors. The use of these optimized inductors is demonstrated to enhance the circuit characteristics of a giga-hertz amplifier.

While metallization in the form of inductors are “friends” that store magnetic energy, as interconnects, they are however, “foes” to RF/mm-Wave IC designers. Without considering these parasitics from interconnects in the design phase, fabricated RF/mm-Wave circuits suffer power loss and shifts in circuit operating frequencies. A new figure of merit, intrinsic factor, IF, has been proposed in this work to provide a convenient quantitative indication as to how interconnects affect the performance of RF/mm-Wave ICs.

One of the key challenges in realizing multi-standard communications is how to design low-noise amplifiers (LNAs) that can operate in different frequency bands with less inductor number and small inductor value. Based on the above-mentioned concerns, the investigation on the designs of dual-band LNAs, as well as a modified architecture used for input matching in CMOS LNAs to make the circuit more compact, is presented. The proposed design approach can achieve compact and fully integrated LNAs for multi-standard communications. This tutorial covers 3 topics: 1) The Past, Present and Future of Semiconductors, Global Semiconductor Market, Early Computers, RF Revolution and CMOS Future Directions; 2) Characterization and Modeling of Spiral Inductors, Circuit Verification of Inductor Design Methodology, Characterization and Modeling RF Interconnects, and RF Interconnects Model Verification; and 3) Review of LNA Designs, A New Input Architecture for CMOS LNAs, A Narrow-Band LNA Design Using the New Input Architecture, and Dual-Band LNA Design Using the New Input Architecture.

This tutorial is aimed at students, instructors, IC designers, engineers, and researchers working in CMOS technology and RFIC design. It will enhance their knowledge of the subject, and inspire them to pursue the subject further. The techniques and methodologies presented in this tutorial can help IC designers to reduce the design cycle time and enable them to achieve first pass silicon success.



Organizer/Speaker

Yeo, Kiat Seng

Singapore University of Technology and Design

Bio.: Professor Kiat Seng YEO, (PhD, PPA, PBS, FSAEng, FSNAS, FCAE, FIEEE, FAAET, FAIIA, FAAIA) received his B.Eng. (EE) in 1993, and Ph.D. (EE) in 1996 both from Nanyang Technological University (NTU), Singapore. He has 37 years of experience in industry, academia, startups and consultancy. Prof. Yeo is a widely known authority in low-power RF/mm-wave IC design and a recognized expert in CMOS technology. He has made many outstanding contributions to advance Singapore's education and research ambitions over the course of his career. As the Founding Director of VIRTUS (now Centre for Integrated Circuits and Systems), a S\$52M National IC Design Centre of Excellence jointly funded by NTU and Singapore Economic Development Board (EDB), he contributed extensively to the economic development of integrated circuit design in Singapore by leading multidisciplinary research, with a focus on industry collaboration. He has secured over S\$70M of research grants as PI, published 14 books (4 Amazon Best Sellers), 7 book chapters, 700+ journal and conference papers, and holds 55 patents. Prof. Yeo holds/held key positions in many international conferences as Advisor, General Chair and Technical Chair. He is one of the rare 7 "double academicians" of the Singapore Academy of Engineering (SAEng) and the Singapore National Academy of Science (SNAS). He is also a Foreign Fellow of the Canadian Academy of Engineering (CAE) and Fellows of the ASEAN Academy of Engineering and Technology (AAET), the Asia-Pacific Artificial Intelligence Association (AAIA), the International Artificial Intelligence Industry Alliance (AIIA), and the IEEE for his contributions to low-power integrated circuit design. He received 2 National Day Awards from the President of Singapore in 2009 (PPA) and 2020 (PBS), and awarded the Nanyang Alumni Achievement Award by NTU in 2009. Prof. Yeo is the principal author of Integrated Circuit Design Research Ranking for Worldwide Universities 2008 and World University Research Rankings (WURR) 2020. He was recognized among the Top 2% Scientists Worldwide by Stanford University from 2020 to 2025, World's AI Top Scientist by the International Artificial Intelligence Industry Alliance in 2023, and Top Scholar by ScholarGPS from 2023 to 2025.



TUTORIALS

Tutorial 05: Design Methodologies for Intelligent & Perceptive Hardware Security - From Circuits to Machine Learning Algorithms



13:30-16:30, May 10



Orchid 4201

Tutorial Description: The rapidly expanding attack surface of connected devices and the decreasing cost of hardware security attacks now demand ubiquitous deployment of security countermeasures down to edge silicon systems. In this tutorial, the on-going and prospective evolution of design methodologies for on-chip hardware security is presented through an original macro-trend analysis, based on a database that is shared with the attendees as part of the tutorial toolkit. The tutorial covers recent and promising directions to protect confidential information/data in intelligent sensors (e.g., on-chip AI model) from network down to physical level. As emerging security trend towards more self-reliant edge devices, on-chip intelligent sensors (intelligent & perceptive) with built-in AI to autonomously detect and respond to attacks are discussed.

Design methodologies for on-chip sensing, attack detection, actuation and attack counteraction are introduced for key classes of attacks from non-invasive to invasive. A wide range of methodologies and architectures of on-chip frontends and actuators is introduced to inexpensively monitor the chip environment continuously, gaining physical context awareness and capturing physical anomalies. This trend is progressively fusing with relentless advances in inexpensive on-chip (AI) intelligence, which can oversee physical signals and events to orchestrate the on-chip security ecosystem.

The road towards ubiquitous intelligent & perceptive hardware security countermeasures is illustrated through the analysis of recent silicon demonstrations with unprecedented capabilities to perceive security events inexpensively, and react to them intelligently, all the time (always-on). The new concept of hardware patching is also discussed where circuit flexibility is introduced to make silicon chips able to evolve over time, and counteract newly discovered vulnerabilities through (machine) learning-based physical protection mechanisms.

OUTLINE/SCHEDULE

- HW security challenges towards trillions of connected devices
 - Edge attack surface extending exponentially
 - Ubiquitous HW security
 - HW security challenges towards the trillions
- Immersed-in-logic and in-memory security primitives (fully-digital fully-automated design)
 - Root of trust and chain of trust
 - On-chip intrinsic root of trust
 - State of the art review through the "hardware security database"
 - Primitive architectures for ubiquitous security
- On-chip sensorization for physical attack detection (fully-digital fully-automated design)
 - Side-channel attack detection
 - Laser Voltage Probing (LVP) attacks and on-chip detection
- Machine learning-based counteraction of side-channel attacks: hardware patching and self-learning
 - HW patchability via ML-based counteraction
 - Neural net reverse engineering counteraction under voltage scaling
 - Online learning-based countermeasure against power analysis attacks
- Conclusions



Organizer/Speaker

Alioto, Massimo

National University of Singapore

Massimo Alioto is Provost's Chair Professor at the ECE Department of the National University of Singapore, where he leads the Green IC group, the Integrated Circuits and Embedded Systems area, and the FD-fAbrICS center on intelligent&connected systems. Previously, he held positions at the University of Siena, Intel Labs – CRL (2013), University of Michigan - Ann Arbor (2011-2012), University of California – Berkeley (2009-2011), EPFL - Lausanne.

He is (co)author of 400 publications on journals and conference proceedings, and four books with Springer (with two more coming). His primary research interests include ultra-low power and self-powered systems, green computing, circuits for machine intelligence, hardware security, and emerging technologies.

He was the Editor in Chief of the IEEE Transactions on VLSI Systems and Deputy Editor in Chief of the IEEE Journal on Emerging and Selected Topics in Circuits and Systems. He was the Chair of the Distinguished Lecturer Program for the IEEE CAS Society, and Distinguished Lecturer for the SSC and CAS Society. He is currently member of the SSC AdCom, and previously member of the Board of Governors and the Chair of the "VLSI Systems and Applications" Technical Committee of the IEEE Circuits and Systems Society (2010-2012). He served as Guest Editor of numerous journal special issues (JSSC, TCAS-I, JETCAS...), Associate Editor (currently in JSSC), Technical Program Chair of several IEEE conferences (ISCAS, SOCC, PRIME, ICECS), and TPC member (ISSCC, ASSCC). His research group contribution has been recognized through various best paper awards (e.g., ISSCC), and in the ten technological highlights of the TSMC annual report, among the others. Prof. Alioto is an IEEE Fellow.



TUTORIALS

Tutorial 06: Design of Energy-Constrained AI Accelerators for Edge Computing

 13:30-16:30, May 10

 Orchid 4202

Tutorial Description: This tutorial presents a comprehensive overview of the design methodologies and emerging techniques for energy-constrained AI accelerators targeting edge computing applications. As artificial intelligence continues to move from cloud-centric platforms to resource-limited edge devices, stringent constraints on energy consumption, area, and latency necessitate fundamentally new approaches to both hardware design and algorithm development. This tutorial is organized into two complementary parts: (1) computing-in-memory (CiM) architectures for energy-efficient AI acceleration, and (2) hardware–algorithm co-design strategies for embodied AI systems.

The first part of the tutorial focuses on computing-in-memory as a promising paradigm to overcome the energy bottleneck associated with conventional von Neumann architectures. By integrating memory and computation within the same physical location, CiM significantly reduces data movement, which is a dominant contributor to energy consumption in modern AI workloads. The tutorial will cover key CiM design principles, including SRAM- and emerging memory-based implementations, analog versus digital CiM trade-offs, and circuit-level challenges such as nonlinearity, device variation, and limited precision. Practical design techniques for improving energy efficiency and robustness—such as variation-aware analog-to-digital conversion, calibration methods, and quantization-aware design—will be discussed.

The second part of the tutorial addresses hardware–algorithm co-design for embodied AI, where intelligent agents interact with dynamic physical environments under tight energy and latency constraints. Unlike traditional AI workloads, embodied AI requires continuous perception, decision-making, and closed-loop control, often in real time and directly on edge devices. This makes it essential to jointly optimize algorithms, architectures, and hardware implementations in order to achieve efficient end-to-end system performance. This part will highlight representative co-design methodologies and case studies spanning multiple embodied sensing modalities and application scenarios. Topics will include: (1) hardware–algorithm co-optimization for real-time 2D/3D vision-based gesture recognition chips for human–computer interaction; (2) the design of low-latency vision chips for fast-moving object detection and tracking based on both frame-based CMOS image sensors and event-driven dynamic vision sensors; (3) event-triggered tactile sensing and processing systems for low-power, scalable robotic e-skins; and (4) an energy-efficient self-organizing map (SOM) processor based on computing-in-memory (CiM) architecture for continuous on-device adaptation to new knowledge. Through these examples, the tutorial will illustrate how embodied AI workloads introduce unique requirements in sensing, data representation, temporal processing, adaptability, and hardware efficiency, and how co-design across the algorithm, architecture, and circuit levels can address these challenges.

Organizer/Speaker(s)

Kim, Tony Tae-Hyoung

Nanyang Technological University

Bio.: Prof. Tony T. Kim received the B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, Korea. He received the Ph.D. degree in electrical and computer engineering from the University of Minnesota,



Minneapolis, MN, USA in 2009. From 2001 to 2005, he worked for Samsung Electronics where he performed research on the design of high-speed SRAM memories, clock generators, and IO interface circuits. In 2007 ~ 2009 summer, he was with IBM T. J. Watson Research Center and Broadcom Corporation where he performed research on isolated NBTI/PBTI measurement circuits and SRAM Mismatch measurement test structure, and battery-backed memory design, respectively. In November 2009, he joined Nanyang Technological University where he is currently an associate professor. His current research interests include in-memory computing for edge computing, emerging memory circuit design, energy-efficient AI accelerators, variation tolerant circuits and systems, and circuit techniques for 3-D ICs.

Title: Computing-in-memory for Energy Constrained AI accelerators

Lu, Yuncheng

Nanyang Technological University

Bio.: Yuncheng Lu received the Ph.D. degree from Nanyang Technological University, Singapore, in 2024. He is currently a Research Fellow with Nanyang Technological University. In 2018, he joined Shenzhen HiSilicon Semiconductor Co., Ltd., Shenzhen, China, as a Digital IC Design Engineer. In 2019, he joined the Center for Integrated Circuits and Systems, Nanyang Technological University, as a Research Associate. He has authored and co-authored around 30 papers in top-tier venues in solid-state circuit design, including ISSCC, CICC, ASSCC, ICCAD, JSSC, TCAS-I/II, and TVLSI. His research interests include energy-efficient AI accelerator design and real-time low-power processor design for embodied AI.

Title: Hardware-Algorithm Co-design for Embodied AI



TUTORIALS

Tutorial 07: Sub-THz Circuits and Systems for Radar Transceivers

 13:30-16:30, May 10

 Orchid 4203

Tutorial Description: With sub-THz frequencies becoming increasingly important for high-resolution radar, imaging, and emerging sensing platforms, there is a timely need for a tutorial dedicated specifically to radar transceiver design beyond conventional mm-wave operation. This tutorial introduces the fundamentals of sub-THz circuits and systems for radar transceivers, with emphasis on wideband FMCW/chirp generation, low-phase-noise LO and frequency synthesis, transmitter and receiver design, phased-array beamforming, beam steering, calibration, and chip-package-antenna co-design. Particular attention will be given to the circuit and system bottlenecks that dominate above 100GHz, including limited device gain, output-power constraints, receiver noise, phase-noise accumulation, broadband signal distribution, on-chip/AiP consideration with loss, and packaging-aware integration. By focusing on radar-oriented sub-THz transceivers and their end-to-end system realization, this tutorial is positioned distinctly from recent ISSCC tutorials that emphasized broader wireless communication topics or general mm-wave design practices, and instead addresses the specific architectural and implementation challenges of fully integrated sub-THz radar systems.

Organizer/Speaker

Zheng, Yuanjin

Nanyang Technological University

Bio.: Yuanjin Zheng received the B.Eng. and M.Eng. degrees from Xi'an Jiaotong University, Xi'an, China, in 1993 and 1996, respectively, and the Ph.D. degree from Nanyang Technological University (NTU), Singapore, in 2002. From July 1996 to April 1998, he was with the National Key Laboratory of Optical Communication Technology, University of Electronic Science and Technology of China, Chengdu, China. From 2001 to 2009, he was with the Institute of Microelectronics, A*STAR, Singapore, where he served as a Technical Manager and Senior Scientist. Since 2009, he has been with NTU, Singapore, where he is currently a Professor, the MediaTek Endowed Professor in Integrated Circuit Design, the Director of the Centre for Integrated Circuits and Systems (CICS) and the IC Design Center of Excellence, the Program Director of VALENS Bio-Instrumentation, Devices and Signal Processing Centre of Excellence, and a Co-director of the Schaeffler-NTU Corporate Lab. He has authored or coauthored more than 650 journal and conference papers, 30 patents, and several book chapters. His research interests include RF, mm-wave and sub-THz integrated circuits and systems, radar and UWB transceivers, biomedical circuits and imaging systems, acoustic/MEMS/SAW devices, and AI-assisted sensing systems. Dr. Zheng has served as an Associate Editor for several journals, including the IEEE Transactions on Biomedical Circuits and Systems, IEEE Transactions on Circuits and Systems, and the IEEE Journal of Electromagnetics, RF, and Microwaves in Medicine and Biology.



BEST PAPER CANDIDATES

The Best Paper Award winners will be announced at the banquet on May 8th.

Paper ID: 33

Paper Title: Reducing SQED Duplicate Mode via Processor-Level Abstraction

Authors: Yufeng Li (Institute of Computing Technology, Chinese Academy of Sciences), Linze Li, Qiusong Yang, Yiwei Ci, Enyuan Tian (Institute of Software, Chinese Academy of Sciences), Xuying Zhao, Yungang Bao, Kan Shi (Institute of Computing Technology, Chinese Academy of Sciences)

Paper ID: 72

Paper Title: LEGO: An LLM Skill-Based Front-End Design Generation Platform

Authors: Jincheng Lou, Ruohan Xu (Peking University), Jiecheng Ma (Xidian University), Runzhe Tao, Xinyu Qu, Yibo Lin (Peking University)

Paper ID: 103

Paper Title: KAN-CHAR: Efficient Cell Library Characterization for DTCO via Kolmogorov-Arnold Graph Learning

Authors: Shubo Xu, Zeyuan Deng, Zhi Li, Peng Cao (Southeast University)

Paper ID: 118

Paper Title: ELITE: Efficient Lookup Table Assisted Routing Engine for Photonic Integrated Circuits

Authors: Xiaofei Yu, Yuchao Wu, Haopeng Yan, Yeyu Tong, Yuzhe Ma (The Hong Kong University of Science and Technology (Guangzhou))

Paper ID: 133

Paper Title: CoCIM: A Collaborative Design Framework for Multi-Network Shared Compute-in-Memory Architectures

Authors: Xiaoxiao Zhou, Wanqian Li, Rui Liu, Yuping Wu, Xinjiang Zhang, Xueqi Li, Xiaoming Chen (Institute of Computing Technology, Chinese Academy of Sciences)

Paper ID: 148

Paper Title: IC-MoE: Interface-Constrained MoE for Fast Thermal Evaluation in 3D ICs

Authors: Xiangqiao Meng (The Hong Kong Polytechnic University), Wenkai Yang (BTD Technology), Siyuan Miao (University of California, Los Angeles), Jinlong Yan (Eastern Institute of Technology), Rumin Zhang, Chen Wu (Engineering Research Center of Chiplet Design and Manufacturing of Zhejiang Province, Ningbo Institute of Digital Twin), Lei He (Eastern Institute of Technology)

Paper ID: 150

Paper Title: CoPU-RL: A Hardware-Friendly Training-Inference Framework For Neuro-Symbolic Reinforcement Learning

Authors: Fengyuan Guo, Hongxiao Zhao, Jiayi Li, Anjunyi Fan, Yihan Fu, Wenshuo Yue, Bonan Yan (Peking University)

Paper ID: 157

Paper Title: Automated Synthesis of Abutable Macros for Analog Layout

Authors: Li Huang (Shanghai Jiao Tong University, Eastern Institute of Technology), Xueying Huang (Eastern



Institute of Technology), Xinpeng Wang (Tsinghua University), Fuya Yang, Junjie Yang, Lin Yang (BTD Technology), Chuanxu Liu, Rui Ye, Zhanfei Chen, Ting-Jung Lin (Ningbo Institute of Digital Twin), Lei He (University of California, Los Angeles)

Paper ID: 162

Paper Title: DTCO Dual-Sided Standard Cells with DP-Based Optimization and Custom Library Development

Authors: Qinzhe Zhi, Peijie Ma, Heng Wu, Tianyu Jia (Peking University)

Paper ID: 219

Paper Title: Flexible Double-Side Pin Redistribution with Efficient Virtual-Net-Guided Cell Type Assignment

Authors: Xun Jiang, Haoran Lu, Yifan Chen, Yibo Zhang, Jianxiang Jin, Jiarui Wang, Chunyuan Zhao, Heng Wu, Runsheng Wang, Yibo Lin (Peking University)

Paper ID: 257

Paper Title: An MLIR-Based Compilation Flow for CGRAs: Optimized CDFG Generation, Control-Flow Handling, and Scalability

Authors: Jiahang Lou, Jiayao Hu, Jianrong Zhang, Yikuan Chen, Zewei Zhong, Yuan Dai, Wenbo Yin, Lingli Wang (State Key Laboratory of Integrated Chips and Systems, Fudan University), Yu He (Hangzhou Phlexing Technology Co., Ltd), Qing He (Tongji University)

Paper ID: 267

Paper Title: HeteroPower: A CPU-GPU Heterogeneous Engine to Accelerate Gate-Level Power Analysis

Authors: Xizhe Shi, Zizheng Guo, Yun Liang, Runsheng Wang, Yibo Lin (Peking University)



TECHNICAL SESSIONS

Technical Session 01: Heterogeneous & Optical Compute-Infrastructure Innovations

🕒 13:30-14:50, May 8, 2026

📍 Orchid 4304

👤 Session Chair: **Yan, Hao**, Southeast University

13:30-13:50 | Paper ID: 109

Title: Differentiable Timing-Aware Tier Partitioning with Hypergraph Extension for Heterogeneous 3D ICs

Author(s): Xilin Qian, Han Zhang, Peng Cao

Presenter: **Qian, Xilin**, Southeast University

Abstract: Heterogeneous 3D ICs expand the power, performance, area, and cost (PPAC) design space by combining tiers with different technology characteristics. Their benefits, however, depend on tier partitioning, which must assign cells to suitable tiers while jointly balancing timing, cross-tier interconnects, and local area-balance constraints. Existing partitioners are either designed for homogeneous tiers or rely on sequential heuristics, which often leads to mismatched objectives and suboptimal solutions. To address this issue, we propose a differentiable tier partitioning framework based on an extended circuit hypergraph with virtual edges weighted by slack, enabling jointly optimization of timing, cross-tier interconnects, and area balance in a unified global optimization space. Compared with state-of-the-art academic solutions on five OpenCores benchmarks, our method guarantees minimum improvements of 18% in total negative slack (TNS), 1% in power delay product (PDP), while reducing 36% hybrid bonding terminals (HBTs).

13:50-14:10 | Paper ID: 118

Title: ELITE: Efficient Lookup Table Assisted Routing Engine for Photonic Integrated Circuits

Author(s): Xiaofei Yu, Yuchao Wu, Haopeng Yan, Yeyu Tong, Yuzhe Ma

Presenter: **Yu, Xiaofei**, The Hong Kong University of Science and Technology (Guangzhou)

Abstract: Optical interconnects have emerged as a promising solution for rack-, board-scale, and even in-package communications, thanks to their high available optical bandwidth and minimal latency. However, optical waveguides are intrinsically different from traditional metal wires, especially in terms of phase-matching constraints, which pose new challenges for routing in photonic integrated circuit (PIC) design. In this paper, we present ELITE, a comprehensive and efficient lookup table (LUT)-assisted routing engine tailored for PICs with phase-matching constraints. We integrate a customized A*-based initial routing algorithm that accounts for potential waveguide crossings, a diffuse-based length-matching technique, and LUT-based strategies that leverage a pre-defined path library for bend reduction and strategic insertion to ensure precise phase alignment. The framework minimizes total insertion loss while adhering to all matching constraints. Experimental results based on real photonic designs show that our optical routing flow can reduce total insertion loss by 25% and maximum insertion loss by 46% compared to manual results, while satisfying matching constraints.

14:10-14:30 | Paper ID: 132

Title: Accuracy-Configurable Floating-Point Multiplier Design for SRAM-Based Compute-in-Memory

Author(s): Yiqi Zhou, Junhao Lu, Jiale Yu, Zhuo Xu, Yang He, Yue Yuan, Shan Shen, Daying Sun

Presenter: **Zhou, Yiqi**, Nanjing University of Science and Technology

Abstract: Digital Compute-in-Memory (DCiM) reduces data movement and has become a promising solution for energyefficient edge AI. However, most existing DCiM frameworks still primarily target integer or



fixed-point arithmetic, and provide limited support for compiler-integrated and accuracyconfigurable floating-point computation. Directly integrating conventional IEEE 754 floating-point units into dense SRAM-based DCiM arrays, however, incurs high area and power overhead. To address this challenge, this work presents an accuracyconfigurable floating-point multiplier integrated into the OpenACM framework for SRAM-based DCiM. An exact IEEE 754- compliant multiplier is first implemented as a baseline, and a mantissa-segmentation-based approximate multiplier is then proposed to reduce hardware cost while preserving numerical fidelity. Post-layout results show up to 69% logic area reduction and 72% power savings over exact floating-point designs without delay overhead. Evaluations on image processing tasks and ResNet-18 inference further demonstrate negligible accuracy degradation. These results indicate that compiler-integrated approximate floating-point multiplication is a practical approach for enabling efficient and configurable floating-point support in SRAM-based DCiM systems, while also providing potential for future extensions toward more learning-oriented workloads. The Floating-Point Multiplier is available on GitHub.

14:30-14:50 | Paper ID: 237

Title: OptINC: Optical In-Network-Computing for Scalable Distributed Learning

Author(s): Sijie Fei, Grace Li Zhang, Bing Li, Ulf Schlichtmann

Presenter: Fei, Sijie, Technical University of Munich

Abstract: Distributed learning is widely used for training large models on large datasets by distributing parts of the model or dataset across multiple devices and aggregating the computed results for subsequent computations or parameter updates. Existing communication algorithms for distributed learning such as ring all-reduce result in heavy communication overhead between servers. Since communication in large-scale systems uses optical fibers, we propose an Optical In-Network-Computing (OptINC) architecture to offload the computation in servers onto the optical interconnects. To execute gradient averaging and quantization in the optical domain, we incorporate optical devices such as Mach-Zehnder-Interferometers (MZIs) into the interconnects. Such a de facto optical neural network (ONN) can effectively reduce the communication overhead in existing distributed training solutions. To reduce dataset complexity for training this neural network, a preprocessing algorithm implemented in the optical domain is also proposed. Hardware cost is lowered by approximating the weight matrices of the optical neural network with unitary and diagonal matrices, while the accuracy is maintained by a proposed hardware-aware training algorithm. The proposed solution was evaluated on real distributed learning tasks, including ResNet50 on CIFAR-100, and a LLaMA-based network on Wikipedia-1B. In both cases, the proposed framework can achieve comparable training accuracy to the ring all-reduce baseline, while eliminating communication overhead.



TECHNICAL SESSIONS

Technical Session 02: Logic and High-level Synthesis

🕒 13:30-14:50, May 8, 2026

📍 Orchid 4201

👤 Session Chair: **Chu, Zhufei**, Ningbo University

13:30-13:50 | Paper ID: 40

Title: A Novel Iterative Pairing Selection (IPS) Framework for FPGA Dual-Output LUT

Author(s): Zhaohan Wang, Yunfan Zuo, Du Yuhang, Ziji Zhao, Yaling Li, Chengao Zhang, Jiajie Xu, Lixin Ge

Presenter: **Wang, Zhaohan**, Southeast University

Abstract: Modern commercial field-programmable gate arrays (FPGAs) employ dual-output look-up tables (LUTs), in which a larger LUT (e.g., six-input) can be decomposed into two smaller LUTs (e.g., two five-input LUTs). This dual-output capability is crucial for reducing logic cell count while improving both area efficiency and circuit delay. In this work, we propose a novel Iterative Pairing Selection (IPS) method for efficient dual output LUT mapping. IPS extends ABC's multi-round iterative cutset optimization framework by introducing a Pairing Flow (PF) mechanism that explicitly captures interconnections among cutsets. By exploiting this connectivity information, PF identifies high-quality, pair-compatible cutsets suitable for dual-output implementation. A weighted graph-matching procedure is then applied to perform global pairing among the selected cutsets, forming dual-output LUTs. Experimental results demonstrate that IPS achieves a 36.78% area reduction compared with ABC's six-input single-output LUT mapping. Furthermore, relative to prior dual-output LUT mapping techniques, IPS provides additional improvements of 2.34% in area and 3.64% in logic depth, highlighting its ability to simultaneously optimize both area and delay.

13:50-14:10 | Paper ID: 189

Title: AXON: An Automated Netlist Optimization Framework for High-Speed Adders

Author(s): Tiantian Yang, Xuanle Ren, Qingdian Wan, Qi Meng

Presenter: **Yang, Tiantian**, ShanghaiTech University

Abstract: Adders are fundamental building blocks in modern digital systems, such as CPUs and accelerators, and their performance, power, and area (PPA) directly affect overall system efficiency. Contemporary adders commonly adopt parallel prefix architectures with well-known PPA trade-offs. However, for specific design objectives, these classical architectures often fail to achieve globally optimal PPA. Prior work often optimizes topology without netlist-/cell-level awareness, and general purpose synthesis heuristics are not adder-specific, leading to suboptimal PPA. To address these limitations, we propose an automated netlist optimization framework for adders, named AXON. The framework performs comprehensive design space exploration from the architectural to the netlist level, integrating prefix topology search with standard-cell-aware mapping strategies. It employs a hierarchical search methodology that rapidly converges to near-optimal solutions under given PPA constraints. Furthermore, we introduce a hybrid ultra-high-speed adder structure that combines the advantages of the parallel prefix and Ling architectures to effectively shorten the critical path. Experimental results based on the TSMC 28nm library show that AXON achieves up to 10.3%, 12.6%, and 32.1% improvement in delay, area-delay product, and energy-delay product, respectively, compared to the synthesis results using commercial synthesis tools.

14:10-14:30 | Paper ID: 211

Title: Timing-Oriented DFF and Splitter Refinement for RSFQ Circuits



Author(s): Qiangqiang Hu, Zhang Hu, Yinshui Xia, Lunyao Wang, Zhufei Chu

Presenter: Hu, Qiangqiang, Ningbo University

Abstract: Rapid single-flux quantum (RSFQ) circuits offer high speed and low energy consumption, making them promising candidates for post-CMOS high-performance computing. However, their strict gate-level pipelining and limited fanout pose significant challenges for logic synthesis. Path balancing requires the insertion of D flip-flops (DFFs) to equalize input arrival times, while high-fanout nodes necessitate deep splitter trees, both of which can severely inflate the circuit overhead and degrade timing performance. In this work, we propose a timing-oriented post-optimization framework for legalized RSFQ netlists. By exploiting level mobility during DFF insertion and applying selective node duplication to reshape splitter trees, our approach redistributes fanout across pipeline stages and structurally reduces splitter depth, thereby improving timing performance without violating global path balancing. Experimental results demonstrate that our method reduces maximum fanout and worst stage delay (WSD) by 70% and 30%, respectively, with only 11% DFF and 9% JJ overhead. Notably, even on high-quality pre-optimized netlists, it still achieves a 27% reduction in WSD, demonstrating its robustness.

14:30-14:50 | Paper ID: 257

Title: An MLIR-Based Compilation Flow for CGRAs: Optimized CDFG Generation, Control-Flow Handling, and Scalability

Author(s): Jiahang Lou, Jiayao Hu, Jianrong Zhang, Yikuan Chen, Zewei Zhong, Yuan Dai, Wenbo Yin, Lingli Wang, Yu He, Qing He

Presenter: Hu, Jiayao, Fudan University

Abstract: Mapping high-level applications with diverse memory access patterns and control structures onto Coarse-Grained Reconfigurable Architectures (CGRAs) remains a significant compilation challenge. This paper proposes a unified MLIR-based compilation flow that bridges multiple frontends (C/C++ and ONNX) to hardware-specific spatial mappings. We introduce an optimized Control-Data Flow Graph (CDFG) generation framework featuring dataflow pruning, control-flow handling via if-conversion and carry nodes, and memory access linearization. Evaluated on fundamental Polybench kernels, our approach significantly reduces mapped node counts and achieves an optimal Initiation Interval (II) compared to a state-of-the-art MLIR compiler. These results highlight the high structural quality of our generated CDFGs and demonstrate practical portability across different CGRA backends.



TECHNICAL SESSIONS

Technical Session 03: Fast Thermal & Reliability Modeling for 3D ICs

13:30-14:50, May 8, 2026

Orchid 4202

Session Chair: **Zhang, Yushen**, Technical University of Munich

13:30-13:50 | Invited Talk



Yu, Han

Huada Emyrean Tech. Corp.

Bio.: Yuhan works for Huada Emyrean Tech. Corp. from 2010 and he has abundant experience in AMS IC design together with EDA tools development. He participated in the architectural design of Emyrean's full-flow AMS EDA system, as well as the R&D and technical promotion of several digital EDA tools. Currently, he leads Emyrean's initiatives in industry-education integration and ecosystem promotion. Yuhan is also one of the deputy heads of EDA2's AMS and Advanced Packaging workgroup.

Title: Highlights to Full Flow 2.5/3D Chiplets Toolsuit and Development Trend

Abstract: The presentation will describe Emyrean's latest 2.5/3D chiplets full flow EDA architecture including physical implementation, Multi Physics simulation, and verification toolsuits. And also will discuss possible development trends of China domestic EDA in this field.

13:50-14:10 | Paper ID: 148

Title: IC-MoE: Interface-Constrained MoE for Fast Thermal Evaluation in 3D ICs

Author(s): Xiangqiao Meng, Wenkai Yang, Siyuan Miao, Jinlong Yan, Rumin Zhang, Chen Wu, Lei He

Presenter: **Meng, Xiangqiao**, The Hong Kong Polytechnic University

Abstract: Fast and accurate thermal evaluation is crucial for design space exploration in 3D integrated circuits (3D ICs), where multi-material stacks introduce complex and discontinuous heat conduction behavior. Conventional numerical solvers, though accurate, are computationally prohibitive for iterative optimization, and existing operator learning approaches struggle with the material discontinuities inherent in realistic 3D IC structures. In this study, we propose IC-MoE, an interface-constrained mixture-of-experts (MoE) framework for fast and reliable 3D IC thermal simulation. IC-MoE integrates an interface-aware flux loss that enforces physically correct heat transfer across discontinuous thermal conductivities, a bilateral interface sampling strategy that generates paired points along interface normals, and a domain-aware MoE architecture that models piecewise thermal behavior using domain-specific residual experts atop a shared global backbone. This design enables fast, label-free, and physically consistent thermal prediction for heterogeneous IC stacks. Extensive experiments demonstrate that IC-MoE achieves up to 63.9x faster inference than commercial solvers while reducing temperature error by 59.4% compared with state-of-the-art operator-learning baselines on multi-material 3D IC cases.

14:10-14:30 | Paper ID: 163

Title: OSF: An Operator-Splitting Framework for Fast Heterogeneous Thermal Simulation

Author(s): Hao Ai, Liang Chen, Jianhua Zhang, Wenxing Zhu

Presenter: **Ai, Hao**, Fuzhou University



Abstract: Fast and accurate thermal simulation is increasingly critical for thermal-aware “shift-left” design in modern EDA. Existing approaches face a fundamental trade-off: high-fidelity numerical solvers are too slow for iterative design workflows, while conventional analytical models rely on simplifying assumptions that break down under spatially heterogeneous material properties. To address this challenge, we propose OSF, an operator-splitting thermal simulation framework that extends fast constant-coefficient solvers to handle heterogeneity efficiently. OSF decomposes the thermal operator into a constant-coefficient base operator and a residual operator that captures spatial variation effects. Based on this decomposition, we develop OSF-PCG, a specialized preconditioned conjugate gradient solver for heterogeneous $\kappa(x, y)$. OSF-PCG leverages explicit finite-volume residual evaluation while preserving the efficiency of the underlying constant-coefficient solver. Experiments show that OSF achieves accuracy comparable to COMSOL (MAE less than 0.5 K), while achieving average speedups of 40× on heterogeneous problem with near $O(N \log N)$ scalability, where N is the total number of grid cells.

14:30-14:50 | Paper ID: 224

Title: Electromigration Stress Prediction for Power Delivery TSVs in 3D ICs Using Hierarchical Heterogeneous Graph Attention Networks

Author(s): Shuo Cui, Yunfan Zuo, Yan Li, Hao Yan, Longxing Shi

Presenter: Cui, Shuo, Southeast University

Abstract: The increasing complexity of 3D Integrated Circuits (3D ICs) has rendered electromigration (EM) a critical reliability challenge in the design of power delivery networks (PDNs). Specifically, heterogeneous materials in different chiplet regions and potential unbalanced arrangements of Through-Silicon Via (TSV) arrays increase finite element-based EM analysis time by multiple orders of magnitude. Although existing predictive methods efficiently analyze EM effects of individual TSVs, their performance cannot be directly extended to large-scale non-uniform TSV arrays. To address these limitations, we propose H2GAT, a method based on the Hierarchical Heterogeneous Graph Attention Network for the efficient prediction of EM-induced hydrostatic stress in complex TSV arrays. We use heterogeneous graphs to resolve modeling difficulties caused by material and structural differences in TSV-containing chiplets, and adopt a divide-and-conquer approach for large-scale non-uniform TSV arrays. Experimental results demonstrate that our method achieves up to 4190× speedup compared to commercial Finite Element Analysis (FEA) tools, while maintaining high fidelity with an average relative error of only 1.12%.



TECHNICAL SESSIONS

Technical Session 04: Hardware Design, Verification & ML-Enabled RTL

🕒 13:30-15:10, May 8, 2026

📍 Orchid 4203

👤 Session Chair: **Jia, Tianyu**, Peking University

13:30-13:50 | Paper ID: 77

Title: Pella: A Coq-based Hardware Description Language Infrastructure

Author(s): Liangliang Huang, Kangqi Zhou, Gang Chen

Presenter: **Huang, Liangliang**, Nanjing University of Aeronautics and Astronautics

Abstract: With the proliferation of open Instruction Set Architectures (ISAs) like RISC-V, achieving rapid yet reliable hardware implementation has become a paramount challenge. Traditional Hardware Description Languages (HDLs) such as Verilog suffer from weak typing, often masking bit-width mismatches and boundary errors. Meanwhile, modern agile HDLs like Chisel postpone critical safety checks until generation or simulation. To bridge this gap, we present Pella, a novel hardware description infrastructure embedded in the Coq proof assistant. The core innovation of Pella is the use of Dependent Types to encode hardware bit-width and indexing constraints directly into type signatures. Built upon a customized vector foundation, Pella enforces strict bit-width alignment across all arithmetic and structural operations at the type level. We validate Pella by constructing two RISC-V RV32I processor models—a single-cycle core and a Wildcat-style 3-stage pipeline with data forwarding—both extracted to executable OCaml code and verified through comprehensive simulation. A 92-test benchmark suite covering instruction compliance, algorithm execution, and cross-verification between the two models achieves a 100% pass rate.

13:50-14:10 | Paper ID: 172

Title: Prompt-Set: Bridging the Semantic and Optimization Gaps in LLM-Assisted HLS via Deep Structural Refactoring and System-Level Closure

Author(s): Hongpeng Wu, Junjie Fan, Mingjun Cai, Bin Huang

Presenter: **Wu, Hongpeng**, Jimei University

Abstract: High-Level Synthesis (HLS) has democratized FPGA design by raising abstraction levels, yet a substantial gap persists between algorithmic descriptions and high-performance hardware implementation, often necessitating deep architectural expertise. While Large Language Models (LLMs) have accelerated code generation, existing methodologies primarily focus on syntax correction or parameter-level Design Space Exploration (DSE), failing to address deep structural refactoring for complex computation patterns and lacking system-level closed-loop verification. To bridge these gaps, we propose Prompt-Set, a multi-stage, gateable SoC design paradigm. Prompt-Set decouples the design workflow into three distinct roles: Translator, Architect, and Integrator. By enforcing hard constraints and executable Gates—spanning C-Simulation, synthesis, and board-level execution—we effectively confine code generation uncertainty. Beyond simple pragma tuning, we introduce the Structural Optimization Library (S-Lib) to guide reusable micro-architectural refactoring, such as banking and systolic array mapping. Validated on benchmarks including LZ4, SHA256, Cholesky, and SR-UKF, Prompt-Set demonstrates significant latency reduction while ensuring interface consistency and successfully intercepting integration errors. Our prompt templates and source code will be publicly released.

Index Terms—LLM (Large Language Models), HLS (High-Level Synthesis), DSE (Design Space Exploration), Micro-Architectural Optimization, System-Level Verification, FPGA.



14:10-14:30 | Paper ID: 194

Title: Error-Constrained Cost-Driven Synthesis of General-Purpose Special Function Units

Author(s): Zikai Zhao, Ke Hu, Yuxuan Qiao, Jun Liu, Lei Xu, Wenhao Cai, Xiankui Xiong, Fan Yang

Presenter: **Zhao, Zikai**, Fudan University

Abstract: Different architectures have different SFU accuracy requirements, with many variables---for example, the target function mix (activation vs. general special functions), practical input domain, internal numeric format, and the chosen acceptance metric (absolute/relative/hybrid error)---shaping the best accuracy--hardware-cost trade-off. However, many Special Function Unit (SFU) designs still follow a sequential methodology that fixes hardware configurations before systematically evaluating approximation quality. To address this limitation, we propose a system-level, error-constrained SFU synthesis framework that treats target accuracy as an explicit design constraint and automatically derives hardware parameters. The proposed HW/SW co-design flow jointly optimizes the internal floating-point format, non-uniform segmentation, and polynomial coefficients via quadratic Remez approximation and custom floating-point validation, and then generates synthesizable RTL using a parameterized backend. The resulting SFU operator integrates built-in pre- and post-processing so that a unified approximation core can support a broad class of special functions. Compared with the state-of-the-art Flex-SFU baseline, the synthesized core achieves up to $1.95\times$ MSE reduction over $[-8, 8]$ while reducing LUT-related area by 11.4%--42.0% in 28 nm technology, demonstrating that error-constrained synthesis can outperform preset-based SFU design in both accuracy and hardware cost.

14:30-14:50 | Paper ID: 164

Title: Lightweight RISC-V ISA Extensions for Efficient ML-KEM and its Ascon Variant

Author(s): Yuxiang Li, Fei Lyu, Bei Wang, Chenghua Wang, Yijun Cui, Weiqiang Liu

Presenter: **Li, Yuxiang**, Nanjing University of Aeronautics and Astronautics

Abstract: ML-KEM and its lightweight variant are post-quantum cryptographic algorithms resistant to quantum attacks and have attracted extensive attention and research in recent years. This work explores the deep optimization space for hardware-software co-acceleration of ML-KEM and its Ascon variant on 32-bit RISC-V devices. We propose a series of custom instructions supported by a resource-efficient microarchitecture to accelerate polynomial operations, together with dedicated Ascon instructions that exploit register-level reuse mechanisms to enable single-round permutations. The entire implementation is based on the 32-bit RISC-V Hummingbird E203 processor and a Xilinx XC7A35T FPGA. Experimental results show that, compared to the C implementation, the proposed instruction set reduces the cycle counts of the NTT, Basemul, and INTT stages by 87.72%, 77.6%, and 91.12%, respectively. The hardware overhead is only 208 LUTs and 2 FFs, without introducing additional multipliers, BRAM, or DSP resources. Furthermore, in the complete implementation of the Ascon variant, our modified SoC platform achieves an overall cycle reduction of more than 86% at the cost of only 740 additional LUTs and 2 FFs.

14:50-15:10 | Paper ID: 150

Title: CoPU-RL: A Hardware-Friendly Training-Inference Framework For Neuro-Symbolic Reinforcement Learning

Author(s): Fengyuan Guo, Hongxiao Zhao, Jiayi Li, Anjunyi Fan, Yihan Fu, Wenshuo Yue, Bonan Yan

Presenter: **Zhao, Hongxiao**, Peking University

Abstract: Deep reinforcement learning (DRL) is highly effective for complex decision-making tasks; however, its trial-and-error paradigm often results in low sample efficiency and limited generalization to environments with minor modifications. To address these limitations, this paper proposes a hardware-software codesign framework that integrates data-driven learning with symbolic reasoning. Central to this framework is a neuro-symbolic algorithm wherein DRL conducts policy learning, while a cognitive architecture provides



symbolic feedback to constrain and refine the policy space. Supporting this algorithm, a dedicated hardware architecture is designed for efficient execution of the hybrid model. Experimental results indicate that the proposed algorithm significantly outperforms conventional DRL baselines in learning speed and adaptability. Furthermore, the specialized hardware achieves latency reductions of 15.18%~87.63% and energy consumption decreases of 60.63%~99.38% compared to mainstream CPU+GPU or CPU+NPU solutions, highlighting the promise for building efficient and reliable reinforcement learning systems. The proposed framework is open-sourced at <https://github.com/a-FCBer/CoPU-RL>.



TECHNICAL SESSIONS

Technical Session 05: Topology/Layout, ADC & LLM-enabled Synthesis



16:00-18:00, May 8, 2026



Orchid 4302



Session Chair: **He, Lei**, Eastern Institute of Technology, Ningbo

16:00-16:20 | Paper ID: 115

Title: A Novel CNN-Based Design Framework for Incorporating Electrical Chip-Package Interaction with Application to Bandgap Reference Design

Author(s): Ziji Zhao, Chengjie Liu, Zhonglei Yang, Pengxiang Li, Yaling Li, Zhaohan Wang, Chengao Zhang, Lixin Ge

Presenter: **Zhao, Ziji**, Southeast University

Abstract: Package-induced stress and the resulting electrical Chip-Package Interaction (e-CPI) present a new challenge for modern analog circuit design, driven by the trends of increasingly thinner silicon dies and more complex packaging. Current EDA tools and design flows, however, lack the capability to account for e-CPI. To bridge this gap, we present a novel Convolutional Neural Network (CNN)-based design framework that directly incorporates e-CPI effects into the analog design flow. The CNN is trained on stress data from Finite Element Analysis (FEA), parameterized precisely by spatial coordinates and operating temperature. This enables the instantaneous prediction of stress distributions for arbitrary design configurations. These distributions are then translated into carrier mobility deltas via a piezoresistive model and back-annotated to the circuit netlist, enabling direct simulation and evaluation of e-CPI impact on circuit performance. The CNN-based design framework is validated through a case study on a CMOS Bandgap Reference (BGR) circuit. Simulations identify the open-loop output stage as the most vulnerable block, showing a worst-case degradation of 36% in its temperature coefficient under bump-induced stress. Finally, we discuss the design insights and optimization strategies derived from this analysis to mitigate package stress impact.

16:20-16:40 | Paper ID: 157

Title: Automated Synthesis of Abutable Macros for Analog Layout

Author(s): Li Huang, Xueying Huang, Xinpeng Wang, Fuya Yang, Junjie Yang, Lin Yang, Chuanxu Liu, Rui Ye, Zhanfei Chen, Ting-Jung Lin, Lei He

Presenter: **Huang, Li**, Shanghai Jiao Tong University

Abstract: Automating physical design for analog circuits remains a formidable challenge, as layout quality still depends heavily on expert knowledge and manual tuning. Inspired by the digital standard-cell paradigm, abutable analog cells (Acells) offer a promising path. However, existing Acell-based flows often fail to rigorously enforce device-matching constraints, resulting in parasitic imbalance and performance loss. This paper presents AmacroSyn, an automated framework that synthesizes abutable analog macros (Amacros) from primitive Accells. The framework integrates symmetry-aware place-and-route (PnR) and constraint-preserving compaction to synthesize Amacros that are design-rule-checking (DRC)- and layout-versus-schematic (LVS)-clean, and fully compatible with standard digital PnR tools. In a 28 nm technology node, AmacroSyn reduces total transistor area by 15.3% on average. For symmetry-critical specifications, Amacro-based layouts exhibit less than 2.5% performance deviation compared to manual Pcell-based layouts, while providing up to a 21.6X improvement in offset voltage over Acell-only baselines.



16:40-17:00 | Paper ID: 177

Title: SpecMind: Automated Analog-to-Digital Converter (ADC) Diagnosis with Agentic Reasoning

Author(s): Zhishuai Zhang, Kezhuo Liu, Aodong Zhang, Shilong Liu, Yi Zhong, Nan Sun, Lu Jie

Presenter: Liu, Kezhuo, Tsinghua University

Abstract: Debugging performance degradation in analog and mixed-signal (AMS) integrated circuits remains manual because failures must be interpreted across waveforms, spectra, and code statistics. Vision language models (VLM) can read spectra, but they lack domain knowledge and tool use and therefore produce unreliable answers. This paper presents SpecMind, a proof-of-concept framework for agentic analog-to-digital converter (ADC) failure diagnosis that preprocesses raw data, extracts quantitative signatures, retrieves diagnostic fragments and patterns from a structured expert-curated knowledge base, and generates interpretable reports. We construct a controlled benchmark of 800 labeled ADC diagnosis cases designed to isolate diagnostic reasoning; within this benchmark, SpecMind achieves more than 88% diagnostic accuracy and outperforms baseline VLMs while providing physically consistent reasoning.

17:00-17:20 | Paper ID: 214

Title: A Hardware-Efficient Adaptive Two-Stage Digital Calibration Method for SAR-Assisted Hybrid ADCs

Author(s): Jiaju Lu, Siqi Zhang, Xinzhe Xie, Wang Ling Goh, Jinhai Hu, Yuan Gao

Presenter: Zhang, Siqi, Nanyang Technological University

Abstract: This paper presents a general two-stage digital calibration method for high-resolution hybrid ADC architectures. Motivated by the increasing complexity of static mismatch, dynamic nonlinearities, and noises in advanced data converters, an error compensation scheme is established to separately model linear and dynamic distortions. The proposed method connects a bit-weight optimized linear projection stage with a normalized LMS-based Adaptive Line Enhancer to compensate for static bit-weight and interstage gain errors and dynamic nonlinearities and noises, respectively. The dedicated stages and the alternating optimization strategy guarantee a small number of parameters, stable convergence, and effective calibration outcome, which makes the calibration method highly compatible with design automation flows. Validated on a 14-bit pipeline-assisted noise-shaping SAR ADC prototype, the proposed calibration method achieves a peak Signal-to-Noise-and-Distortion Ratio (SNDR) improvement of 21.4dB in post-layout simulations with only 62 parameters and 150 MACs required.

17:20-17:40 | Paper ID: 199

Title: Causally-Grounded Generation: LLM-Based Analog Circuit Synthesis via Causal Hierarchy

Author(s): Wenjin Mai, Xuan Zeng, Zhaori Bi

Presenter: Mai, Wenjin, Fudan University

Abstract: Large Language Models (LLMs) show growing promise in electronic design automation, yet their application to analog circuit synthesis faces a fundamental gap: LLMs operate as associational models that capture statistical regularities from training data, while analog circuit behavior is governed by device physics and causal mechanisms that pattern matching alone cannot reliably predict. We propose Causally-Grounded Generation (CGG), a multi-agent pipeline organized along Pearl's causal hierarchy, a framework that stratifies reasoning from statistical association through intervention to counterfactual inference. To validate the framework, we apply CGG to specification-driven synthesis of analog circuits from academic publications, where the source papers provide ground-truth specifications for objective assessment. Without simulator grounding, pure LLM generation produces no functional circuit in any run. With the full CGG pipeline, across four benchmark circuits each tested with multiple independent runs, three of four circuits achieve 80% synthesis success rate. The results demonstrate that organizing the pipeline along the causal hierarchy effectively bridges the gap between statistical generation and physics-grounded circuit synthesis.



17:40-18:00 | Paper ID: 204

Title: TopoSynth: Explainable and Sample-Efficient Topology Synthesis of Analog/Mixed-Signal Circuits via an LLM-powered Multi-Agent System

Author(s): Jiaqi Wang, Georges G.E. Gielen

Presenter: Wang, Jiaqi, KU Leuven

Abstract: Automated topology synthesis of analog/mixed-signal circuits remains an open challenge due to the vast combinatorial design space, the high simulation costs, and the lack of interpretability in existing approaches. This paper presents TopoSynth, an explainable and sample-efficient framework based on a general-purpose large language model (LLM)-powered multi-agent system that emulates an expert design team. It integrates causal critique and refinement to provide interpretable, human-readable feedback that explains why designs succeed or fail, and employs a resource manager agent to intelligently allocate the computational budget, dynamically balancing LLM-driven heuristics with reinforcement learning optimization. Demonstrated for the case of $\Delta\Sigma$ ADCs, the system explores over 8,000 candidate topologies to determine the best suited topology for the given specifications, while jointly optimizing its high-level design parameters. Experimental results show that, for demanding specifications, the proposed method discovers successful designs up to 3.6x faster than a state-of-the-art bi-level Bayesian optimization baseline, while achieving a large reduction in power consumption. These results demonstrate the potential of LLM-powered multi-agent systems to scale joint topology and parameter synthesis beyond existing paradigms while providing critical gains in efficiency, solution quality, and especially design transparency and explainability.



TECHNICAL SESSIONS

Technical Session 06: Formal Verification & Equivalence Checking



16:00-18:00, May 8, 2026



Orchid 4304



Session Chair: **Liu, Tianji**, Institute of Computing Technology, Chinese Academy of Sciences

16:00-16:20 | Paper ID: 100

Title: CondEC: Equivalence Checking under Conditions

Author(s): Wenbin Che, Changyuan Yu, Hongce Zhang

Presenter: **Che, Wenbin**, The Hong Kong University of Science and Technology (Guangzhou)

Abstract: Combinational equivalence checking (CEC) is a fundamental technique in hardware verification, widely used to validate functional equivalence between design revisions or optimizations. However, conventional CEC enforces equivalence over the entire input space and cannot effectively handle scenarios where equivalence is required only for a subset of inputs under specific conditions. This paper proposes CondEC, a constraint-aware framework that enables conditional equivalence checking by explicitly integrating additional constraints into the CEC workflow. CondEC incorporates constraint-guided simulation and constraint-aware checking of equivalent pairs to reason about equivalence under a restricted functional domain. To the best of our knowledge, this is the first work that systematically supports conditional equivalence checking within a gate-level CEC framework. Experimental results on representative benchmarks show that CondEC achieves effective speedups compared to the primary CEC tool and state-of-the-art SAT solvers.

16:20-16:40 | Paper ID: 110

Title: Symbolic Method for Static CDC Checking

Author(s): Qin He, Ziqi Liu, Yingmeng Li

Presenter: **He, Qin**, Ennocad Electronics Technology Company, Ltd.

Abstract: With the increasing complexity and scale of industrial circuits, Clock Domain Crossing (CDC) violation detection in synthesized circuits has become increasingly challenging. The primary difficulty is to detect CDC issues and synchronization/unsynchronization (sync/unsync) patterns within a multidimensional problem domain involving sources, qualifiers, and destinations. We present a symbolic method for precise static CDC checking. First, a CDC State (CS) Symbolic Algorithm performs calculations based on BDD to mark CS-Symbols at each node. A subsequent consolidation pass refines it to reflect multi-qualifier effects. Second, a Pattern Specific (PS) Symbolic Algorithm leverages CS-Symbols to perform calculations and propagate PS-Symbols across the domain for each source. The symbols serve as the basis for determining the sync/unsync. Compared with an industrial CDC tool, our approach achieves 100% accuracy in both CDC sync/unsync detection for a real-life design.

16:40-17:00 | Paper ID: 178

Title: A Phase-Coordinated Optimization Framework for Accelerating SAT-Sweeping

Author(s): Haihang Wang, Yue Wu, Xiaoyan Yang

Presenter: **Wang, Haihang**, Hangzhou Dianzi University

Abstract: Modern combinational equivalence checking (CEC) flows heavily rely on SAT sweeping, which interleaves logic simulation with SAT solving to merge functionally equivalent nodes. However, conventional SAT sweeping suffers from two major bottlenecks: random simulation quickly loses power in large designs,



generating many spurious candidates, and counterexamples (CE) are exploited locally, leading to excessive SAT calls. This paper proposes a phase-coordinated optimization framework that jointly improves the simulation and refinement phases. In simulation, we introduce a critical-PI-based strategy that ranks inputs using fanout and transitive-fanout (TFO) metrics, applies exhaustive stimuli to high-impact inputs, reducing pseudo-equivalences. In refinement, we guide SAT solving towards top-level nodes and expand each counterexample through combinational perturbations along high-impact inputs to maximize distinguishing power. We implement the proposed techniques in ABC and evaluate them on large ISCAS'85/89, IWLS'05, and EPFL benchmarks. Compared with the publicly available ABC SAT-sweeping engine "abc-sweep", our framework preserves the same final AIG sizes while reducing the geometric mean number of SAT calls by about 11.5% and the SAT-sweeping runtime by about 12.3%.

17:00-17:20 | Paper ID: 228

Title: Efficient SAT Sweeping via Direct Circuit-Based SAT Solving

Author(s): Junjie Xu, Zhang Hu, Yinshui Xia, Lunyao Wang, Zhufei Chu

Presenter: Xu, Junjie, Ningbo University

Abstract: SAT sweeping is a widely used technique for combinational equivalence checking that merges equivalent nodes in And-Inverter Graphs (AIGs) through a large number of SAT queries. Most existing flows rely on CNF-based SAT solving, which introduces conversion overhead and discards circuit structural information. This paper presents an efficient SAT sweeping framework built upon a native circuit-level SAT solver operating directly on AIGs. By exploiting circuit topology for propagation and decision making, the proposed approach avoids CNF translation and is particularly effective for the small, repetitive queries that dominate SAT sweeping workloads. To support iterative equivalence checking, we further integrate cone-based solving and a selective reset mechanism that preserves proven UNSAT outputs across queries. Experimental results show that, on standalone SAT instances constructed from MCNC-derived miter circuits, our solver achieves a 2.85× geometric mean speedup over MiniSat and a 2.55× speedup over Kissat. When integrated into SAT sweeping and evaluated on 13 benchmarks derived from the EPFL arithmetic suite and ISCAS'89 circuits, the framework achieves 2.68× and 3.12× geometric mean speedups in total runtime and SAT solving time, respectively, over ABC's abc-sweep. These results demonstrate that circuit-level reasoning provides consistent and robust performance improvements for SAT sweeping.

17:20-17:40 | Paper ID: 242

Title: AutoPDR: Circuit-Aware Solver Configuration Prediction for Hardware Model Checking

Author(s): Guangyu Hu, Chen Chen, Xiaofeng Zhou, Jiaxi Zhang, Wei Zhang, Hongce Zhang

Presenter: Zhang, Hongce, The Hong Kong University of Science and Technology (Guangzhou)

Abstract: Property Directed Reachability (PDR) is a powerful algorithm for formal verification of hardware and software systems, but its performance is highly sensitive to parameter configurations. Manual parameter tuning is time-consuming and requires domain expertise, while traditional automated parameter tuning frameworks are not well-suited for time-sensitive verification tasks like PDR. This paper presents a circuit-aware solver configuration framework that employs graph learning for intelligent heuristic selection in PDR-based verification. Our approach combines graph representations with static circuit features to predict optimal PDR solving configurations for specific circuits. We incorporate expert prior knowledge through constraint-based parameter filtering to eliminate invalid and inefficient configurations and reduce 78% search space. Our feature extraction pipeline captures structural, functional, and connectivity characteristics of circuit topology and component patterns. Experimental evaluation on a comprehensive benchmark suite demonstrates significant performance improvements compared to default configurations and commonly-used settings. The system successfully identifies circuit-specific parameter patterns and automatically selects the



most suitable solving strategies based on circuit characteristics, making it a practical tool for automated formal verification workflows.

17:40-18:00 | Paper ID: 193

Title: Orion: Automated Verification of Randomized Dynamic Schedules for Dataflow HLS

Author(s): Lianghui Wang, Zijian Jiang, Ziqing Zhang, Weijie Weng, Yungang Bao, Kan Shi

Presenter: **Jiang, Zijian**, Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences

Abstract: Dataflow high-level synthesis (HLS) automatically translates untimed C/C++ code into dynamically scheduled dataflow circuits. While dynamic scheduling substantially boosts performance for dataflow HLS designs, it also introduces a vast schedule space that must be verified, posing challenges to comprehensive functional verification. Existing tools and methodologies exercise only a narrow subset of dynamic schedules and lack quantitative metrics to characterize verification coverage over the schedule space, raising concerns about design robustness. To address these challenges, we present Orion: an agile verification framework for dataflow HLS designs that extensively explores the schedule space by automatically instrumenting enhanced stalling units and coverage logic. Orion adopts a novel stalling unit design that decouples stalling length from stalling probability, enabling broader exploration of the schedule space and thereby increasing the likelihood of exposing schedule-dependent bugs. Orion also defines a schedule coverage metric based on temporal signatures, offering a measurable basis for assessing verification completeness. Furthermore, Orion supports FPGA deployment for accelerated simulation. Experimental results on realistic benchmarks demonstrate that Orion achieves an average 16.8% improvement in schedule coverage while providing a simulation speedup of 2859–5462 $\times$.



TECHNICAL SESSIONS

Technical Session 07: Timing Analysis & Optimization



16:00-18:00, May 8, 2026



Orchid 4201



Session Chair: **Tan, Chee Wei**, Nanyang Technological University

16:00-16:20 | Paper ID: 22

Title: Sparse-GCS: A High-Performance GPU Timer for Composite Current Source Model-Based STA

Author(s): Haoran Li, Yukun Wang, Hu Zhou, Guangyu Sun, Guojie Luo

Presenter: **Li, Haoran**, Peking University

Abstract: Composite Current Source (CCS) timing model plays an important role in modern static timing analysis (STA) because it precisely captures the timing behavior of a design at advanced nodes. However, CCS is extremely time-consuming due to its accurate but complicated timing models. To overcome this challenge, prior work has attempted to accelerate CCS models using GPUs with moderate success. However, significant computational resources are wasted during the matrix preprocessing stage. The existing algorithms assume dense matrices, whereas, the admittance matrices exhibit sparsity. This allows for optimization in storage based on CSR format and computation based on sparse LU factorization, thereby improving both space and time efficiency of the algorithms. Experimental results show that our Sparse-GCS Timer achieved up to 2.57× faster runtime than baseline in the preparation stage, and the space complexity is reduced to $O(n + |E|)$ in the simulation stage.

16:20-16:40 | Paper ID: 52

Title: GraphDTA: Fast Dynamic Timing Analysis for Circuits with Graph Representation Learning

Author(s): Guangxi Fan, Tianliang Ma, Xuguang Sun, Xun Wang, Cheng Wang, Zhiping Yu, Xiaolei Zhu, Leilai Shao

Presenter: **Fan, Guangxi**, Shanghai Jiao Tong University

Abstract: The conservative nature of static timing analysis (STA) in delay estimation consistently results in overdesign, leading to suboptimal power efficiency and increased area overhead, particularly in compute-intensive AI applications. To address these limitations, this paper introduces GraphDTA, a learning-based dynamic timing analysis (DTA) framework tailored for functional units in AI accelerators. GraphDTA combines graph-based representation learning and downstream model to predict workload-induced dynamic timing slack with high accuracy. We evaluate our approach on 10 benchmarks, including multiplier units (MULs), multiply-accumulate units (MACs) and matrix multiplier units (MMUs), across 45 nm and 7 nm technologies. Our framework outperforms existing machine-learning methods by 37.52% for 45 nm and 42.60% for 7 nm, and achieves roughly a 50X speedup compared to gate-level simulation.

16:40-17:00 | Paper ID: 59

Title: Pre-routing Timing Prediction with Global Netlist Topology Awareness

Author(s): Mingle Wu, Zhongdong Qi, Qi Peng

Presenter: **Wu, Mingle**, Xidian University

Abstract: Accurate and efficient pre-routing timing prediction is crucial for enhancing chip design efficiency. Machine learning based methods have achieved great progress in estimating the timing delay after layout, avoiding the significant time overhead caused by routing and timing optimization. However, existing methods lack use of global netlist topological information that is important for timing metrics. To address this



limitation, we propose a dual-branch model to collectively capture global netlist topology and local circuit and layout features. Specifically, we design a linear-complexity graph transformer with Mamba inspiration, which can efficiently extract information from large scale designs. In addition, a customized graph neural network is leveraged to propagate circuit information as well as layout features extracted by U-Net. The rich information gathered by these two branches is used to predict endpoint arrival time. Experiments on open-source circuits with advanced 7 nm technology node validate the effectiveness of our proposed method.

17:00-17:20 | Paper ID: 156

Title: ElmoreCeff: A GPU-Friendly Elmore-Like Delay Calculator with a Closed-Form Effective Capacitance Model

Author(s): Haichuan Liu, Zizheng Guo, Runsheng Wang, Yibo Lin

Presenter: Liu, Haichuan, Peking University

Abstract: Accurate and efficient delay calculation is critical for timing-driven optimization, yet existing methods require a trade-off between fast but inaccurate Elmore timers and accurate but slow model order reduction (MOR) techniques. To bridge this gap, we introduce ElmoreCeff, a GPU-friendly Elmore-like delay calculator based on a closed-form effective capacitance model. ElmoreCeff can achieve speedups of up to 40.95 $\times$ and 5.95 $\times$ over CPU- and GPU-based MOR timers, respectively, while achieving comparable accuracy. This advantage enhances final quality of results (QoR) of timing optimization, improving worst negative slack (WNS) and total negative slack (TNS) by up to 8.4% and 5.4% compared to optimizations guided by the standard Elmore timer.

17:20-17:40 | Paper ID: 233

Title: Fast Timing Analysis Framework Considering Manufacturing and In-field Operation for Silicon Lifecycle Management

Author(s): Yewei Zhang, Huizhen Qiu, Feiyang Shu, Muyan Jin, Pengpeng Ren, Zhigang Ji

Presenter: Zhang, Yewei, Shanghai Jiao Tong University

Abstract: Manufacturing-induced process variations and workload-dependent aging jointly affect circuit timing throughout the silicon lifecycle. For silicon lifecycle management (SLM), timing analysis must move beyond conventional sign-off and support rapid, silicon-aware assessment under chip-specific process signatures and time-varying aging states. This paper presents a fast timing analysis framework for SLM by integrating an addressable test circuit (ATC) with a hybrid graph neural network (GNN). The ATC enables high-throughput, fine-grained measurement of on-chip process characteristics, while the hybrid GNN performs rapid aging-aware delay inference conditioned on the measured process signatures. The inferred standard-cell delay values are incorporated into gate-level timing propagation to estimate the critical-path delay under both manufacturing variations and in-field aging conditions. Unlike conventional process-corner- or margin-based analysis, the proposed framework leverages measured process information, supports rapid adaptation to new cell types and process signatures through few-shot fine-tuning, and provides timing-oriented decision support for operating-condition management in SLM. Experimental results on a 7 nm test vehicle show that the framework achieves over 90% cell-delay prediction accuracy for most standard cells with only 300 fine-tuning samples, reducing retraining time by more than 98% compared with direct training.

17:40-18:00 | Paper ID: 270

Title: Gradient-Guided Hierarchical Lagrangian Optimization for Timing ECO

Author(s): Hao Yan, Jiajie Xu, Yanfang Zhang

Presenter: Zhang, Yanfang, Southeast University

Abstract: Timing Engineering Change Order (ECO) plays a pivotal role in modern physical design for timing closure. As circuit complexity continues to increase in advanced technology nodes, the combinatorial



explosion of candidate modifications increases the complexity of timing optimization. Our study reveals the uneven distribution of timing optimization potential, indicating that effective improvements are concentrated on a small subset of critical cells. Motivated by this observation, we propose a gradient-guided hierarchical Lagrangian optimization framework. Timing gradients propagated from violated endpoints identify critical cells, which are subsequently optimized in a hierarchical batch-wise manner, with the Lagrangian formulation providing reliable evaluation of candidate replacements. Experiments on four open-source designs demonstrate that our method reduces the number of modified cells by 51.7% while achieving 11.0% better total negative slack (TNS), 19.6% lower power, and 15% lower cell area than [9].



TECHNICAL SESSIONS

Technical Session 08: Lithography & Mask Optimization



16:00-18:00, May 8, 2026



Orchid 4202



Session Chair: **Zhu, Binwu**, Southeast University

16:00-16:20 | Paper ID: 13

Title: U-WNO-ILT: U-Net Enhanced Wavelet Neural Operator with Spectral-Spatial Decoupling for Inverse Lithography

Author(s): Xinyu Shang, Tianyi Xu, Hongwei Yao, Bingyi Ye, Yang Shen, Xiaojin Li, Yabin Sun, Yuhang Zhang

Presenter: **Shang, Xinyu**, East China Normal University

Abstract: Inverse Lithography Technology (ILT) is vital for advanced chips but has high computational costs. Current deep learning methods face challenges in balancing global diffraction, edge fidelity, and speed. To address these challenges, we propose U-WNO-ILT (U-Net Enhanced Wavelet Neural Operator for ILT), a physics-aligned spectral-spatial decoupling framework for ILT. While inheriting the mathematical foundation of wavelet operators, we reshape the architecture and high-frequency processing strategies. Global diffraction is modeled in the low-frequency band (complemented by spatial processing), while local edges are refined in the high-frequency bands. The architecture incorporates Large Kernel Attention for long-range optical proximity effects and an edge-aware training loss aligned with lithography goals. Experiments on LithoBench show that U-WNO-ILT outperforms state-of-the-art baselines by 17% in L2 and over 50% in EPE, with an 8--9 $\times$ speedup over BBDM-ILT.

16:20-16:40 | Paper ID: 20

Title: TC-FNO: Fourier Neural Operator with Learnable Time-Step Embedding for 3D Lithography Exposure Simulation

Author(s): Yuqian Zang, Yiqing Wang, Silin Chen, Jiale Li, Zhili Xu, Yufeng Zhang, Hao Zhang, Chia-Yen Li, Ningmu Zou

Presenter: **Zang, Yuqian**, Nanjing University

Abstract: In chemically amplified photoresist, optical exposure transforms the spatial light intensity distribution into a three-dimensional photoacid distribution, which in turn drives the post-exposure bake (PEB) and development processes. Rigorous exposure simulation is typically formulated as solving a system of coupled partial differential equations on a dense three-dimensional mesh. However, solving these equations is computationally expensive and can become a bottleneck in photoresist simulation and optimization. Inspired by the physical structure of the exposure process, we propose a time-conditioned Fourier neural operator as a neural PDE surrogate for exposure simulation. The key innovation lies in explicitly encoding the time-step size Δt to constrain the learned evolution operator, since discrete exposure updates inherently depend on Δt . This model learns the mapping from the current three-dimensional photoacid field and light intensity field to the photoacid field at the next time step. Compared to traditional numerical solvers, it can achieve accurate prediction of three-dimensional photoacid evolution with significantly reduced runtime. This method provides efficient exposure modeling support for photoresist simulation, illumination/mask optimization, and layout co-optimization.

16:40-17:00 | Paper ID: 18

Title: Adaptive Frequency-Domain Optimization for Lithography Mask Synthesis with Topological Constraints



Author(s): Jiale Liu, Wenjing He, Fengyun Long, Ruifan Gao, Luying Liang

Presenter: **Shen, Yijiang**, Guangdong University of Technology

Abstract: Inverse Lithography Technology (ILT) is critical for advancing semiconductor scaling, yet conventional pixel-based optimization often suffers from convergence stagnation and topological violations, such as disconnected SRAFs and pinhole defects. To address these challenges, this paper proposes a robust two-stage optimization framework that bridges the frequency and spatial domains. In the first stage, we introduce a Symmetric Frequency-Domain Optimization (SFDO) strategy. By optimizing spectral coefficients with a progressive frequency release mechanism, this stage circumvents the "curse of dimensionality" and generates a high-quality gray-scale seed map with globally consistent topology. In the second stage, we implement a Sequential Region-Activated Refinement strategy. Instead of treating the pixel domain uniformly, the gray-scale seed is stratified into distinct regions based on intensity and sequentially integrated into the optimization process. This progressive inclusion protocol explicitly prioritizes the retention and growth of weak sub-resolution signals, significantly enhancing the extraction capability for SRAFs. Furthermore, to ensure topological integrity, we incorporate a Differentiable Morphological Guidance (DMG) mechanism. Complementing the sequential optimization, DMG employs a soft loss function derived from morphological closing targets to seamlessly heal void defects and enforce pattern continuity without breaking the gradient flow. Experimental results demonstrate that this coarse-to-fine approach effectively synthesizes robust, SRAF-rich masks with defect-free main features, achieving superior fidelity and process window stability compared to traditional spatial-only methods.

17:00-17:20 | Paper ID: 57

Title: Efficient Mask Optimization via Brownian Bridge Diffusion Model

Author(s): Yihan Li, Wanlin Zhang, Liying Yang, Hu Ding

Presenter: **Zhang, Wanlin**, University of Science and Technology of China

Abstract: In the modern VLSI manufacturing process, mask optimization is a critical step to compensate for optical distortions. However, traditional mask optimization methods suffer from high computational costs and extremely long turn-around times. This paper introduces a novel method that uses a Brownian Bridge diffusion model for mask generation to improve efficiency and quality. We treat mask optimization as an image-to-image translation problem, learning the mapping between pre-lithography and post-lithography image data. In this case, we can regard the layout as a post-lithography image to generate the corresponding pre-lithography mask image. This approach enables the generation of high-quality masks just through a single inference pass. Our approach achieves a 40% reduction in time cost over leading ILT frameworks and superior performance on the crucial L2 metric.

17:20-17:40 | Paper ID: 215

Title: MODiff: Layout-Guided Mask Optimization via Diffusion Model

Author(s): Jiale Li, Silin Chen, Yiqing Wang, Kangjian Di, Li Du, Ningmu Zou

Presenter: **Li, Jiale**, Nanjing University

Abstract: Inverse lithography technology (ILT) is one of the resolution enhancement techniques (RETs) that is very promising in modern chip manufacturing. However, traditional ILT methods need to consume a lot of computing resources, and the existing models based on deep learning are unstable in training due to the lack of a smooth prediction process. As a result, the generated mask is relatively discontinuous, which leads to a loss of mask accuracy and printability. In this paper, we proposed the MODiff based on the diffusion model. The model can make strong control over both the global layout map and each detailed object in the process of mask generation to obtain a high-quality mask. Firstly, we use a "noise to map" generative paradigm to make predictions by removing noise from a random Gaussian distribution step by step, guided by the target



layout. Secondly, we propose a forward lithography simulation-based module to improve the mask accuracy. Finally, We introduce a cosine scheduling strategy specifically designed for mask optimization. Experimental results demonstrate that our method provides a clear and interpretable visualization of the mask generation process, and achieves superior performance compared with other state-of-the-art approaches.

17:40-18:00 | Paper ID: 254

Title: Differentiable Optical Proximity Correction for Metasurfaces Using a Pitch-Conditioned Neural Operator Surrogate

Author(s): Zaifeng Yang, Ayan Nussupbekov, Ray Jia Hong Ng, Viet Phuong Bui, Wei Bin Ewe, Hong Son Chu, Ching Eng Png

Presenter: Yang, Zaifeng, A*STAR

Abstract: Metasurface fabrication relies on dense periodic arrays of sub-wavelength meta-atoms whose printed photoresist profiles can deviate substantially from intended mask geometries due to diffraction-driven optical proximity effects. These effects make repeated high-fidelity lithography simulation a major computational bottleneck for design-space exploration and mask correction. To enable fast, accurate, and differentiable prediction of printed resist patterns on periodic domains, we develop a pitch-conditioned Fourier Neural Operator (FNO) surrogate that learns the mapping from binary mask patterns to continuous-valued resist profiles for periodic circular (negative-tone) meta-atoms over a wide pitch range (200–650 nm). The surrogate is trained on paired mask/resist data generated by Prolith at 1 nm/pixel resolution across pitch-dependent unit-cell sizes and is augmented with periodic positional encodings and an explicit normalized pitch-conditioning input to accommodate variable periodic domains. On the validation set, the surrogate achieves low prediction error ($L1 = 1.651 \times 10^{-3}$ and $MSE = 3.57 \times 10^{-4}$) and remains fully differentiable, enabling gradient-based OPC through backpropagation. Across the OPC benchmarks, the mean IoU between the printed resist and the target increases from 0.548 to 0.988, indicating consistent correction of both under- and over-printing across feature scales. RCWA-based device-level evaluation further confirms that OPC-corrected patterns reduce transmission and phase deviations over 400–650 nm, improving agreement with the intended optical responses.



TECHNICAL SESSIONS

Technical Session 09: Placement & Floorplanning Algorithms

🕒 16:00-18:00, May 8, 2026

📍 Orchid 4203

👤 Session Chair: **Zhou, Chencan**, Nantong University

16:00-16:20 | Paper ID: 4

Title: ILSDP: An Iterated Local Search Algorithm for Routing-Aware FPGA Detailed Placement

Author(s): Canhui Luo, Zhouxing Su, Zhipeng Lü

Presenter: **Luo, Canhui**, Huazhong University of Science and Technology

Abstract: Modern FPGAs face growing routing congestion and wirelength challenges due to increasing logic density and heterogeneous resources. Detailed placement therefore plays a decisive role in coordinating placement quality with post-routing performance and maintaining sufficient margin for routing closure. This paper presents ILSDP, an iterated local search algorithm for routing-aware detailed placement. The framework integrates a two-level evaluation strategy to support effective cell movement, a matching-based defragmentation scheme that reorganizes cells inside configurable logic blocks, and a net-compaction perturbation operator to escape local optima. Evaluated on the 2024 integrated circuit EDA Challenge benchmarks, ILSDP attains first place with a 56.04% overall quality score lead over all participants. Notably, ILSDP outperforms the state-of-the-art placer CAPlace on all instances, reducing overall critical and non-critical wirelength by 51.55% and 14.80%, respectively, while avoiding density degradation. These results demonstrate that ILSDP effectively addresses the wirelength and congestion challenges in modern FPGA detailed placement.

16:20-16:40 | Paper ID: 107

Title: SWIFT-Place: Sequential Wirelength, Interposer Footprint, and Thermal-Aware Placement for 2.5D Advanced Packaging Systems

Author(s): Yubo Hou, Sezin Ata Kircali, Gen Liang Lim, Richard Chang, Mihai Dragos Rotaru, Dutta Rahul, Ashish James

Presenter: **Hou, Yubo**, A*STAR

Abstract: As 2.5D chiplet integration emerges as a de facto standard for high-performance computing, the placement optimization has evolved into a complex multi-physics problem. Existing placement approaches predominantly emphasize wirelength minimization and thermal hotspot mitigation, often overlooking interposer footprint optimization, a critical determinant of manufacturing cost. To address this, we present SWIFT-Place, a sequential wirelength, interposer footprint, and thermal-aware placement framework for 2.5D heterogeneous systems. At the core of SWIFT-Place is a mask-assisted placement paradigm that translates disparate physical constraints into a unified spatial decision space. We introduce a novel Area Mask designed to steer the placement toward high-density configurations, effectively minimizing the interposer footprint. This is integrated with Wire and Thermal Masks to provide a holistic evaluation of each candidate location. To further ensure thermal reliability, we propose an adaptive thermal-spacing mechanism that dynamically adjusts the separation between high-power dies during the sequential placement process, guaranteeing compliance with peak temperature constraints. Experimental evaluations demonstrate that SWIFT-Place achieves superior area utilization and reduced wirelength compared to existing methods, offering a scalable and robust solution for thermally reliable 2.5D package co-design.



16:40-17:00 | Paper ID: 160

Title: Reinforcement Learning-based 3D Floorplanner with Die Swapping Mechanism

Author(s): Qin Luo, Hailiang Li, Evangeline F.Y. Young

Presenter: Luo, Qin, The Chinese University of Hong Kong

Abstract: 3D ICs offer significant advantages in performance, power and area over their 2D designs but introduce floorplanning challenges requiring the co-optimization of block positions and die assignments. While reinforcement learning (RL) has shown promising performance in 2D floorplanning, existing RL-based 3D methods rely on static and pre-assigned die placements, preventing true simultaneous optimization. To overcome this, we propose a novel RL-based 3D floorplanner that features a die-swapping mechanism, enabling the dynamic co-optimization of block placement and die assignment in RL-based approach for the first time. We develop an efficient spatial checking algorithm for available space checks in large layouts to support the die-swapping mechanism. Evaluation on the GSRC benchmark and a new benchmark generated by the OpenROAD suite demonstrates that our approach outperforms existing methods by significantly reducing cross-die wirelength, showcasing the critical importance of joint optimization and strong potential of RL for 3D physical design.

17:00-17:20 | Paper ID: 179

Title: Macro Placement Algorithm Based on Monte Carlo Tree Search and Bayesian Optimization

Author(s): Huicheng Yu, Yue Wu, Xiaoyan Yang

Presenter: Yu, Huicheng, Hangzhou Dianzi University

Abstract: Macro placement is a critical stage in physical design, directly impacting power consumption, performance, and chip area. While existing approaches—whether optimization-based or learning-driven—have achieved certain progress, they still suffer from inefficiencies in search exploration and limitations in overall optimization capability. To address these challenges, this paper proposes a novel macro placement framework that integrates Monte Carlo Tree Search (MCTS) with Bayesian Optimization (BO). A search tree is constructed in which each node represents a macro placement state, enabling global exploration of the solution space for improved search efficiency. Simultaneously, BO is employed to perform iterative local refinement of macro positions, with the scale of the search tree regulated by optimizing leaf nodes. Through continuous iterations, the proposed method is capable of generating high-quality placement solutions. Experimental results on the ISPD2005 benchmark demonstrate that the proposed approach significantly reduces wirelength and improves placement quality, outperforming conventional BO, MCTS, and reinforcement learning-based methods.

17:20-17:40 | Paper ID: 180

Title: ESPlace: Fast Macro Placement Using Evolution Strategy with Partial Evaluation

Author(s): Limei Ma, Handing Wang, Hongzhong Wu

Presenter: Ma, Limei, Xidian University

Abstract: In very large-scale integration physical design, macro placement is critical for determining chip performance. Macro placement can be viewed as an optimization problem that requires evaluating the quality of different layouts during the search process. However, in most existing methods, a single evaluation needs a complete generated layout, which is time-consuming. This leads to inefficient search due to the wastes of significant computational resources on low-potential layouts. Therefore, we propose a fast macro placement optimization method named ESPlace, which adopts (1+1)-evolution strategy as its optimization backbone and incorporates a partial-evaluation guided early stopping mechanism to eliminate low-potential layouts, thereby improving the search efficiency. Experimental results on the ISPD2005 and ICCAD2015 benchmarks show that,



with a 2-hour wall clock time constraint, ESPlace achieves remarkable half-perimeter wire length (HPWL) reductions of 24.73% and 13.77%, respectively, compared to the-state-of-the-art methods. Therefore, ESPlace can produce high-quality layouts in an efficient manner.

17:40-18:00 | Paper ID: 219

Title: Flexible Double-Side Pin Redistribution with Efficient Virtual-Net-Guided Cell Type Assignment

Author(s): Xun Jiang, Haoran Lu, Yifan Chen, Yibo Zhang, Jianxiang Jin, Jiarui Wang, Chunyuan Zhao, Heng Wu, Runsheng Wang, Yibo Lin

Presenter: Jiang, Xun, Peking University

Abstract: The backside process has become a key enabling technology for sub-2nm nodes. To explore the design boundaries of double-side chips, leveraging cells with double-side pins is now at the forefront of design technology co-optimization (DTCO), whose key challenge is redistributing pins to the chip frontside or backside. In this work, we formulate this new problem mathematically and propose virtual-net-guided cell type assignment to solve it. Compared with the state-of-the-art (SOTA) method, our algorithm outperforms in all metrics and supports broader DTCO exploration. We believe our studies will largely stimulate research on double-side physical design and further push cutting-edge VLSI technology.



TECHNICAL SESSIONS

Technical Session 10: Compact Modeling & Parameter Extraction

13:30-16:10, May 9, 2026

Orchid 4301

Session Chair: **Chen, Quan**, Southern University of Science and Technology

13:30-13:50 | Invited Talk



Shen, Chen

Shanghai PFTN Semiconductor Co., Ltd.

Bio.: Dr. Shen obtained his doctoral degree in the Department of Electrical and Computer Engineering at the National University of Singapore in 2009. In 2011, he co-founded Suzhou Kejingda Electronics Co., Ltd. During his 20-year entrepreneurial journey, influenced by the spirit of independent and controllable technological

innovation passed down through three generations of engineers in his family, he has been firmly committed to the R&D of semiconductor device and process TCAD simulation software. After the company's restructuring in 2021, he serves as the Chairman of Peifeng Tunan Semiconductor (the parent company), continuing to devote himself to the R&D and promotion of EDA software for wafer fabrication.

He has published more than 40 papers in international SCI/EI-indexed journals and conferences, with an h-index of 15. His most highly-cited paper has received over 100 citations, and is among the most-cited literatures in the field of MOSFET reliability in recent years.

Title: Modernizing TCAD: High-Performance Solvers and GPU-Accelerated Architectures for Sub-2nm Technology Development

Abstract: As semiconductor scaling progresses beyond the 2nm node, the aggressive introduction of novel materials, complex processing equipment, and 3D device architectures has created a multidimensional design space. The exponential increase in permutations across the material-process-device-design interface presents a significant computational bottleneck, necessitating advanced simulation frameworks to sustain the pace of innovation. Traditional TCAD methodologies are increasingly strained by the rigorous topographical and physical modeling requirements of the nano-scale regime.

In this presentation, we detail a comprehensive modernization of the TCAD framework designed to address these challenges. We discuss the transition to GPU-accelerated computing architectures, the implementation of novel solvers capable of resolving multi-scale, multi-physics phenomena, and the integration of these tools into a new paradigm for co-optimization. By bridging the gap between fundamental physics and industrial-scale development, this work establishes a scalable foundation for the next generation of semiconductor process technology.



13:50-14:10 | Invited Talk**Liu, Wenchao****Primarius Technologies CO., LTD**

Bio.: Dr. Liu Wenchao, currently serving as Senior Vice President at Primarius and General Manager of Primarius Technologies Guangzhou Branch. He has more than 20 years of experience in technology research and management at international semiconductor companies, also his work areas cover chip manufacturing, IC design, and electronic design automation (EDA). In 2004, he obtained a doctor's degree in microelectronics and solid state electronics from Shanghai Institute of Microsystem and Information Technology; From 2004 to 2005, he served as the Principal Engineer of HHGrace; From 2005 to 2009, he served as the Principal Engineer of IBM Semiconductor R&D Center in the United States; From 2009 to 2018, he also served as Senior Technical Manager in Singapore Chartered Semiconductor and GLOBALFOUNDRIES; From 2018 to 2019, served as the Design Platform Director of Unisoc Technologies.

Title: DTCO: Synergizing Design and Manufacturing, Building Application-Driven EDA Flows

Abstract: As semiconductor innovation becomes increasingly application-driven, stronger synergy between design and manufacturing is essential to EDA development. This presentation discusses the importance of DTCO at advanced nodes, where tight co-optimization across process, device, circuit, and layout is critical to achieving PPA and manufacturability targets. It also highlights the value of process exploration at mature nodes, where application-specific optimization continues to unlock competitive advantages. From TCAD and SPICE simulation to cell generation and characterization, Primarius provides a full-flow methodology that connects process innovation with design enablement. By bridging manufacturing insights and downstream EDA optimization, this end-to-end flow helps accelerate technology development, improve design quality, and support application-driven semiconductor innovation across both advanced and mature process nodes.

14:10-14:30 | Paper ID: 74

Title: PINN-2DFET: A Physics-Informed Neural Network for Compact Modeling of 2D FETs

Author(s): Xu Zhang, Yu Geng, Mingyang Kou, Weiqing Ji, Zheng Zhang, Hailong Yao

Presenter: Yao, Hailong, University of Science and Technology Beijing

Abstract: Compact modeling of two-dimensional field-effect transistors (2D FETs) is essential for circuit-level simulation, yet remains highly challenging due to their complex and unique physical mechanisms. While machine learning offers a promising route for rapid compact model development, conventional data-driven approaches often suffer from poor generalization and lack physical interpretability. To address these limitations, this work presents PINN-2DFET, a physics-informed neural network-based compact modeling model for 2D FETs. Unlike most existing methods that directly predict terminal current, PINN-2DFET employs a neural network to infer the internal channel potential distribution, from which the drain current is subsequently computed using physical equations. The model embeds semiconductor physics constraints specifically adapted for two-dimensional materials, incorporating both quantum capacitance and short-channel effects. Evaluated on a small-sample dataset in real-world, PINN-2DFET achieves superior modeling precision, outperforming most comparable approaches, with a LogRMSE of less than 0.001.

14:30-14:50 | Paper ID: 95

Title: A Sensitivity-Guided Progressive Layered Extraction Neural Network for BSIM-CMG Parameter Extraction

Author(s): Yaling Li, Xinyu Gao, Changqin Dong, Junjie Li, Ziji Zhao, Zhaohan Wang, Chengao Zhang, Lixin Ge



Presenter: Li, Yaling, Southeast University

Abstract: Transistor compact models are fundamental to circuit simulation and design, yet conventional model parameter extraction workflows are highly dependent on expert knowledge and require extensive manual calibration. To address those challenges, we propose a novel Sensitivity-Guided Progressive Layered Extraction Neural Network (SGPLE-NN) for efficient and accurate parameter extraction of the BSIM-CMG model. Unlike existing deep learning-based approaches that largely ignore the varying sensitivities of model parameters to device characteristics, our method first perform a comprehensive sensitivity analysis to quantify each parameter's impact on device behavior. Based on this analysis, we design a physics-aware loss function that assigns parameter-dependent weights according to their sensitivity levels. The proposed SGPLE-NN further incorporates shared and task-specific expert networks to mitigate gradient conflicts among parameters with differing sensitivities. Experimental results on the ASAP7 FinFET technology demonstrate excellent accuracy in fitting IV characteristics, achieving an average mean absolute percentage error (MAPE) of approximately 1% across all bias conditions. The proposed framework completes parameter extraction within seconds while preserving physical consistency, offering a practical and scalable solution for rapid device modeling.

14:50-15:10 | Paper ID: 171

Title: AgenticPE: A Dual-Agent LLM Framework with Physics-Gated Reasoning and Trajectory-Aware Optimization for Model Parameter Extraction

Author(s): Yuxiang Zhou, Yu Li, Guibo Luo, Runsheng Wang, Lining Zhang

Presenter: Zhou, Yuxiang, Peking University

Abstract: Parameter extraction for complex compact models, such as BSIM-CMG, involves fitting hundreds of physical parameters to complex measurement data. Due to severe parameter coupling, this process inherently requires a complex, multi-step extraction flow. Traditionally, this relies on labor-intensive manual tuning or black-box numerical optimizers. While the latter can minimize numerical errors within isolated steps, they lack physical understanding and the heuristics to navigate the holistic sequential flow, often trapping the model in non-physical regimes or local minima. Consequently, the process remains heavily dependent on expert orchestration. To address this, we propose AgenticPE, a Human-AI collaborative framework that democratizes expert-level device modeling. AgenticPE features a dual-agent cognitive architecture: a Strategist Agent that emulates expert reasoning via a physics-gated hierarchical protocol to ensure causal consistency, and an Executor Agent that performs trajectory-aware numerical optimization. Unlike traditional solvers, the Executor targets Key Operating Parameter (KOP) discrepancies and utilizes historical trajectory memory to mitigate the aggressive bias of Large Language Models (LLMs). Experimental results demonstrate that AgenticPE successfully decouples physical reasoning from numerical execution, achieving industrial-grade physical accuracy with significantly reduced manual effort.

15:10-15:30 | Paper ID: 182

Title: A Physics-Based OTS Compact Model with Force-Driven Adaptive Damping and Differentiable Hysteresis

Author(s): Yukun Li, Jiajun Qiu, Sannian Song, Lining Zhang

Presenter: Li, Yukun, Peking University

Abstract: This paper presents a physics-based, numerically robust Verilog-A compact model for Ovonic Threshold Switching (OTS) devices, addressing the critical trade-off between physical fidelity and EDA simulation convergence. By analogizing OTS filamentary evolution to ferroelectric polarization, we propose a modified Landau-Khalatnikov (LK) equation featuring a novel force-dependent dynamic damping coefficient $pdyn$. This physically-inspired approach accurately captures complex non-linear switching kinetics and nanosecond-scale transitions. To eliminate convergence failures inherent to traditional "if-else" logic, an



infinitely differentiable smooth clamping function is implemented to ensure derivative continuity, while subthreshold leakage is efficiently described by a simplified Poole-Frenkel mechanism. The model is rigorously validated against experimental data from GeAsSe-based devices with varying As₂Se₃ concentrations (50%, 40%, and 22%). Results demonstrate accurate reproduction of static DC I-V characteristics across six orders of magnitude and high-precision fitting of transient switching times (ton and toff) under 2us triangular pulse actuation. This highly generalizable and stable framework is well-suited for large-scale simulations of 3D stacked memory and neuromorphic systems.

15:30-15:50 | Paper ID: 232

Title: Fast Generation Strategies for cell library characterization in Advanced Technology Node

Author(s): Zonghan Lei, Chaozheng Qin, Wenli Huang, Wenchao Liu, Bin Li, Zhaohui Wu

Presenter: Lei, Zonghan, South China University of Technology

Abstract: Composite Current Source (CCS) models are critical for static timing analysis (STA) and standard cell library characterization. At advanced technology nodes, process variations significantly affect chip performance. To accurately capture statistical distribution with on-chip variations variability during sign-off analysis, the Liberty Variation Format (LVF) has been introduced as the current industrial standard. However, LVF relies on massive Monte Carlo SPICE simulations, leading to enormous computational overhead that severely limits the efficiency of advanced process design flows. To address this challenge, a novel fast library generation framework for advanced technology nodes is proposed in this paper. A Graph Transformer Networks (GTN) model is first employed to efficiently generate CCS models, significantly reducing the number of required SPICE simulations. Subsequently, Gaussian process (GP) surrogate models are utilized to accelerate statistical characterization in the LVF flow. Experiments show that, compared with SPECTRE simulations, the proposed GTN approach achieves prediction errors below 3% for power and delay across different standard cells. Meanwhile, the proposed GP-based statistical modeling significantly accelerates the LVF characterization process, reducing sign-off runtime by over 70%.

15:50-16:10 | Paper ID: 234

Title: IBISgen: An Agent-based Framework for Automated IBIS Model Generation

Author(s): Wentai Huang, Ziwen Wang, Yiming Huang, Shuilong Chen, Lei He

Presenter: Huang, Wentai, The Hong Kong Polytechnic University

Abstract: Accurate I/O buffer modeling is essential for system-level SI analysis, where IBIS models provide a standardized, efficient representation with IP protection. Generating high-quality IBIS models from SPICE descriptions remains labor-intensive and error-prone, often involving substantial manual specification/configuration, iterative checker-driven debugging, and repeated waveform correlation to meet both compliance and behavioral fidelity requirements. To address these challenges, we propose IBISgen, an agent-based framework for automated and verification-driven IBIS model generation. Starting from a SPICE description of an I/O buffer, IBISgen runs DC and transient characterization to obtain reference I--V and V--T responses, constructs candidate IBIS models via a template-driven generator, and iteratively refines the specification using compliance diagnostics and waveform-fidelity metrics. Experimental results on a 1000-case benchmark dataset demonstrate that IBISgen produces IBIS models that pass IBIS v7.x compliance checks while closely matching SPICE reference behaviors. Across diverse buffer types, the generated models achieve an average NRMSE of 2% and an overall FOMcom of 98.97%, while reducing end-to-end model generation time by 66.7× and 9.2× compared to manual and conventional workflows, respectively.



TECHNICAL SESSIONS

Technical Session 11: AI-Driven Front-End & Circuit Optimization

🕒 13:30-15:50, May 9, 2026

📍 Orchid 4303

👤 Session Chair: **Yan, Hao**, Southeast University

13:30-13:50 | Paper ID: 53

Title: Discriminative Joint Representation Learning via High-Fidelity Point Clouds for Pattern Clustering

Author(s): Longyue Wei, Changzhe Jiao, Rongfang Wang

Presenter: **Wang, Rongfang**, Xidian University

Abstract: Traditional layout pattern analysis, whether based on rasterized high-resolution images or downsampled rasters, is fundamentally limited. Rasterization inevitably discards nanometer-scale details, while purely geometric methods struggle with semantic ambiguity in hard negatives—patterns with nearly identical shapes but opposite manufacturability (Hotspot vs. Non-Hotspot). To address these challenges, we propose Point-JR, a discriminative joint representation learning framework for pattern clustering. Point-JR first converts layout data into high-fidelity point clouds using a Lithography-Aware Point Sampling strategy that preserves native vector precision by encoding coordinates and topological normals. It accurately captures fine geometric details critical to lithographic behavior. On this basis, Point-JR jointly optimizes self-supervised geometric invariance and supervised hotspot classification, explicitly embedding lithographic semantics into the feature space. As a result, the learned embeddings can effectively separate visually similar but manufacturability-opposite patterns. Evaluated on the ICCAD 2012 benchmark, Point-JR achieves state-of-the-art performance, reaching 99.38% Cluster Purity and 96.76% Hotspot Neighborhood Consistency, significantly outperforming pixel-, vector-, and graph-based baselines. These results confirm that vector-native point cloud representations combined with joint geometric-semantic learning are essential for generating robust and discriminative features required for design for manufacturability. Dataset, code and trained models are available at <https://github.com/Code-RFGroup/Point-JRL>.

13:50-14:10 | Paper ID: 72

Title: LEGO: An LLM Skill-Based Front-End Design Generation Platform

Author(s): Jincheng Lou, Ruohan Xu, Jiecheng Ma, Runzhe Tao, Xinyu Qu, Yibo Lin

Presenter: **Lou, Jincheng**, Peking University

Abstract: Existing LLM based EDA agents are often isolated task-specific systems. This leads to repeated engineering effort and limited reuse of successful design and debugging strategies. We present LEGO, a unified skill-based platform for front-end design generation. It decomposes the digital front-end flow into six independent steps and represents every agent capability as a standardized composable circuit skill within a plug-and-play architecture. To build this skill library, we survey more than 100 papers, select 11 representative open-source projects, and extract 42 executable circuit skills within a six-step finite state machine formulation. Circuit Skill Builder automates skill extraction with linear scalability. Agent Skill RAG achieves submillisecond retrieval without relying on embedding models. Empirical evaluation on a hard subset of 41 VerilogEval v2 problems that gpt-5.2-codex fails to solve under extra-high (xhigh) reasoning effort shows that individual circuit skills constructed within LEGO raise Pass@1 from 0.000 to 0.805. This is an 80.5% gain over the baseline. Cross project skill compositions also reach 0.805 Pass@1. They outperform hierarchy-verilog by 14.6% and VerilogCoder by 2.5%. They also match MAGE. These results show that modular skill composition supports



both effective and flexible RTL design automation. The LEGO platform and all circuit skills are publicly available at GitHub (<https://github.com/loujc/LEGO-An-LLM-Skill-Based-Front-End-Design-Generation-Platform>).

14:10-14:30 | Paper ID: 113

Title: GT-Fusion: Synergizing LLM Semantics, Topology, and Geometry for PPA Prediction and Optimization

Author(s): Xinhua Lai, He Liu, Miao Liu, Jungang Xu, Xingquan Li

Presenter: Lai, Xinhua, University of Chinese Academy of Sciences

Abstract: Existing AI-based PPA predictors often falter in reconciling the disparity between logical netlists and physical layouts, lacking a holistic view to balance conflicting design constraints. We propose GT-Fusion, a tri-modal framework establishing a Unified Multi-Objective Optimization Paradigm to bridge this gap. The framework synergizes an LLM-driven semantic encoder (Qwen3), a GNN-based topological extractor, and a UNet-based spatial analyzer via a Cross-Modal Injection Mechanism. Extensive experiments on Skywater 130nm benchmarks demonstrate that our tri-modal approach significantly outperforms both single-modal (Pure UNet) and bi-modal (UNet+GNN) baselines. Specifically, GT-Fusion reduces visual congestion prediction error (MAE) by 28.3% and path delay RMSE by 37.5%, proving the necessity of fusing semantic, topological, and geometric views. Furthermore, deployed within the iEDA open-source infrastructure, our model successfully guides placement optimization, delivering significant reductions in timing violations.

14:30-14:50 | Paper ID: 135

Title: OpenOpt: An Open-Source SRAM Optimizer Based on Equivalent Circuit Model

Author(s): Yikai Wang, Yiheng Wu, Can Wang, Bohao Liu, Junhao Ma, Zhuohua Liu, Qinxin Mei, Shan Shen

Presenter: Wang, Yikai, Nanjing University of Science and Technology

Abstract: This paper proposes a co-optimization framework that jointly optimizes SRAM architecture and transistor sizing using equivalent circuit models. The framework simplifies inactive SRAM cells into equivalent RC loads and static power models, achieving up to 61.4X simulation speedup while maintaining high fidelity (read/write delay error <0.22%, power error <1.68%). A joint search space encompassing architecture parameters and device sizing integrates seven algorithms including SA, PSO, Bayesian Optimization variants, and multi-objective evolutionary algorithms. Based on FreePDK45, ablation experiments confirm complementary gains from architecture selection and transistor sizing. Among all algorithms, MOEA/D achieves the best Figure of Merit (8.2721), yielding 6.2% improvement in SNM, 73.6% reduction in area, and 42.3% reduction in peak power. The framework is publicly available at OpenOpt.

14:50-15:10 | Paper ID: 155

Title: CircuitMind: Achieving Human-Level Circuit Efficiency via Multi-Agent Collaboration and Gate-Aware Benchmarking

Author(s): Haiyan Qin, Jiahao Feng, Xiaotong Feng, Wang Kang

Presenter: Qin, Haiyan, Beihang University

Abstract: Large language models (LLMs) often generate behavioral RTL that synthesizes to physically inefficient implementations, creating a gap to human-optimized designs in power, performance, and area (PPA). We propose CircuitMind, a framework that exposes gate-level structure to synthesis tools. It integrates (i) Syntax Locking to enforce synthesis-optimized idioms, (ii) a Self-Evolving Knowledge Base for autonomous optimization learning, and (iii) Dual-Reward Optimization to balance correctness with PPA. On the TC-Bench (our proposed human-aligned benchmark), CircuitMind achieves expert-level performance. Specifically, on a 4 x 8-bit register file, it reduces area, power, and delay by 95.3%, 94.4%, and 58.0%, respectively, with CM(Phi-4) attaining the highest optimization efficiency (SEI=0.115), surpassing strong baselines and matching top-tier human designers. Code and data are available at <https://github.com/BUAA-CLab/CircuitMind.git>.



15:10-15:30 | Paper ID: 176

Title: Magic Gourd: An AI Agent for Interactive Chip Front-End Design

Author(s): Min Cui, Hao Deng, Kexin Chen, Liangliang Huang, Jiaojiao Xu, Yang Liu, Mingxu Song, Chenghua Xu, Chuqiao Zhao, Gang Chen

Presenter: Cui, Min, Nanjing University of Aeronautics and Astronautics

Abstract: The advent of Large Language Models (LLMs) has revolutionized chip design automation, but the human-AI collaboration process remains time-consuming and iterative. We present Magic Gourd, an interactive AI agent that streamlines chip front-end design workflows. Magic Gourd features a three-panel integrated development environment (IDE) with multi-task orchestration and novel Golden Model integration using Coq formal specifications. The system supports multiple LLM providers (DeepSeek, OpenAI, Claude) and generates synthesizable RTL code, testbenches, and design documentation. Experimental results demonstrate 15–25× productivity improvements in design tasks.

15:30-15:50 | Paper ID: 227

Title: RTLPilot: Skill-Driven Multi-Agent System for RTL Code Migration

Author(s): Xiao Tan, Chenyue Li, Sunan Zou, Wei Yan, Guojie Luo

Presenter: Tan, Xiao, Peking University

Abstract: The rapid advancement of large language models (LLMs) has shown great promise in automating Register-transfer level (RTL) code generation. However, existing tools remain confined to single-module designs and ignore opportunity of repository-level migration. This paper introduces RTLPilot, a multi-agent system (MAS) built for RTL project migration, where seven specialized agents, equipped with an integrated skill library, work collaboratively across the planning and generation stages. We also release a RTL migration benchmark comprising 75 module-level pairs and 30 repository-level instances. Experiments show our framework reaches up to 98.7% success for module-level migration and 73.3% for repository-level tasks on various models. On GPT-4o, our work improves over the state-of-the-art RTL generation framework by 14.7% and 10.0% in absolute accuracy, respectively.



TECHNICAL SESSIONS

Technical Session 12: Routing & Layout Engines

🕒 13:30-15:30, May 9, 2026

📍 Orchid 4201

👤 Session Chair: **Feng, Lang**, Sun Yat-sen University

13:30-13:50 | Paper ID: 61

Title: GeoTrans: A Geometry-Driven Transitive Connectivity Resolution Engine for Efficient Layout Net Tracing

Author(s): Chunrun Su, Boyu Shi, Shuo Liu, Chao Wang

Presenter: **Su, Chunrun**, Southeast University; National Center of Technology Innovation for EDA

Abstract: As semiconductor process nodes scale to the nanometer regime, the exponential growth in layout complexity poses formidable challenges to net tracing efficiency. This paper proposes a high-performance Layout Net Tracing (LNT) framework centered on an innovative Transitive Connectivity Resolution Engine. The engine specifically addresses the discrepancy between geometric isolation and electrical continuity by identifying active regions physically bisected by polysilicon gates. By treating high-logic-level gates as conductive bridges, it restores logical connectivity across these geometric gaps, ensuring absolute electrical correctness in complex gate-level structures. To support this core engine, we incorporate context-aware stream filtering to optimize data ingestion and utilize adaptive spatial indexing with dynamic task scheduling to manage computational throughput. Experimental results on official datasets from a national EDA competition demonstrate that the framework achieves 100% accuracy and a speedup of up to 49.8% in single-threaded mode and 32.0% in multi-threaded mode compared to the competition's first-place winner, establishing a new performance baseline for large-scale net tracing.

13:50-14:10 | Paper ID: 83

Title: SNTRoute: Scalable Negotiation-Based Transistor Routing with Dense-Grid Exploration for Standard Cell Layout Synthesis

Author(s): Yinuo Wu, Keyu Peng, Zhengzhe Zheng, Xiqiong Bai, Chao Wang, Ziran Zhu

Presenter: **Peng, Keyu**, Southeast University

Abstract: As semiconductor technology advances to the 7nm node and beyond, the continued scaling of transistor dimensions and reduction in standard cell height have led to a substantial decrease in available routing resources during standard cell synthesis. To address this challenge, we present SNTRoute, a high-performance and scalable negotiation-based transistor routing engine that leverages a denser grid graph to enhance routing flexibility and rapidly achieve desired routing solutions. The proposed framework employs source-region-expansion-based congestion-driven 3D routing to dynamically navigate resource-constrained lower metal layers, enabling flexible path selection and improving routing efficiency under advanced design rules. Additionally, a negotiation-based rip-up and reroute mechanism is integrated to iteratively refine routing paths, resolve congestion, and ensure design rule compliance with minimal runtime overhead. Furthermore, we propose a frontier-based checking technique with linear-time complexity that significantly reduces redundant searches and increases routing flexibility. Experimental results on the ASAP7 cell library demonstrate that SNTRoute achieves a 24% average wirelength reduction compared to hand-crafted layouts. Furthermore, compared to the state-of-the-art router, SNTRoute achieves an average 13% wirelength reduction and 12.92× runtime speedup, showcasing substantial advances in both routing quality and computational efficiency.



14:10-14:30 | Paper ID: 97

Title: Unified Sequence Modeling for Joint PCB Placement and Routing via Causal Transformers

Author(s): Hao Wang, Jienan Chen, Meiqi Li, Silu Xiong, Xiantuo He

Presenter: **Chen, Jienan**, University of Electronic Science and Technology of China

Abstract: Automatic placement and routing (PR) of printed circuit boards (PCBs) is a fundamental step in electronic design, as it directly impacts manufacturability, signal integrity, and overall performance. However, the task is highly challenging due to the vast solution space and the large number of geometric and electrical design rules that must be satisfied simultaneously. We propose a unified learning-based framework for joint PCB placement and routing. Unlike conventional toolflows that separate placement and routing, our method formulates joint PR as conditional next-token prediction using a domain-specific vocabulary. Specifically, we design a tokenizer that serializes board geometry, component context, and routing decisions into a single sequence, which is then modeled by a GPT-style decoder-only Transformer. This autoregressive architecture enables unified generation of placement and routing, while masked training allows the model to learn PCB design rules implicitly from large-scale data. Furthermore, we implement double-layer PCB layout with DRC-compliant layer switching vias across top and bottom layers. Experiments show our approach achieves 94.3% placement similarity to commercial-tool references, 99.7% routing completion, and 1.36× speedup (on GPU) over a commercial EDA tool (on CPU).

14:30-14:50 | Paper ID: 143

Title: Preventing Clock Routing Short with Noisy Label Correction under Worst IR Drop Constraints before Power Planning and Placement

Author(s): Qi Liu, Ming Ling

Presenter: **Liu, Qi**, Southeast University

Abstract: This paper presents a unified early-stage screening framework, termed Binned Gating Fusion with Multiple Kernel Learning (BGF-MKL), which integrates Support Vector Machines (SVMs) enhanced by Multiple Kernel Learning (MKL) with a confidence-binned, sample-adaptive fusion mechanism to jointly screen clock routing short risk and worst static/dynamic IR drop risk before power planning and placement. Unlike prior work focusing solely on clock routing short risk, the proposed framework performs joint risk modeling to enable conservative final screening decisions. To improve robustness against noisy labels induced by tool randomness, CleanLearning is applied during training, while short violation evaluation labels are refined using rule-based statistics of Design Rule Checking (DRC) violations. Experimental results across five open-source benchmarks show that the proposed method consistently achieves around 0.9 accuracy and 0.9 F1-score on average across different noise levels, demonstrating strong cross-design generalization and effectiveness in early filtering of unfavorable configurations, thereby reducing unnecessary downstream design iterations.

14:50-15:10 | Paper ID: 251

Title: RePart: Efficient Hypergraph Partitioning with Logic Replication Optimization for Multi-FPGA System

Author(s): Zizhuo Fu, Yifan Zhou, Zhaoxin Lu, Guangyu Sun, Runsheng Wang, Meng Li, Yibo Lin

Presenter: **Fu, Zizhuo**, Peking University

Abstract: Multi-FPGA systems (MFS) are widely adopted for VLSI emulation and rapid prototyping. In an MFS, FPGAs connect only to a limited number of neighbors through bandwidth-constrained links, so inter-FPGA communication cost depends on network topology. This setting exposes two fundamental limitations of existing MFS-aware partitioning methods: conventional hypergraph partitioners focus solely on cut size and ignore topological structure, and they leave substantial FPGA resources unused due to conservative balance margins. We present RePart, a fully customized multilevel hypergraph partitioning framework for MFS that



integrates logic replication with topology-aware optimization. RePart introduces three coordinated innovations across the multilevel pipeline: FPGA-aware dynamic coarsening, heat-value guided assignment, and replication-deletion supported refinement. Extensive experiments on the Titan23 and EDA Elite Challenge Contest benchmarks show that RePart reduces total hop distance by 52.3% on average over state-of-the-art hypergraph partitioners with an 11.1× speedup, and outperforms the EDA Elite Challenge winners. Code is available at: <https://github.com/Welement-zyf/RePart>.

15:10-15:30 | Paper ID: 269

Title: Delayed Exposure of Clock Routing Short Risk Across Technology Nodes Under PDN Configuration Variations

Author(s): Qi Liu, Ming Ling

Presenter: Liu, Qi, Southeast University

Abstract: Clock routing short violations have been used as early indicators for screening improper Power Distribution Network (PDN) configurations before costly downstream implementation. However, when extending this event-based screening strategy across technology nodes under PDN configuration variations, we observe that clock routing short violations are not consistently exposed at the clock routing snapshot. In some technology node and PDN configuration combinations, short violations appear at the clock routing snapshot, whereas in others delayed exposure defers the clock routing short risk to post-routing clock-net Design Rule Checking (DRC) involvement. To address this exposure-stage shift, we introduce a stage-aware evaluation strategy that pairs an early clock routing short label with a posterior reference defined by post-routing clock-net DRC violations. Across multiple designs and configuration regimes under a fixed implementation flow, cross-design leave-one-design-out (LODO) experiments show that ranking-based screening remains robust despite exposure-stage shift. In particular, the Area Under the Receiver Operating Characteristic curve (AUC-ROC), as a ranking metric, consistently remains around 0.9 on average under both evaluation labels, and under the posterior label the best-performing classifier achieves Accuracy no lower than 0.84.



TECHNICAL SESSIONS

Technical Session 13: Reliability, Device-Circuit Co-Optimization & Emerging Tech

🕒 13:30-16:30, May 9, 2026

📍 Orchid 4202

👤 Session Chair: **Yu, Bei**, The Chinese University of Hong Kong

13:30-13:50 | Paper ID: 9

Title: GNN-reliability: A Function-aware Graph Neural Network for Reliability Prediction of Gate-level Netlist

Author(s): Zongta Li, Liang Yang, Hao Li, Mian Lou, Yingqi Li

Presenter: **Li, Zongtai**, Xi'an Microelectronic Technology Institute

Abstract: High-energy particles in the space radiation environment tend to induce soft errors in digital circuits, which seriously threaten the reliability of aerospace integrated circuits. The triple modular redundancy hardening strategy is widely adopted to ensure circuit reliability, yet it remains challenging to effectively identify which logic gates require hardening. Recent methods leverage graph neural networks (GNNs) to extract only the structural features of netlists for logic gate reliability prediction, resulting in suboptimal prediction accuracy. This paper proposes a novel machine learning-based framework named GNN-reliability, which first models the netlist as a directed acyclic graph (DAG) and then employs a customized GNN to simultaneously extract the structural and functional features of post-mapping netlists, thus enables rapid reliability prediction of logic gates in unseen netlists at the early stages of circuit design. Experimental results demonstrate that the proposed GNN-reliability outperforms not only four typical GNNs and six directed GNN variants, but also the state-of-the-art (SOTA) combinational representation method FGNN, which validates the model's practicality for netlist reliability evaluation.

13:50-14:10 | Paper ID: 146

Title: Graph Neural Network-based Prediction of SEU-type Sensitive Registers in Gate-level VLSI Circuits

Author(s): Yiqing Li, Huixin Peng, Yan Liang, Fei Xiao, Huimeng Guo, Mingwei Mi, Liang Wang, Yuanfu Zhao

Presenter: **Li, Yiqing**, Beijing Microelectronics Technology Institute

Abstract: Soft errors induced by single event upsets (SEUs) pose growing reliability challenges in very large-scale integration (VLSI) circuits. This paper presents a graph neural network (GNN)-based method for predicting SEU-sensitive registers in gate-level netlists under a partially labeled setting. By leveraging sensitivity labels from a subset of registers within the same circuit, the proposed approach infers the vulnerability of unlabeled nodes, thereby reducing fault simulation overhead. The netlist is modeled as a directed graph that captures structural connectivity and logic-depth features. Three representative GNN architectures—GCN, GAT, and GraphSAGE—are evaluated on an AES circuit, where GAT achieves the best performance with a precision of 90.18% and an F1 score of 0.95. Compared with commercial SSIM-based fault injection, the proposed method achieves a 17× speedup while maintaining competitive prediction accuracy. The framework enables fast and scalable reliability assessment within electronic design automation (EDA) flows.

14:10-14:30 | Paper ID: 188

Title: Physics-Informed and Uncertainty-Aware Surrogate Modeling of Single-Event Transients for Aerospace Applications

Author(s): Yaoyu Li, Bi Wang, Hanbin Wang, Chao Wang, Zhaohao Wang, Yingqi Ma

Presenter: **Li, Yaoyu**, Beihang University



Abstract: Single-event transient (SET) effects are a major reliability concern in deeply scaled CMOS technologies, where a single energetic particle can induce transient currents that propagate to logic nodes and cause soft errors. Although technology computer-aided design (TCAD) provides accurate SET waveform simulation, its extremely high computational cost limits large-scale circuit-level reliability analysis. This work presents a physics-informed, time-aware convolutional neural network (CNN) that serves as a fast and reliable surrogate model for SET current prediction. The framework integrates four physics-motivated constraints into a composite loss function—emphasizing peak regions, temporal alignment, waveform monotonicity, and pre-strike silence—to preserve the underlying charge-collection physics. In addition, Monte Carlo Dropout (MC-Dropout) enables quantitative uncertainty estimation in each predicted waveform, providing confidence awareness that is critical for reliability sign-off. Trained on TCAD-generated data for a 12-nm FinFET process, the model achieves $R^2 > 0.99$ for key physical metrics such as peak current and collected charge, while offering a speedup of $> 10^6\times$ over TCAD simulations. These results demonstrate that embedding physical priors and uncertainty quantification into neural surrogate models can significantly enhance both the accuracy and trustworthiness of SET reliability evaluation.

14:30-14:50 | Paper ID: 264

Title: On the Understanding and Modelling of Negative Bias Temperature Instability (NBTI) under AC Conditions

Author(s): Yongkang Xue, Chang Liu, Menghua Wang, Niuyi Sun, Dan Yang, Na Mei, Keqing Ouyang

Presenter: Xue, Yongkang, Sanechips Technology Co., Ltd.

Abstract: The reliability assessment of AC Negative Bias Temperature Instability (NBTI) at GHz frequencies faces significant challenges due to measurement limitations and complex defect kinetics. This work presents a comprehensive investigation into the frequency-dependent NBTI degradation mechanisms of HKMG pMOSFETs using a unified physics-based model calibrated against measurement data from DC to 1 MHz. By deconvoluting the total degradation into contributions from oxygen vacancies (VO), hydrogen bridges (HB), nitrogen interstitials (Ni), and border traps, the distinct frequency-dependent responses of capture and emission dynamics for each defect species are elucidated. The analysis reveals that while VO-induced degradation at End-of-Stress (EOS) decreases with increasing frequency due to insufficient capture time, HB and Ni induced degradation at EOS remain frequency-insensitive owing to their ultra-fast τ_c . However, all defect types exhibit increased residual degradation at End-of-Recovery (EOR) at higher frequencies, attributed to incomplete emission from traps with large τ_e . Crucially, extrapolation to the GHz regime predicts a fundamental shift in the dominant degradation mechanism: as frequency-dependent components (V_o) are suppressed, border traps emerge as the primary reliability limiter, exhibiting negligible frequency dependence. Consequently, the total AC NBTI degradation at 10 GHz converges with that at 1 MHz. This finding validates that physics-based models calibrated at conventional test frequencies (e.g., 1 MHz) can be reliably extrapolated to predict circuit reliability in the GHz domain, offering a robust and efficient framework for the design of next-generation high-speed integrated circuits.

14:50-15:10 | Paper ID: 93

Title: Reliability-aware DTCO with Physics-Constrained Machine Learning Framework for MOSFETs Trap Extraction

Author(s): Jinghan Xu, Zheng Zhou, Yongjia Wang, Shuhan Wang, Ligong Zhang, Xiaoyan Liu,

Presenter: Xu, Jinghan, Peking University

Abstract: Traps in MOSFETs pose severe reliability issues and affect circuit performance, yet extracting trap distribution is difficult due to diverse trap types and their complex energy and spatial distributions. In this work, we propose a general Physics-Constrained Machine Learning (PCML) framework, which inherently



applies strong physics constraints without requiring extensive training, enabling accurate trap extraction applicable to diverse process technologies. Leveraging these results, we provide a reliability-aware Design-Technology Co-Optimization (DTCO) flow that includes a trap-informed compact model for circuit-level reliability assessments and targeted technology optimization, achieving a 30% reduction in BTI effects and a tenfold decrease in low-frequency noise. The benefits of this optimization were validated on ring oscillator (RO) circuits, reducing aging-induced frequency degradation to one-half of its initial value, demonstrating significant improvements in circuit-level reliability and performance.

15:10-15:30 | Paper ID: 197

Title: Toward Aging-Aware Silicon Lifecycle Management in Advanced Technology: In-Situ Monitor Design, Deployment and AVS Architecture

Author(s): Hongbei Jin, Feiyang Shu, Huizhen Qiu, Yewei Zhang, Pengpeng Ren, Zhiguo Yu, Zhigang Ji

Presenter: Jin, Hongbei, Jiangnan University

Abstract: In recent years, Silicon Lifecycle Management (SLM) has gained significant attention due to the growing demands for system performance, reliability, and sustainability. By deploying in-situ monitors embedded within silicon designs, real-time chip performance data are acquired, which in turn can be used to guide Adaptive Voltage Scaling (AVS) to eliminate pessimistic voltage guardbands. However, traditional monitoring schemes suffer from excessive area overhead and complex hardware integration. Furthermore, monitor efficacy is compromised by path inactivity, as timing alerts are triggered only when signal toggling occurs. To address these issues, the proposed in-situ early-warning monitor integrates signal toggle detection to mitigate the risk of erroneous voltage scaling caused by path inactivity, while employing a multiplexer-based sharing structure to minimize the overall monitoring cost. Additionally, a monitor deployment strategy based on path structural similarity is introduced to eliminate redundant monitoring points, further enhancing area efficiency. Finally, an AVS controller with status aggregation logic is implemented to consolidate outputs from multiple monitors for guiding global voltage regulation. Implemented on a 7-nm RISC-V processor, the design achieves a 22.4% initial power reduction while maintaining 7.3% savings at the end-of-life stage.

15:30-15:50 | Paper ID: 168

Title: H-COSY: Hardware-Calibrated Co-Optimization for Device-Circuit Synergy in Ferroelectric SNNs

Author(s): Qincheng Yang, Xianyu Wang, Jiajun Qiu, Guangyu Sun, Runsheng Wang, Lining Zhang

Presenter: Yang, Qincheng, Peking University

Abstract: This paper proposes an automated application driven Design-Technology Co-Optimization (DTCO) framework that bridges the gap between ferroelectric device physical parameters, circuit design parameters, and system-level performance metrics. The framework integrates three key components: a physics-informed fast evaluation engine, a commercial circuit solver (Cadence Virtuoso), and an optimizer. Rather than relying on computationally expensive full-circuit simulations, we employ the evaluation engine as a hardware-calibrated proxy. Leveraging a dynamic Look-Up Table (LUT) mechanism to extract accurate weight variation and power characteristics from transistor-level simulations, this engine enables the rapid assessment of accuracy and power consumption for ferroelectric-based Spiking Neural Networks (SNNs). Targeting the specific nucleation-limited switching dynamics of ferroelectrics, the framework automatically explores the high-dimensional design space to co-optimize device switching dynamic (τ_0 , E_c) and neuron integration characteristics (C_{mem} , V_{th}). Experimental results on the MNIST dataset demonstrate that the co-optimized solution achieves a peak accuracy of 85.5% (a 5.5% improvement over the baseline) while simultaneously reducing power consumption by 45.6%. Furthermore, sensitivity analysis reveals distinct physical drivers—identifying coercive field as the primary determinant for accuracy and membrane capacitance for



power—validating the effectiveness of integrating precise device- circuit modeling into the architectural exploration loop.

15:50-16:10 | Paper ID: 206

Title: Enabling High-Performance VGAAFET SRAM through Gate-Via Placement and Coupling Interconnect Co-Optimization

Author(s): Xinlong Guo, Tao Liu, Lewen Qian, Meicheng Liao, Min Xu, David Wei Zhang

Presenter: Guo, Xinlong, Fudan University

Abstract: Vertical gate-all-around FETs (VGAAFETs) have emerged as a promising device architecture for advanced-node SRAM, owing to their superior electrostatic control and the capability to enable zero diffusion break (ZDB) layouts for enhanced area scalability. In this work, a comprehensive design-technology co-optimization (DTCO) framework combining TCAD and SPICE simulation is employed to systematically investigate VGAAFET 6T SRAM. The analysis focuses on gate-via configurations located on the long side of the nanosheet, considering different via placements and coupling interconnect layers. The results show that relocating the vias from the middle to the side reduces the capacitance of storage nodes Q and QB by 11.8%, which can be further improved to 22.2% when the coupling interconnects are implemented in the M0 layer. These parasitic reductions translate into significant SRAM performance gains, including up to 6.1% and 5.6% reductions in read and write delay, respectively, along with decreases of 12.3% and 10.8% in read and write energy-delay product (EDP). Overall, this study provides meaningful insights into the via-aware co-optimization of device, interconnect, and circuit design, offering practical guidance for future high-speed and low-power SRAM technologies.

16:10-16:30 | Paper ID: 220

Title: Gradient-Driven Yield Optimization for Carbon Nanotube Circuits

Author(s): Xingyang Li, Wei W. Xing, Yuanqing Cheng, Yuquan Sun

Presenter: Cheng, Yuanqing, Beihang University

Abstract: Carbon nanotube (CNT) circuits promise superior energy efficiency beyond silicon, yet their adoption is blocked by severe manufacturing yield challenges. Traditionally, optimizing this yield is computationally intractable, as state-of-the-art methods treat complex, physics-based CNT simulators as opaque black boxes, requiring hundreds of slow simulations. Our core insight is that this black-box assumption is flawed: the CNT yield objective can be reformulated into a differentiable form. We propose a novel framework that exploits this property. We construct a differentiable Gaussian Process surrogate model from sparse simulations and, for the first time, perform direct gradient ascent on the yield function. By introducing an antithetic sampling technique to stabilize the statistical gradient, our method achieves near-optimal (99.7%) yield on complex CNT-SRAM and OTA circuits while delivering up to approaching 10× speedup over state-of-the-art Bayesian optimization. This work transforms CNT yield optimization from an intractable, multi-hour task to a practical, interactive process, removing a key barrier to postsilicon technology adoption.



TECHNICAL SESSIONS

Technical Session 14: In-Memory Computing: Architectures for AI, Security, and Beyond

🕒 13:30-15:50, May 9, 2026

📍 Orchid 4203

👤 Session Chair: **Li, Xueqi**, Institute of Computing Technology, Chinese Academy of Sciences

13:30-13:50 | Paper ID: 17

Title: BioAP: A Bidirectional Digital-Slope ADC-Based In-Memory Architecture for Output-Aware Point Cloud Feature Computation

Author(s): Xinsong Sheng, Xuliang Yu, Wei-Han Yu, Zhichao Tan, Liang Zhao

Presenter: **Sheng, Xinsong**, Zhejiang University

Abstract: Point cloud neural networks (PCNNs) are dominated by massive fine-grained MLP operations over irregular point neighborhoods, making their performance in CIM-based accelerators largely limited by ADC sensing latency and bit-serial computation depth. Existing RRAM-based CIM designs rely on independent full-range ADC conversions for each output, failing to exploit strong output-domain numerical locality under shared MLP weights. This work proposes BioAP, a charge-domain RRAM-based CIM architecture that targets these two bottle necks. BioAP integrates a bidirectional digital-slope ADC (BDS ADC) with a 2T1R1C CIM array to reduce per-output ADC conversion latency by exploiting output locality, and incorporates a sensitivity-guided bit-allocation (SGBA) framework to shorten the bit-serial activation computation path using 4-bit inputs while enabling sub-4-bit average weight precision with minimal accuracy loss. Implemented in commercial 65 nm technology, BioAP achieves 1.14 TOPS peak throughput and 51.92 TOPS/W energy efficiency, delivering up to 1.94× higher throughput and 2.22× lower energy consumption than conventional SAR-ADC based CIM designs on representative PCNN benchmarks.

13:50-14:10 | Paper ID: 37

Title: Tackling KV-Cache Stress with Heterogeneous Dynamic LLM Inference

Author(s): Wanqian Li, Zerun Li, Xinyu Wang, Xiaotian Sun, Xinjiang Zhang, Xiaoyu Zhang, Xueqi Li, Xiaoming Chen

Presenter: **Chen, Xiaoming**, Institute of Computing Technology, Chinese Academy of Sciences

Abstract: Processing-near-memory (PNM) architectures have been explored to mitigate memory bottlenecks in LLM inference through heterogeneous integration with GPUs/NPUs. However, they still face two key challenges as inference scales. First, the increasing KV-Cache stress limits throughput by system memory capacity. Second, diverse workloads reduce efficiency under static scheduling, which fails to adapt to dynamic workload variations. We propose DyInfer, a heterogeneous dynamic inference system integrating GPUs and DIMM-based PNM (DIMMPNM). DyInfer leverages the scalable capacity of DIMM-PNM to relieve KV-Cache stress and employs workload-adaptive dynamic scheduling to adjust policies in real time, sustaining efficient serving. Experiments show that DyInfer achieves higher LLM inference performance than state-of-the-art designs.

14:10-14:30 | Paper ID: 44

Title: Triple-Encrypted Stochastic Computing-in-Memory Architecture for Secure CNN Inference

Author(s): Yaqi Wang, Jiayao Wu, Xinrui Duan, Wen Shi, Weisheng Zhao, Yijiao Wang

Presenter: **Wang, Yaqi**, Beihang University

Abstract: User authentication plays an increasingly critical role in security applications, where convolutional



neural network (CNN)-based authentication has achieved superior accuracy. However, model theft in such scenarios poses severe risks: attackers can harvest user data with the cloned model, or craft malicious inputs to bypass authentication. Therefore, secure model deployment on edge devices is essential. This paper proposes a Triple-Encrypted Stochastic Computing-in-Memory (TE-SCIM) architecture based on a spin-orbit torque magnetic random-access memory (SOT-MRAM) array, which employs encrypted weights to prevent attackers from acquiring model weights and enhances resilience against malicious inputs. The main contributions include: 1) A dual-mode architecture, which supports key restoration via the Physical Unclonable Function (PUF) mode and in-situ decryption through the SCIM mode. It ensures that keys and model weights remain in protected (masked or encrypted) form during transmission and are recovered only on-chip, thereby mitigating eavesdropping risks over insecure edge-cloud channels. 2) Triple encryption for model weights, which expands the key space by 4× compared to prior CIM encryption. 3) A two-layer defense against adversarial attacks, which reduces the attack success rate (ASR) by ~30% compared to the conventional paradigm. Besides, with 22-nm FDX, the TE-SCIM reaches an energy efficiency of 198 TOPS/W, a 1.5× gain over state-of-the-art encrypted CIM works.

14:30-14:50 | Paper ID: 133

Title: CoCIM: A Collaborative Design Framework for Multi-Network Shared Compute-in-Memory Architectures

Author(s): Xiaoxiao Zhou, Wanqian Li, Rui Liu, Yuping Wu, Xinjiang Zhang, Xueqi Li, Xiaoming Chen

Presenter: Zhou, Xiaoxiao, Institute of Computing Technology, Chinese Academy of Sciences

Abstract: Compute-in-memory (CIM) architectures based on crossbar arrays offer a promising solution to the memory wall bottleneck for neural network acceleration. While existing design space exploration frameworks effectively optimize hardware for single dedicated networks, they fail to address the deployment of multiple networks on a CIM accelerator. This limitation causes significant efficiency degradation (up to 68% in our experiments), when reusing a single-network optimized architecture due to conflicting hardware preferences. To bridge this gap, we propose CoCIM, a collaborative design framework for multi-network shared CIM architectures. CoCIM maximizes average and worstcase energy efficiency across target networks by solving a resource-constrained optimization problem. The framework employs a decoupled two-stage search strategy: first generating topk candidate configurations for individual networks, then executing global joint optimization to resolve cross-network contention. This decoupled approach ensures computational feasibility while retaining diverse high-potential candidates to approximate global optimality. By distinguishing between architecture-level fixed parameters such as crossbar size and device resolution versus deployment-time configurable parameters like mapping strategy, CoCIM identifies shared configurations balancing diverse workload requirements. Experimental results show that, when running diverse CNNs, CoCIM improves the energy efficiency by 29.6% in average case and 19.5% in worse case, compared with baselines which are optimized based on a single network.

14:50-15:10 | Paper ID: 169

Title: An MRAM-Based TCAM with Static Match-Line and Search-Line Gating for Low-Power and High-Speed Search

Author(s): Shiheng Li, Humiao Li, Longyang Lin

Presenter: Li, Shiheng, Southern University of Science and Technology

Abstract: This paper proposes a high-performance, energy-efficient, and robust STT-MRAM-based 14T-4MTJ non-volatile TCAM (nvTCAM) in 40-nm CMOS that addresses the dynamic power and reliability limitations of prior MRAM-based designs. A segmented static match-line (SML) architecture adopting a match-to-discharge scheme minimizes ML switching activity, while a data-aware search-line gating (SLG) mechanism detects identical consecutive search keys and suppresses redundant search operations. Compared with a conventional



dynamic ML (DML) baseline, the proposed SML and SLG techniques reduce power consumption by 11.5 percent and 75.6 percent, respectively, for a total reduction of 87.1 percent. The design occupies 4.09 square micrometers per cell and achieves a search delay of 0.151 ns with a search energy of 0.84 fJ/bit/search. Furthermore, Monte Carlo simulations demonstrate a 9.02 sigma search yield with zero search error rate. Compared with state-of-the-art nvTCAMs, this work improves the energy–delay–area product (EDP-A) by up to 216.5 times, making it well suited for energy-constrained associative computing applications.

15:10-15:30 | Paper ID: 244

Title: NICE: 3D-NAND-based In-Memory-Computing with In-Situ ECC-Protection for Fault-Tolerant and Efficient LLM Inference

Author(s): Boyi Fu, Weikai Xu, Tong Xie, Jin Luo, Yaoyu Tao, Meng Li

Presenter: Fu, Boyi, Peking University

Abstract: Large language models (LLMs) demand massive weight storage, making edge deployment challenging. 3D NAND Flash offers high density for hosting LLM weights, and analog compute-in-memory (CIM) can improve throughput by reducing data movement and enabling multi-wordline parallelism. However, analog NAND-CIM reliability is limited by raw bit error rate (RBER) and read-out variation, while conventional NAND ECC and prior CIM-oriented codes are not directly compatible with in-situ MAC pipelines. We propose NICE, an ECC-protected analog 3D NAND-CIM architecture that interleaves parity/check-bit columns along the page direction and computes syndromes in parallel with MAC outputs, enabling on-die correction without reducing CIM parallelism. On HellaSwag with Qwen3-30B-A3B (W8A4, 6-bit ADC), NICE reduces codeword FER by 66.4%/34.9% versus No-ECC/single error correction at the highest error point, improving accuracy to 62.0% from 24.7% and 51.9%. End-to-end, NICE incurs only 0.7% throughput loss versus CIM w/o ECC (7.63->7.58 tokens/s) while achieving a 4.59× speedup over a von Neumann baseline.

15:30-15:50 | Paper ID: 258

Title: Toward Reliable Analog In-Memory Computing: Co-Design of Hardware-Aware Input Processing and Accuracy-Optimized Weight Mapping

Author(s): Guangyao Wang, Guilin Zhao, Yuxin Xie, Meng Wang, Wenying Zhou, Hanjie Liu, Yuannuo Feng, Yuexi Lv, Wang Kang

Presenter: Wang, Guangyao, Beihang University

Abstract: Analog in-memory computing (AIMC) has emerged as a promising architecture for accelerating neural network (NN) inference due to its high energy efficiency and parallelism. However, large-scale deployment remains challenged by input processing methods that fail to capture hardware trade-offs, together with accuracy degradation under analog noise. This work presents a hardware-aware input processing and weight mapping framework for AIMC, implemented and evaluated on a commercial 40nm eFlash IMC SoC. The proposed scheme dynamically selects input-processing strategies according to latency, area, and accuracy requirements, while the weight mapping techniques mitigate quantization and analog distortions. On-chip evaluations show that the three input-processing methods achieve distinct trade-offs, with sign reversal improving single-layer accuracy by up to 42.6%, and weight doubling and replication strategies yielding average gains of 54.8%, 41.1%, and 41.0%. End-to-end deployment of the DCCRN speech model further demonstrates up to 35.9% reduction in per-layer error deviation and a 0.886 increase in PESQ. These results highlight the effectiveness of the proposed methodology in enabling system-level AIMC deployment under real-time and resource-constrained AI scenarios.



TECHNICAL SESSIONS

Technical Session 15: Device Characterization, Cryogenics & Cross-Layer Design



10:00-12:20, May 10, 2026



Orchid 4301



Session Chair: **Shao, Leilai**, Shanghai Jiao Tong University

10:00-10:20 | Paper ID: 70

Title: Compact Modeling and Characterization Strategies for CMOS Technology Down to 9 mK: From Physics to Simulation

Author(s): Zhizhao Ma, Shilong Li, Yuhuan Lin, Kai Chen, Quan Chen

Presenter: **Ma, Zhizhao**, Southern University of Science and Technology

Abstract: The absence of accurate Electronic Design Automation (EDA) models at milli-Kelvin (mK) temperatures constitutes a major bottleneck for scaling cryo-CMOS quantum interfaces. Conventional compact models lose physical validity in the deep cryogenic regime due to unmodeled non-ideal quantum and transport effects. This work proposes a physics-augmented compact modeling framework tailored for commercial 40 nm and 180 nm technologies, calibrated from 300 K down to 9 mK. By introducing temperature-dependent analytical corrections to the core model formulation, we successfully overcome the convergence issues inherent in standard models at near-zero Kelvin. The proposed methodology demonstrates excellent agreement between SPICE simulations and experimental measurements, enabling a reliable design library for hybrid quantum-classical circuits.

10:20-10:40 | Paper ID: 104

Title: Tail-Aware Variability Simulation for Planar CMOS under Cryogenic Condition: From Compact Modeling to SPICE Integration

Author(s): Yongqiang Zhong, Junwei Guo, Chenyang Zhang, Xiaochang Zhang, Pengpeng Ren, Runsheng Wang, Zhigang Ji

Presenter: **Zhong, Yongqiang**, Shanghai Jiao Tong University

Abstract: Cryogenic variability in planar CMOS can exhibit strong non-Gaussian behavior, significantly complicating circuit sign-off. Statistical measurements on 40-nm HKMG planar nMOS devices over the temperature range 300–77 K reveal long-tailed threshold voltage distributions that emerge at cryogenic temperatures. Analysis indicates that these tails arise from the coupling between work-function variation (WFO) and band-tail-state variation (BTV), which we model using a mixed normal-lognormal distribution. To enable circuit-level evaluation, we develop a BSIM4-compatible cryogenic compact model that incorporates subthreshold-slope saturation through an effective-temperature formulation and embeds WFO–BTV coupling into the carrier statistics. This model is integrated into NGSPICE to support efficient corner and Monte Carlo analysis. Application to a 6T SRAM demonstrates that the coupled distribution tails can increase the read V_{min} by more than 0.5 V (6σ), while native integration within NGSPICE achieves a 13.6× speedup compared with a Verilog-A HSPICE flow.

10:40-11:00 | Paper ID: 126

Title: First Demonstration of Cell Level Electro-Thermal Coupled Full TCAD Simulation Based on Quantum Transport Framework

Author(s): Suteng Zhao, Hongwei Zhou, Zhongzhen Tong, Lang Zeng

Presenter: **Zhao, Suteng**, Beihang University

Abstract: Complementary Field-Effect Transistors (CFETs) are the leading candidates for sub-2nm nodes, yet



their vertical stacking architecture induces severe self-heating effects (SHE). Accurate thermal evaluation at this scale requires a rigorous framework that simultaneously accounts for quantum transport, phonon-scattering-based heat generation for replacing the inadequate Joule heating mode and full-cell coupling with the interconnect hierarchy. In this work, we developed and demonstrated the first 3D Cell-level electro-thermal coupled full TCAD simulation framework based on the Non-Equilibrium Green's Function (NEGF) method. This tool enables the simultaneous solution of quantum transport and thermal equations across a complete 3D inverter cell, encompassing FEOL, BEOL, and BSPDN structures. Our 3D simulations reveal a significant thermal disparity: the NFET exhibits a peak lattice temperature 21 K higher than the PFET. This is attributed to the quasi-ballistic nature of transport, where 45.8% of heat in the NFET is localized within the intrinsic device region, compared to only 12.7% in the PFET, which dissipates most of its heat into the macroscopic drain extensions. Furthermore, the self-heating effects result in a 7.54% drive current degradation for the NFET versus 5.4% for the PFET. These results underscore the necessity of our 3D quantum transport based thermal framework for the precise design and thermal management of sub-2nm CFET technology.

11:00-11:20 | Paper ID: 103

Title: KAN-CHAR: Efficient Cell Library Characterization for DTCO via Kolmogorov-Arnold Graph Learning

Author(s): Shubo Xu, Zeyuan Deng, Zhi Li, Peng Cao

Presenter: Xu, Shubo, Southeast University

Abstract: Design-Technology Co-Optimization (DTCO) is crucial for achieving optimal power, performance, and area (PPA) in advanced technology nodes. However, its efficacy is severely bottlenecked during the exploration of fine-grained process parameter space by the prohibitive cost of iterative standard cell library characterization. Existing GNN-based methods typically rely on Multi-Layer Perceptrons (MLPs) for feature transformation and prediction, which are inefficient at approximating the highly non-linear physical response of transistors to process parameter fluctuations. To address this, we propose KAN-CHAR, an efficient cell library characterization framework that integrates Kolmogorov-Arnold Networks (KANs) into a heterogeneous graph attention mechanism. By replacing linear projections with learnable spline functions and explicitly embedding physics-derived sensitivity priors into heterogeneous node features, KAN-CHAR captures the distinct responses of devices to process parameter fluctuations with minimal data. Validated on the ASAP7 PDK, our framework achieves a MAPE of 1.61% and a MAE of 0.54ps for cell delay on unseen process parameter configurations, delivering a 98.8% runtime reduction compared to traditional flow.

11:20-11:40 | Paper ID: 117

Title: High-Density NAND-Like In-Plane SOT-MRAM with Robust and Efficient Write Scheme

Author(s): Xiaoyang Xu, Chenyi Wang, Mingche Li, Weimeng Zhao, Zhongzhen Tong, Yaling Wang, Chao Wang, Zhaohao Wang

Presenter: Xu, Xiaoyang, Beihang University

Abstract: Spin-orbit torque magnetic random-access memory (SOT-MRAM) demonstrates potential to overcome the write speed and energy bottlenecks of spin transfer torque MRAM (STT-MRAM). Among various SOT-MRAMs, the in-plane magnetic anisotropy MTJ (iMTJ) is widely adopted in tape-outs due to its intrinsic field-free switching capability, superior write/read performance, and mature process. However, the area efficiency of SOT-MRAM is significantly constrained by the need of two access transistors, and most existing area-optimized designs target perpendicular magnetic anisotropy MTJ (pMTJ). This work presents a NAND-like in-plane SOT-MRAM (iNAND-SPIN) utilizing a robust and efficient write scheme. The design leverages the intrinsic disparity between SOT and STT switching currents to establish well-defined write operation windows. Compared to typical SOT architectures (2T-1MTJ), the proposed iNAND-SPIN achieves an area reduction of up to 37.5%. Unlike conventional NAND-SPIN, which suffer from a reliability-area-power trade-off that limits the write margin to only 3.5, the proposed iNAND-SPIN delivers a 2.89× larger write margin while reducing per-bit



write energy by 64% in 2-bit multi-level cell (MLC). Furthermore, it successfully enables 4-bit MLC, which is unattainable with conventional iMTJ based designs, while maintaining a transistor count of only 1.25 per bit. A dedicated write scheme is devised for the 4-bit MLC, significantly lowering the required write voltage and the average write cycles. Compared to the proposed 2-bit MLC, the optimized 4-bit write scheme further reduces write power consumption by 26.7%.

11:40-12:00 | Paper ID: 239

Title: Parasitic-aware SRAM Designs with 2nm Predictive Technology Incorporating Backside Contact

Author(s): Sihao Chen, Baicheng Lin, Jie Lin, Yushen Zhang, Ulf Schlichtmann, Lining Zhang, Runsheng Wang

Presenter: Chen, Sihao, Peking University

Abstract: In this work, we propose a holistic, parasitic-aware design flow for backside contact (BSC)-integrated high-density (HD) and high-performance (HP) static random-access memory (SRAM) designs, featuring accurate layout-dependent parasitic extraction calibrated by an industry-standard 3D field solver. The proposed flow integrates both self-heating effect (SHE) compact modeling and cell-level parasitic extraction, enabling model-driven, comprehensive evaluation of power, performance, area, and temperature (PPAT) metrics across diverse SRAM bitcell architectures. Compared to conventional 3nm frontside power delivery network (FSPDN) SRAM counterparts, the proposed BSC based HD and HP SRAMs achieve 34% and 30% cell height (CH) reduction via complete power-signal routing decoupling. Quantitative results demonstrate that the BSC-based SRAM achieves systematic reductions in parasitic capacitance and improvements in energy efficiency under the coupled constraints of parasitic effects and SHE. Specifically, it delivers up to a 16.9% reduction in read power, a 2.1% reduction in write access time, and a 16.8% improvement in read EDP, with negligible change in read access time. This work validates the energy-delay advantages of BSC-based architectures across both density- and performance oriented SRAM designs, providing critical guidance for design technology co-optimization (DTCO) exploration in advanced technology.

12:00-12:20 | Paper ID: 260

Title: A Cross-Layer Inverse Design Framework from Device to Circuit Using Artificial Neural Networks and Pareto Optimization

Author(s): Yuhan Jiang, Haoqing Xu, Kun Luo, Qi Sun, Cheng Zhuo, Huaxiang Yin, Zhiqiang Li, Zhenhua Wu

Presenter: Jiang, Yuhan, Institute of Microelectronics, Chinese Academy of Sciences

Abstract: Advanced nanosheet gate-all-around (GAA) devices introduce highly coupled and high-dimensional design spaces, making conventional TCAD-driven optimization increasingly costly and inefficient. Although data-driven surrogate models and inverse neural networks have been extensively studied at the device level, the consistency between device-level inverse design and circuit-level performance optimization remains insufficiently addressed. To overcome this limitation, this work proposes a cross-layer inverse design framework that integrates TCAD simulations, artificial neural network (ANN) surrogate modeling, Pareto optimization, and SPICE-level circuit evaluation to enable consistent device-circuit co-optimization. An ANN-based forward surrogate model is first trained using large-scale TCAD data to rapidly generate dense I_d - V_g characteristics and automatically extract key electrical metrics, including subthreshold swing (SS), on-state current (I_{on}), and off-state leakage current (I_{off}). Based on these metrics, a device-level closed-loop inverse neural network with hard physical constraints on the output parameter space is constructed and validated through Pareto-front analysis. The trained ANN surrogate is subsequently converted into a Verilog A compact model and embedded into HSPICE for circuit-level performance evaluation. Using circuit-level performance metrics, a compact circuit-level inverse neural network is further developed to infer physically interpretable device parameters from target specifications. By combining output hard constraints with performance-space sampling, the proposed framework enables efficient dataset expansion and systematic Pareto-based analysis for circuit-level inverse design.



TECHNICAL SESSIONS

Technical Session 16: Automated Circuit Design & Optimization

🕒 10:00-12:00, May 10, 2026

📍 Orchid 4302

👤 Session Chair: **Wang, Zhiang**, Fudan University

10:00-10:20 | Paper ID: 26

Title: TopMat: Knowledge Transfer and Hierarchical Optimization for Analog Circuit Sizing via Sub-circuit Topology Matching

Author(s): Yuan Yao, Yuxuan Zhang, Yuanqi Hu, Wei W. Xing

Presenter: **Zhang, Yuxuan**, Beihang University

Abstract: Analog transistor sizing remains a critical challenge in integrated circuit design, often hindered by vast design spaces and inefficient utilization of historical data. To address these issues, we introduce sub-circuit Topology Matching (TopMat) for knowledge transfer and Hierarchical Optimization for analog circuit sizing. Our approach employs a graph-based algorithm to identify structurally similar components between the target circuit and a database of optimized designs. Unlike previous works that transfer all source data with adaptation that incurs potential error and bias, we synthesize an appropriate design space for the target circuit to avoid introducing errors at the data level. Building upon TopMat, we implement a hierarchical optimization strategy that employs a divide-and-conquer approach, significantly reducing the vast design space of large circuits. Moreover, TopMat is a versatile optimization framework tailored for transistor sizing and can be readily integrated with most existing sizing methods. Our transfer learning approach demonstrates a 2.30x improvement in performance and a 2.59x increase in optimization speed compared to other state-of-the-art (SOTA) knowledge transfer methods. When integrated with the SOTA sizing techniques, our method achieves remarkable results, including performance enhancements of 1.60x and speed increases of 2.43x across various analog circuit designs.

10:20-10:40 | Paper ID: 138

Title: MEMO: Dynamic Routing Mixture-of-Experts Reinforcement Learning for Multi-Objective Analog Circuit Sizing

Author(s): Yanan Han, Bijian Lan, Jing Wan

Presenter: **Han, Yanan**, Fudan University

Abstract: Multi-objective analog integrated circuit design in high-dimensional spaces is challenged by strongly coupled and competing performance metrics, where conventional single-policy reinforcement learning often suffers from gradient interference and unstable optimization. This paper proposes MEMO (Mixture-of-Experts for Multi-Objective Optimization), a goal-conditioned reinforcement learning framework that introduces a multi-expert policy with dynamic routing to explicitly decouple heterogeneous optimization behaviors across the performance space. By enabling region-aware expert specialization through a staged soft-sparse gating mechanism, MEMO effectively mitigates gradient conflicts and improves policy expressiveness. To further enhance optimization precision, a multi-stage reward formulation integrating bottleneck-prioritized weighting and nonlinear error shaping is developed to balance global convergence and local refinement. In addition, an adaptive training threshold mechanism dynamically regulates the success criterion to maintain sufficient effective samples throughout training. The proposed routing schedule facilitates a smooth transition from early collaborative exploration to late-stage sparse specialization while preventing



expert collapse via entropy and load-balancing regularization. Experiments on high-dimensional multi-objective optimization of operational amplifiers demonstrate that, compared to single-policy baselines, MEMO reduces normalized error by 46.7% and 63.9% in 2D and 3D target scenarios, respectively, with a more concentrated error distribution, exhibiting significant convergence stability and validating the effectiveness and practical value of the proposed method in high-dimensional multi-objective design spaces.

10:40-11:00 | Paper ID: 174

Title: Portable Expertise for Analog Sizing: A Data-Agnostic Foundation Model Approach

Author(s): Zhuohua Liu, Kaiqi Huang, Qinxin Mei, Yuanqi Hu, Wei W. Xing

Presenter: Liu, Zhuohua, Beihang University

Abstract: The productivity of analog circuit design is severely hampered by the challenge of automating transistor sizing, a task that traditionally relies on non-transferable expert knowledge. Current data-driven approaches are caught between two impractical paradigms: data-hungry models requiring massive, proprietary offline datasets, and transfer learning methods that rely on brittle, monolithic knowledge. This work introduces PortAbLe Expertise for Analog Sizing (ALIGN), a new paradigm that bypasses both bottlenecks by decomposing design knowledge into a data-agnostic, universal reasoning prior and lightweight, task-specific expertise modules. Our framework realizes this by integrating a Prior-Fitted Network (PFN), pre-trained on synthetic function priors to serve as the universal prior, with online-trained Low-Rank Adaptation (LoRA) modules that encapsulate design expertise. Evaluations across multiple analog circuits show that our method achieves state-of-the-art sample efficiency, requiring up to 14.69× fewer simulations than existing methods, all without needing any prior circuit data. Crucially, to the best of our knowledge, we are the first to show in analog sizing that these lightweight LoRA expertise modules can act as portable assets, successfully accelerating design convergence when transferred across both technology nodes and circuit topologies. This approach marks a shift from monolithic models to a modular and reusable knowledge framework, offering a scalable path to accelerate analog design automation.

11:00-11:20 | Paper ID: 241

Title: PARADIGM: Pareto-Optimized Analog Design via Intelligent Database-Guided Methodology

Author(s): Anlan Peng, Chengjie Liu, Yuan Du, Li Du

Presenter: Peng, Anlan, The Chinese University of Hong Kong

Abstract: Traditional manual analog integrated circuit (IC) design is inefficient and often lacks quality assurance, while existing automated approaches remain limited and fragmented. To address these challenges, this paper proposes PARADIGM—a Pareto-optimized methodology for constructing analog IC databases and enabling rapid automated design. By pre-establishing a high-quality circuit database, PARADIGM automates both topology selection and parameter optimization, significantly enhancing design efficiency and quality. Key innovations include a cascaded design space exploration strategy, a multi-class classification model for topology prediction, and a neural network-accelerated parameter optimization algorithm. PARADIGM is validated on seven amplifier topologies using an 180-nm technology node. Results show that it completes design space exploration for a given topology within tens of minutes, achieves up to 95% accuracy in topology prediction, and significantly outperforms traditional methods in both parameter optimization speed and final circuit performance. Code and reproducibility materials are publicly available.

11:20-11:40 | Paper ID: 261

Title: SP-BO: Simulation-Pruning Bayesian Optimization via Conformal Filtering for Efficient Analog Sizing

Author(s): Xuhua Liu, Zhuofu Tao, Yichen Shi, Zhiyuan Wang, Zhongyi Li, Ze Zhang, Hongyang Wang, Shuilong Chen, Ting-Jung Lin, Quan Chen, Lei He

Presenter: Liu, Xuhua, Southern University of Science and Technology



Abstract: Analog circuit sizing is dominated by expensive SPICE simulations. While Bayesian optimization (BO) improves search efficiency, many simulated candidates still violate design specifications, leading to wasted computation. We propose Simulation-Pruned Bayesian Optimization (SP-BO), a specification-aware framework that integrates conformal prediction with BO to filter infeasible candidates before simulation. A surrogate model trained on an offline Sobol-sampled database is equipped with weighted conformal prediction and online recalibration to provide reliable uncertainty bounds under distribution shift. Candidates whose predictive bounds cannot satisfy specifications are safely rejected using approximate probabilistic screening, while safeguard mechanisms ensure robust exploration. Across nine analog circuit topologies, SP-BO reduces wall-clock optimization time by up to 86.6% (59.5% on average) without degrading final design quality. We further demonstrate scalability in a hierarchical 12-bit SAR ADC flow, achieving a 50% runtime reduction while preserving system-level performance.

11:40-12:00 | Paper ID: 165

Title: GRIND: GNN and Reinforcement Learning Enhanced Multilevel Nested Dissection for Circuit Matrix Reordering

Author(s): Jianfei Song, Zhenhua Wu, Zhou Jin, Cheng Zhuo

Presenter: Song, Jianfei, Zhejiang University

Abstract: Sparse matrix reordering is essential in circuit simulation, as it directly influences the efficiency of sparse linear solvers. However, the problem is NP-hard, and existing reordering algorithms typically rely on heuristics, making them susceptible to local optima. To address this issue, we propose GRIND, an AI-driven matrix reordering algorithm based on multilevel nested dissection (ND). By integrating Graph Neural Networks (GNNs) with deep reinforcement learning (DRL), GRIND employs a transfer agent for initial partitioning and a refiner agent for uncoarsening, enabling a global optimization process that transcends the local optima limitations inherent in traditional algorithms, thereby improving adaptability to diverse matrix patterns. Experiments on various matrices show that GRIND achieves higher-quality reorderings and substantially reduces fill-ins, outperforming METIS, NESDIS, RCM, and COLAMD with respective reductions of 17.80%, 9.52%, 91.01%, and 67.54%.



TECHNICAL SESSIONS

Technical Session 17: Prediction & Netlist Analytics



10:00-11:20, May 10, 2026



Orchid 4303



Session Chair: **Ma, Yuzhe**, The Hong Kong University of Science and Technology (Guangzhou)

10:00-10:20 | Paper ID: 108

Title: GSGNN: A novel GSegment-based Graph Neural Network for VLSI Pre-Routing Congestion Prediction

Author(s): Xuan Peng, Yi Zou

Presenter: **Peng, Xuan**, South China University of Technology

Abstract: The congestion problem remains one of the most critical issues the very-large-scale integration (VLSI) industry faces during physical design. Modern placement and routing tools optimize performance based on congestion conditions, and an accurate congestion prediction model can significantly enhance routing quality while reducing time-consuming computational iterations. In this paper, we propose GSGNN, a graph neural network model based on Steiner-tree graph structure for binary classification of global-routing overflow congestion. Our approach incorporates two key innovations: a GSegment-based GNN model and a congestion-aware net-density preprocessing method. Experimental results on real-world designs demonstrate that our model achieves approximately a 23% improvement in F1 score compared with GNet-based models. Moreover, our preprocessing method further improves classification performance. Our dataset and code are available at <https://github.com/pxgh02/GSGNN>.

10:20-10:40 | Paper ID: 203

Title: ParaGate: Parasitic-Driven Domain Adaptation Transfer Learning for Netlist Performance Prediction

Author(s): Bin Sun, Jingyi Zhou, Jianan Mu, Zhiteng Chao, Tianmeng Yang, Ziyue Xu, Jing Ye, Huawei Li

Presenter: **Sun, Bin**, Institute of Computing Technology, Chinese Academy of Sciences

Abstract: In traditional EDA flows, layout-level performance metrics are only obtainable after placement and routing, hindering global optimization at earlier stages. Although some neural-network-based solutions predict layout-level performance directly from netlists, they often face generalization challenges due to the black-box heuristics of commercial placement-and-routing tools, which create disparate data across designs. To this end, we propose ParaGate, a three-step cross-stage prediction framework that infers layout-level timing and power from netlists. First, we propose a two-phase transfer-learning approach to predict parasitic parameters, pre-training on mid-scale circuits and fine-tuning on larger ones to capture extreme conditions. Next, we rely on EDA tools for timing analysis, offloading the long-path numerical reasoning. Finally, ParaGate performs global calibration using subgraph features. Experiments show that ParaGate achieves strong generalization with minimal fine-tuning data: on openE906, its arrival-time R^2 from 0.119 to 0.897. These results demonstrate that ParaGate could provide guidance for global optimization in the synthesis and placement stages.

10:40-11:00 | Paper ID: 249

Title: LSnce: Large-Scale Netlist Constraint Extraction via Graph Similarity Learning

Author(s): Tengjun He, Yuhe Wang, Jialiang Ren, Zhicheng He, Jianguo Hu

Presenter: **He, Tengjun**, Sun Yat-sen University

Abstract: The automation of analog integrated circuit layout generation necessitates accurate and efficient extraction of design constraints such as circuit key structures and symmetry constraints from circuit netlists.



While existing methods based on Graph Neural Networks have improved accuracy, they struggle with scalability and generalization, particularly when it comes to large-scale netlists containing thousands of devices. This paper introduces LSnce, a Large-Scale netlists constraint extraction framework that fundamentally addresses these challenges. LSnce transforms constraint extraction into a graph similarity learning problem. It employs a novel heterogeneous graph embeddings for circuits, incorporating a specialized structure vector to encode the functional role of each device within the topology. A fused-attention heterogeneous GNN then learns rich node embeddings. For extraction, LSnce performs efficient node similarity comparison between the target circuit and constraint templates (or the circuit itself for symmetry), followed by a confidence-filtered mapping inference. This approach also eliminates the need for retraining when new constraint types are introduced. Extensive experiments on multiple datasets demonstrate that LSnce significantly outperforms methods in both accuracy and scalability. Notably, it maintains robust performance on large-scale netlists where conventional methods degrade, achieving high F1-scores in both circuit key structure extraction and symmetric constraint detection. In a word, our framework provides a solution for advancing analog Electronic Design Automation.

11:00-11:20 | Paper ID: 273

Title: AccUnit: Accelerating Unit Level Verification for RISC-V Processors Using FPGA

Author(s): Chenang Zhu, Weidong Li, Weijie Weng, Yungang Bao, Kan Shi

Presenter: Jiang, Zijian, Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences

Abstract: Functional verification can account for up to 70% of the chip development cycle and has become a major bottleneck in the development of complex chips, especially for emerging open-source RISC-V processors. Existing agile verification research mainly focuses on system-level verification, while defect repair costs increase exponentially as the design progresses. Unit-level verification is seen as a key approach for shifting verification left. To improve the efficiency and completeness of unit-level verification, we propose AccUnit, a hybrid verification method for unit-level chip verification that combines FPGA hardware acceleration with formal verification. This approach builds a parallel verification framework with FPGA as the high-speed execution core, introduces a synthesizable coverage collection mechanism to quantify the verification process, and uses formal verification to cover the coverage gaps that traditional methods fail to address. This forms an end-to-end, coverage-driven random-formal hybrid verification loop. Experimental results show that this method achieves up to 75× speedup in unit-level chip functional verification and increases coverage in complex control logic modules by more than 20%.



TECHNICAL SESSIONS

Technical Session 18: Test, Emulation, Layout & Lifecycle Tools

10:00-12:20, May 10, 2026

Orchid 4304

Session Chair: **Zhang, Hongce**, The Hong Kong University of Science and Technology (Guangzhou)

10:00-10:20 | Paper ID: 41 | Invited Talk

Yang, Wu

Siemens EDA



Bio.: Wu Yang is the Director of Technical Programs at Siemens EDA. Bringing 27 years of extensive experience in design-for-test (DFT), 2D and 3D IC testing, silicon learning, failure analysis, and reliability, Mr. Yang is a recognized expert in the field. He is a frequent speaker at industry conferences and a regular contributor to technical papers, articles, and public discussions. His current responsibilities have expanded to address new industry challenges in silicon lifecycle management and AI/ML related applications. Mr. Yang has delivered multiple tutorials at prestigious events including ITC, ITC_Asia, ATS, ETS, and VTS. He holds an M.S. degree in Electrical Engineering from Portland State University.

Title: Industrial Trends for Silicon Lifecycle Management

Abstract: The semiconductor industry faces continuous challenges with technology scaling and increasing design complexity. This necessitates a more integrated and intelligent Silicon Lifecycle Management (SLM) strategy, encompassing design-for-test (DFT), testing, time-to-market acceleration, in-field operation and monitoring, and post-mortem analysis.

The design phase is undergoing a transformation, driven by the demand for near-zero defects and significant test cost reduction. This requires defect-oriented fault modeling, embedded IPs for in-life monitoring, and efficient DFT architecture like the Streaming Scan Network (SSN). As silicon transitions into its learning phase—including bring-up, production yield ramp-up, and customer return analysis—advanced scan diagnosis and diagnosis-driven yield analysis are crucial for accelerating defect resolution and physical failure analysis. Furthermore, in-field SLM focuses on enhancing safety and reliability, particularly for mission-critical applications. Robust in-system deterministic tests, combined with continuous and different types of monitoring, reinforce reliability and provide crucial feedback to the design process. This comprehensive approach ensures long-term product viability.

10:20-10:40 | Invited Talk

Huang, Zheng

Semitronix Corporation



Bio.: Mr. Zheng Huang holds a Master's degree in Microelectronics from Xidian University and brings over 20 years of experience in full-custom IC design, with a focus on the DFT (Design for Test) area. He began his career at Infineon in 2004 as a DFT engineer, then served as an SoC architect engineer at Intel Communications from 2011. He is currently the Product Director at the EDA department of Semitronix Corporation. His expertise lies in EDA flow solutions and DFT technology development, and the EDA flow and products he has contributed to have been widely adopted across various chip design domains.



Title: Addressing DFT Challenges in Advanced Packaging: Implementation and Application of EDA Solutions

Abstract: Advanced packaging technologies, such as 2.5D/3D integration, chiplet-based designs, and heterogeneous integration, introduce unprecedented complexities in terms of thermal management, signal integrity, power delivery, mechanical stress, and design-for-Test (DFT). Traditional EDA tools and design flows often fall short in addressing these multi-physics and multi-domain challenges within a unified environment. At Semitronix, we've already developed dft solutions and will demonstract the practical implementation and application of dedicated dft EDA solutions in real-world projects to tackle these hurdles.

10:40-11:00 | Paper ID: 33

Title: Reducing SQED Duplicate Mode via Processor-Level Abstraction

Author(s): Yufeng Li, Linze Li, Qiusong Yang, Yiwei Ci, Enyuan Tian, Xuying Zhao, Yungang Bao, Kan Shi

Presenter: Li, Yufeng, Institute of Computing Technology, Chinese Academy of Sciences

Abstract: Symbolic Quick Error Detection (SQED) streamlines processor verification by checking a microarchitecture-agnostic self-consistency property using bounded model checking (BMC). While effective in detecting bugs without manual property specification, SQED suffers from severe scalability limitations due to state explosion in complex designs. This paper introduces RDM-SQED to mitigate this bottleneck by reducing the resource-intensive duplicate mode with a lightweight Processor-Level Abstraction (PLA). The PLA captures software-visible behaviors through a concise set of Elementary Instructions (EIs). To further constrain the verification logic, we propose a recursive refinement algorithm that generates a minimal EI set. Experimental evaluation on an out-of-order RISC-V processor demonstrates that RDM-SQED significantly outperforms existing variants in both scalability and bug detection efficiency, successfully identifying bugs that cause timeouts in other methods.

11:00-11:20 | Paper ID: 65

Title: AgentFA: The first Agent-assisted Failure Analysis framework for DRAM Test Efficiency Improvement

Author(s): Xiaojing Lu, Daoda An, Shuyang Hu, Yutong Wu, Jingyi Chang, Kangyi Chen, Yundong Liang, Liujia Du, Chong Xu, Hao Wang, Yanfeng Li, Zhigang Ji

Presenter: Lu, Xiaojing, ChangXin Memory Technologies, Inc

Abstract: As DRAM scales to higher density with tighter design margins, rapid yield ramp-up increasingly depends on efficient failure analysis (FA). Core-array FA remains a major bottleneck because it involves analog/mixed-signal behaviors and spatially correlated fail bit maps (FBMs), and it often relies on expert experience. This paper proposes an agent-assisted FA framework to speed up hypothesis-driven diagnosis. We first develop DATIS-FBM, a fault-injectable behavioral DRAM simulator that connects circuit-level fault characterization to bank-level FBM generation under test-pattern-driven execution. We then introduce an LLM-driven workflow that coordinates tool execution, supports fault injection, and guides the analysis in a human-in-the-loop manner. A knowledge-graph (KG) module is further integrated to reuse historical cases and enable continuous knowledge accumulation. Experimental results show that the proposed framework can reproduce real failure cases and reduce FA turnaround time.

11:20-11:40 | Paper ID: 120

Title: Automated SVA Generation with LLMs

Author(s): Lik Tung Fu, Qihang Wang, Shaokai Ren, Mengli Zhang, Sichao Yang, Jun Liu, Xi Wang

Presenter: Fu, Lik Tung, National ASIC Center, Southeast University

Abstract: Functional verification remains a dominant cost in modern IC development, and SystemVerilog Assertions (SVAs) are critical for simulation-based monitoring and formal property checking. However, writing



SVAs by hand is time-consuming and error-prone. Directly prompting general-purpose large language models (LLMs) is also unreliable: the generated properties are often syntactically invalid or semantically incorrect, and the problem is exacerbated by scarce, high-quality domain training data. We present SVA Generator, a data-centric framework that translates natural-language SVA Descriptions (SVADs) into executable SVAs. It uses AST-grounded constraint injection and an automated supervision pipeline that enforces structural consistency and reduces hallucinations via de-duplication and constraint checks. To enable rigorous evaluation, we introduce a benchmark suite stratified by AST depth and use formal property equivalence checking to quantify semantic correctness separately from syntax validity, by checking mutual implication between the generated and reference properties under the same clocking and environment assumptions. Across all difficulty tiers, SVA Generator achieves comparable Syntax Pass Rate (SPR) to strong general LLM baselines, while delivering substantially higher Semantic Equivalence Rate (SER) on deeper tiers: +24.5 pp on D2, +26.0 pp on D3, and +17.5 pp on D4 relative to the best-performing general LLM, corresponding to a +22.7 pp SER improvement on average over D2--D4. These results highlight that high-fidelity data construction and depth-stratified benchmarking are key to reliable, semantics-preserving SVA generation.

11:40-12:00 | Paper ID: 166

Title: MFchip: A Multi-FPGA Emulation Platform for Chiplet-Based Systems with Latency and Bandwidth Adaptation

Author(s): Jinwei Zhou, Xiping Dong, Yi Ren, Chenhao Xue, Jiaying Zhang, Yihan Yin, Guangyu Sun, Xinnan Lin

Presenter: Ren, Yi, Peking University

Abstract: Multi-FPGA emulation is a critical technique for system-level Very Large Scale Integration (VLSI) verification. Existing multi-FPGA emulation methods are inefficient as their generic approach fails to leverage design-specific characteristics. In this paper, we propose MFchip, a multi-FPGA emulation platform for chiplet-based architectures. It aims to map each complete chiplet onto a single FPGA, minimizing inter-FPGA communication and significantly accelerating emulation speed. Co-designed MFlink provides an abstraction layer that dynamically adapts to inter-chiplet latency and bandwidth, ensuring emulation accuracy. On a U200-based platform, MFchip achieves 50 MHz, validating its practicality for complex chiplet designs.

12:00-12:20 | Paper ID: 259

Title: Efficient and Scalable Parallel Net Tracing for Large-Scale Industrial Layouts via Hybrid Spatial Indexing

Author(s): Kaifeng Yang, Minghua Shen

Presenter: Yang, Kaifeng, Sun Yat-sen University

Abstract: Net tracing is a key process in electronic design automation (EDA), and is essential for layout-versus-schematic (LVS) verification, finding defects, and analyzing yield. However, as circuit layouts become larger and more complex, traditional tree or grid methods face significant challenges, such as checking too many overlapping boxes and struggling with objects that span multiple areas. This paper proposes an efficient and scalable net tracing framework designed for large-scale industrial layouts. The core of our framework is a two-stage polygon intersection method using a hybrid grid-based R-tree index, which combines the local grouping of grids with the fast searching of R-trees to speed up geometric queries. Based on this index, we build an exact connectivity tracing algorithm using breadth-first search (BFS). To make effective use of modern multi-core processors, we use a work-stealing parallel scheduling method to balance the workload and handle the uneven layout connections. Tests on industrial datasets with up to 30 million polygons show that our method is about 10% faster than traditional R-tree methods and scales very well, achieving a 5x to 8x speedup on multi-core systems. Our framework provides exact and reliable connection results, making it highly practical for real industrial use.



TECHNICAL SESSIONS

Technical Session 19: Power, IR Drop & Waveform Prediction (ML-Accelerated)



10:00-12:00, May 10, 2026



Orchid 4201



Session Chair: **Lin, Yibo**, Peking University

10:00-10:20 | Paper ID: 27

Title: PowerCube: A Formula-in-the-Loop Framework for Cross-Stage Power Map Prediction

Author(s): Yuxiang Zhao, Jing Mai, Zishu Li, Haoyi Zhang, Jincheng Lou, Qing He, Runsheng Wang, Yibo Lin

Presenter: **Zhao, Yuxiang**, Peking University

Abstract: Accurate early-stage power estimation is critical in modern IC design, yet existing cross-stage predictors yield only a single total power value, lacking granular 2D maps for localized optimization. Thus, we propose PowerCube, the first cross-stage prediction framework dedicated to fine-grained 2D power maps. To capture complex post-routing variations while preserving physical laws, our novel "formula-in-the-loop" architecture decouples transient physical parameter inference from deterministic power computation. First, a Spatial-Signal Dual-Branch Network calibrates early layout features into precise load and slew anchors. These anchors drive a differentiable calculation engine (Diff-LUT) to evaluate internal power, combined with exact analytical equations to compute switching and leakage components. Subsequently, a Synergistic Module bridges these localized metrics with layout-scale contexts to synthesize high-fidelity maps. Evaluated on ASAP7 7nm across 8 designs, PowerCube eliminates Out-of-Memory bottlenecks on massive layouts (e.g., 1.5-million-cell nvdla-large and achieves an average 414x runtime speedup over commercial flows. Furthermore, it outperforms baselines by improving average R^2 by 26.44% and reducing NMAE by 28.72%, delivering a fast, highly scalable, and sign-off-accurate solution.

10:20-10:40 | Paper ID: 36

Title: Aging Aware Adaptive Voltage Scaling for Reliable and Efficient AI Accelerators

Author(s): Tong Xie, Zuodong Zhang, Chao Yang, Yuan Wang, Runsheng Wang, Meng Li

Presenter: **Xie, Tong**, Peking University

Abstract: Deep neural networks (DNNs) have showcased remarkable performance across various tasks and are widely deployed on AI accelerators fabricated in advanced technology nodes for efficiency. As aging effects become more pronounced, timing and voltage guardbands are increasingly applied. Aging-aware adaptive voltage scaling (AVS), which adjusts supply voltage based on on-chip aging scenarios, has emerged as a promising solution to avoid excessive guardbanding. However, conventional AVS techniques overlook the inherent resilience of DNNs and frequently raise the supply voltage unnecessarily, thereby exacerbating aging and increasing power consumption. To enable reliable and efficient AI inference with AVS, in this paper, we develop an accurate aging prediction framework that incorporates historical effects and iterative extrapolation for full-lifetime modeling. Building on this framework, we propose a fault-tolerant voltage scaling policy that exploits DNN resilience and defers voltage increases accordingly. Experiments show that our framework mitigates the pessimism of maximum-voltage baselines, reducing predicted threshold voltage shift (ΔV_{th}) by 19.4% for PMOS and 19.1% for NMOS, respectively. Furthermore, evaluation on representative DNN workloads demonstrates that our optimization reduces aging degradation by up to 45.8% (NMOS) and 30.6% (PMOS) while achieving 14.0% average lifetime power savings compared to resilience-agnostic methods.



10:40-11:00 | Paper ID: 170

Title: Highly Accurate Power Estimation in FPGA HLS via Adaptive and Enhanced Subgraph Learning

Author(s): Yuchen Liu, Xiangfei Hu, Yuyang Ye, Tinghuan Chen, Zhe Lin, Hao Yan

Presenter: **Liu, Yuchen**, Southeast University

Abstract: Power estimation is fundamental to hardware design and optimization. At the high-level synthesis (HLS) stage, accurate prediction of dynamic power remains challenging. Traditional methods model the entire design as a graph, and leverage GNN for fast power prediction. However, their neglect of critical arithmetic nodes and relevant local structures constrains prediction accuracy and scalability. To handle this, we propose a dynamic power prediction framework based on adaptive and enhanced subgraph learning. It integrates power-critical subgraph mining with graph learning to capture the impact of local power features on global dynamic power. Specifically, RL agents adaptively extract subgraphs around critical arithmetic nodes with appropriate depths, enabling the subgraph GNN to learn local power features. This GNN cooperates with a whole-graph GNN that captures graph-level topology to predict the global dynamic power. According to comprehensive experiments, our framework predicts the dynamic power of designs with an average error of 7.34% and delivers an $8.13\times$ speedup over Vivado, outperforming the state-of-the-art PowerGear in accuracy and efficiency.

11:00-11:20 | Paper ID: 218

Title: MiniViT: A Lightweight and Physics-Aware Vision Transformer for Efficient IR Drop Prediction in Multi-Layer PCBs

Author(s): Rui Chen, Jienan Chen

Presenter: **Chen, Jienan**, University of Electronic Science and Technology of China

Abstract: Rapid and accurate IR-drop prediction is essential for power integrity analysis in modern IC design. Traditional numerical solvers suffer from super-linear computational complexity, while data-driven Convolutional Neural Networks (CNNs) struggle to capture global electrical coupling due to their limited receptive fields. To address these challenges, we propose MiniViT, a lightweight and physics-constrained Vision Transformer. Unlike standard Transformers, MiniViT couples PDN-aware tokenization (directional resistances and source/sink indicators) with a streamlined convolutional embedding and efficient windowed self-attention with global register tokens, matching the resistive mesh topology while keeping compute tractable. Furthermore, we introduce a Layer-wise Decomposition and Bias Correction (LDBC) strategy, enabling the efficient prediction of complex multi-layer PCB designs using 2D surrogates. Experimental results on 10,000 industrial designs demonstrate that MiniViT reduces the Average Maximum Relative Error (AMRE) by over 71% compared to U-Net baselines while requiring only 0.85M parameters ($\sim 21\times$ fewer than PDNNet). Crucially, prediction errors across multi-layer configurations remain strictly within the 25% industrial design margin.

11:20-11:40 | Paper ID: 243

Title: SpecWave: Spectral-Enhanced Graph Learning for Crosstalk-Aware Waveform Prediction

Author(s): Yanlong Mao, Jiajie Xu, Bowen Shi, Yaning Jia, Hao Yan

Presenter: **Mao, Yanlong**, Southeast University

Abstract: With technology nodes continue to scale, crosstalk effect has become indispensable to accurate static timing analysis (STA). Existing machine learning-based approaches struggle to achieve high-fidelity waveform prediction under crosstalk due to the intricate time-domain characteristics of signal waveforms. To address these challenges, we propose SpecWave, a spectral-enhanced graph learning framework for crosstalk-aware waveform prediction. SpecWave first extracts spectral representations of crosstalk waveforms in the frequency domain via the Fast Fourier Transform. Hierarchical graph learning is then employed to



capture the input-specific transfer function. Subsequently, high-precision output waveforms are reconstructed via a frequency-domain fusion mechanism and a differentiable inverse transform. Experimental results on open-source designs demonstrate that SpecWave achieves an average waveform error of 3.04%, while delivering a 40.94× speedup over HSPICE and a 3.40× acceleration compared to commercial STA tool.

11:40-12:00 | Paper ID: 267

Title: HeteroPower: A CPU-GPU Heterogeneous Engine to Accelerate Gate-Level Power Analysis

Author(s): Xizhe Shi, Zizheng Guo, Yun Liang, Runsheng Wang, Yibo Lin

Presenter: Shi, Xizhe, Peking University

Abstract: Gate-level power analysis is critical for modern SoC signoff yet imposes a severe computational burden, especially in the vectorless mode where switching activities must be propagated across the entire design. We present HeteroPower, a CPU-GPU heterogeneous gate-level power analysis engine built upon a prior GPU-accelerated static timing analysis (STA) framework. HeteroPower reuses GPU-resident timing data, including slews, capacitances, and the levelized timing graph, to seamlessly integrate power analysis with STA, minimizing redundant data transfers. For vectorless analysis, we exploit the structural isomorphism between activity propagation and STA forward traversal, deploying a lightweight bytecode-driven probabilistic engine that maps naturally to the GPU's SIMT model. Four dedicated GPU kernels then compute switching, internal, and leakage power, featuring fused per-cell evaluation and a postfix bytecode VM for state-dependent leakage. Experiments on industrial benchmarks demonstrate that HeteroPower achieves an average speedup of 6.36x over PrimeTime PX and 10.97x over OpenSTA in the annotated mode, scaling to 21.80x over PrimeTime PX and 5.44x over OpenSTA in the vectorless mode, and maintains an average gate-level total-power mean absolute relative error (MARE) of 0.001% (annotated) and 6.76% (vectorless) against PrimeTime PX.



TECHNICAL SESSIONS

Technical Session 20: Layout, Pattern Matching & Front-End Cell/Mask Flow

10:00-11:40, May 10, 2026

Orchid 4202

Session Chair: **Zhang, Yuhang**, East China Normal University

10:00-10:20 | Paper ID: 73

Title: Layout Pattern Matching Using Edge-Vector Sparse Sequences and Dual-Anchor Probing

Author(s): Shuo Liu, Chunrun Su, Jiaju Li, Chao Wang

Presenter: **Liu, Shuo**, Southeast University

Abstract: As layout designs continue to scale in size and geometric complexity, efficient and accurate layout pattern matching becomes increasingly challenging. This paper presents a fast and exact layout template matching method based on sparse edge vector sequences and a dual-anchor probing mechanism. By encoding polygon boundaries into sparse edge vector sequences, the proposed method enables jump-based pruning to rapidly eliminate infeasible candidate regions. A cache-efficient linearized grid index is further employed to accelerate local geometric queries, while dual-anchor probing ensures accurate geometric verification. These components are integrated into a pruning-first matching framework that significantly reduces the candidate search space without sacrificing correctness. In standard benchmark evaluations, the proposed method achieves 100% matching accuracy while delivering approximately $7.4\times$ – $20.5\times$ runtime speedup over state-of-the-art academic approaches; meanwhile, it attains up to $193.9\times$ speedup compared with the industrial standard tool Calibre, demonstrating significant advantages in performance and scalability.

10:20-10:40 | Paper ID: 127

Title: CAPCell: Standard Cell Layout Synthesis with Parasitic Capacitance Aware Parallel Sampling

Author(s): Kairong Guo, Jiechen Huang, Wenjian Yu, Yibo Lin

Presenter: **Guo, Kairong**, Peking University

Abstract: As technology nodes scale into the angstrom era, the correlation between geometric heuristics (e.g., total wirelength) and actual circuit performance has significantly degraded. Consequently, traditional standard cell synthesis frameworks, which rely heavily on geometric proxies, fail to capture intricate parasitic effects, resulting in sub-optimal Power, Performance, and Area (PPA). To bridge this gap, we propose CAPCell, a neuro-symbolic framework that directly targets parasitic capacitance optimization during layout synthesis. CAPCell integrates a pre-trained CNN as a high-fidelity surrogate for expensive field solvers and employs a hybrid solver-guided sampling strategy to navigate the routing search space. This approach allows for the rigorous optimization of electrical metrics while guaranteeing strict design rule compliance (DRC-clean). Validated on the ASAP7 7nm technology node, CAPCell demonstrates the capability to break the limitations of geometric heuristics. Experimental results show that our method achieves an average delay reduction of 0.36 ps (0.2%) per cell under strict iso-area conditions. Crucially, these cell-level improvements translate to the system level, yielding 2% improvement in maximum operating frequency and 7% reduction in TNS for block-level designs.

10:40-11:00 | Paper ID: 162

Title: DTCO Dual-Sided Standard Cells with DP-Based Optimization and Custom Library Development

Author(s): Qinzhe Zhi, Peijie Ma, Heng Wu, Tianyu Jia

Presenter: **Jia, Tianyu**, Peking University



Abstract: Leveraging dual-sided standard cells is the emerging development trend for advanced technology nodes. By utilizing backside metal and pins, superior PPA can be obtained for each standard cell. However, the diverse cell design configurations also complicate the timing optimization and library development. In this work, we first introduce the design challenges for dual-sided standard cells and then present corresponding DTCO methodology. First, a dynamic programming (DP) based timing optimization is developed to systematically optimize critical paths with dual-sided cells. The cell utilization further drives custom library development to shrink dual-sided cell library size. Compared with FS-only and tool-based dual-sided baseline, our DP-based timing optimization achieves 10.3% effective efficiency improvement, 37.3% WNS, 8.7% area and 67% nTSV count reduction. With our custom library development, the dual-sided standard cell library size is reduced by 1.89 $\times$, shorten the design cycle by 1.46 $\times$ with no performance degradation.

11:00-11:20 | Paper ID: 175

Title: Lithography Test Layout Generation Using DRC-Constrained Large Language Models

Author(s): Jinheng Sun, Tianshuai Qi, Haoxin Zuo, Yingying Shang, Qingchen Cao, Zhenjie Yao, Jiangliu Shi

Presenter: Sun, Jinheng, University of Chinese Academy of Sciences

Abstract: In very-large-scale integration (VLSI) lithography manufacturing, the generation of layout patterns is crucial for supporting key applications such as optical proximity correction (OPC), source-mask optimization (SMO), and hotspot detection. However, in the logic-to-physical design flow, the complexity of timing closure and lengthy iterative cycles result in high costs in constructing large-scale layout test pattern libraries. This study proposes a novel method for layout test pattern generation based on DRC-constrained large language model (DRCLLM). We first encode the patterns into structured token sequences to facilitate model processing. At the same time, we design a loss function that incorporates specific design rule constraints, which guide the generation of rule-compliant layouts. Experimental results on 4 different LLM backbones show that the average legality rate increased by up to 6.6%, while the diversity keeps stable. Notably, the improvement consistently appears across all evaluated LLM backbones, indicating that the DRC constraint is model-agnostic. The results suggest that DRCLLM is capable of generating layouts with high legality without compromising diversity.

11:20-11:40 | Paper ID: 226

Title: Area-Efficient Synthesis of ESD Local Clamps via Floorplan-Aware Optimization and Conduction-Duration Constrained Sizing

Author(s): Baoyang Zhang, Guangyi Lu, Zelong Huang, Haoyu Xia, Qi Wu, Haiming Wang

Presenter: Zhang, Baoyang, Southeast University

Abstract: With the continued scaling of modern integrated circuit technologies, internal gate oxides (GOX) exhibit heightened vulnerability to Charged Device Model (CDM) electrostatic discharge (ESD) events. While main clamps manage high-energy pulses, they often fail to adequately protect deep internal nodes due to severe parasitic bus impedances. Local clamps provide a vital secondary protection path, yet their manual design is highly iterative and heavily constrained by strict area limits. Unlike main clamps, local clamps do not require prolonged conduction durations; instead, given the severe area constraints of internal cores, they demand only sufficient conduction durations (T_{on}) to suppress fast overshoots. By embedding the (T_{on}) as a strict penalty within a genetic algorithm and combining analytical area estimation, the optimized local clamps strictly confine the (T_{on}) error to within 5%. Furthermore, full discharge-path CDM transient simulations demonstrate that the local clamp optimized for minimum on-resistance decisively suppresses the destructive peak voltage at the protected internal node by 49.5% (from 6.64~V down to 3.35~V), while occupying merely 680.26 μm^2 . This methodology significantly accelerates the design cycle while ensuring highly reliable, area-efficient internal ESD protection.



TECHNICAL SESSIONS

Technical Session 21: Architecture, Placement, Routing & Security

10:00-12:00, May 10, 2026

Orchid 4203

Session Chair: **Yao, Hailong**, University of Science and Technology Beijing

10:20-10:40 | Invited Talk



Zhao, Yi

Zhuhai Silicon Chip Technology Ltd.

Bio.: Holding a Ph.D. from the University of Southampton, UK, under the supervision of Royal Society Fellow Professor Bashir Hashimi, he embarked on 2.5D/3D stacked IC design research in 2008 as part of one of the world's earliest pioneering research teams exploring advanced chip architecture methodologies, having collaborated with IMEC for 3D-IC technology validation. With 15 years of dedicated R&D in 3D integrated circuit design, he has published several excellent papers and awarded the VLSI-SOC Best Paper.

Currently serving as Founder and Chief Scientist at Zhuhai Silicon Chip Technology Ltd. , he has been a majority of leading national research initiatives. He directs the team in developing self-proprietary 2.5D/3D Stacked IC EDA tools, enabling critical advancements in the semiconductor industry through backend full-flow EDA tools and solutions.

Title: 2.5D/3D Advanced Packaging EDA+STCO New Paradigm: architecting & physical implementation, simulation, and testing

Abstract: This report aligns with the cutting-edge trends in system-level modeling and design methodologies, addresses the core design challenges of heterogeneous microsystem integration, and focuses on the pain points and demands of cross-device, cross-process, and cross-hierarchy collaborative design in heterogeneous integration from a system-level modeling perspective. It shares how SiChip Technology, with its independently developed STCO (System Technology Co-Optimization) framework, has redefined advanced packaging design flows and established a new 2.5D/3D EDA+ paradigm.

Based on the 3Sheng Integration Platform for stacked chip design, SiChip has built a full-flow tool chain covering five core modules: 3Sheng Zenith (Architecture Design) – 3Sheng Ranger (Physical Implementation) – 3Sheng Ocean (Multi-die DFT) – 3Sheng Volcano (Co-Simulation) – 3Sheng Stratify (Multi-die Physical Verification) (LVS and DRC). The platform also implements a dual-system approach: “chiplet-interposer-packaging co-design” and “performance-cost-testability co-optimization”, enabling efficient full-flow collaboration and global optimization for 2.5D/3D advanced packaging from system planning to mass production.

Combined with industrial cases, the report demonstrates the technological breakthroughs and application value of the STCO-based EDA+ paradigm in advanced packaging heterogeneous integration, and explores how independent 2.5D/3D EDA technologies empower the collaborative development of the advanced packaging industry chain.



10:20-10:40 | Paper ID: 79

Title: Fast, Feasible Chiplet Placements via Warm-Start Multi-Objective Reinforcement Learning

Author(s): Sezin Kircali Ata, Yubo Hou, Mihai Dragos Rotaru, Rahul Dutta, Ashish James

Presenter: **Ata Kircali, Sezin**, A*STAR

Abstract: Heterogeneous System-in-Package (SiP) designs have intensified the need for AI-driven chiplet placement, yet many existing approaches depend on random or biased heuristics, which often yield suboptimal initial layouts and slow optimization convergence. To address sample inefficiency, limited exploration diversity, and slow reinforcement-learning (RL) optimization convergence, we propose a warm-start multi-objective RL framework (FFX-MORL) that integrates fast and feasible space-filling design-space exploration with actor-critic policy optimization. The method uses diverse initial placements and penalty mechanisms to encourage effective exploration while avoiding repetitive patterns. A key result is that well explored warm-starts, followed by policy refinement during training, substantially improve placement efficiency and deliver high quality multi-objective trade-offs. Experiments on advanced packaging systems show that FFX-MORL achieves strong temperature-wirelength trade-offs and up to 3.2x faster convergence compared with state-of-the-art methods. We further analyze design-dependent initialization behavior, which motivates the selection of sampling strategies aligned with chiplet heterogeneity.

10:40-11:00 | Paper ID: 263

Title: RouteFlow: A Routing-Efficient CGRA Framework Through Accelerated Layout-Aware Mapping

Author(s): Hongyu Pang, Jiajun Qin, Song Chen, Qi Xu

Presenter: **Pang, Hongyu**, University of Science and Technology of China

Abstract: Coarse-Grained Reconfigurable Arrays (CGRAs) provide a compelling hardware architecture that balances computational flexibility with energy efficiency. The effectiveness of application mapping onto CGRA fabrics is fundamentally governed by their interconnect design. Among prevailing architectures, the neighbor-to-neighbor (N2N) scheme stands as a fundamental topology that delivers superior area and power characteristics. However, N2N-only CGRAs suffer from prolonged compilation durations because of the constrained routing resources when compared to more richly interconnected architectures. To overcome the mapping limitations of N2N-based CGRAs, we introduce a layout-aware mapping algorithm that utilizes spatial information in dataflow graphs (DFGs) to improve schedulability. We further develop RouteFlow, a parameterized framework for efficient implementation and design space exploration. Experimental results demonstrate that our approach enables practical compilation times and achieves 1.24x lower latency with a 26.2x mapping speedup compared to island-style architectures while retaining area efficiency.

11:00-11:20 | Paper ID: 136

Title: GNN-Accelerated Magnitude Vector Fitting Method for Rapid Pole Identification and Transient Simulation of 2.5D Power Delivery Networks

Author(s): Jiacheng Shen, Yuanyuan Duan, Guangxi Fan, Lyu Lyu, Cheng Wang, Xiaolei Zhu, Leilai Shao

Presenter: **Shen, Jiacheng**, Shanghai Jiao Tong University

Abstract: With the rapid advancement of heterogeneous integration, 2.5D architectures introduce new challenges for power delivery network (PDN) design. Efficient modeling and fast transient analysis of PDN impedance are essential to ensure the signal and power integrity of high-performance electronic systems. This paper presents a novel GNN-accelerated Magnitude Vector Fitting (MVF) framework for the efficient modeling and passivity-constrained analysis of PDN transfer functions. The framework integrates a shared Graph Neural Network (GNN) backbone for topological and electrical feature extraction, a transformer-based pole predictor optimized via bipartite matching loss, and an MLP-based impedance magnitude predictor. Not only does it support frequency-domain impedance prediction, but it can also be applied to perform rapid time-domain



simulations, significantly enhancing the efficiency of power integrity analysis. Through comprehensive evaluations with diverse PDNs, in the frequency-domain, the proposed GNN-accelerated MVF method accurately predicts the impedance responses with an average Mean Absolute Percentage Error (MAPE) smaller than 1% and a passivity pass rate of 98.91% through verification of the Hamiltonian matrix; in the time-domain, our method achieves a speedup over 16 times while maintaining a high R2 bigger than 0.95 compared to NGSPICE simulations.

11:20-11:40 | Paper ID: 195

Title: Reuse, Reduce, Compute: A Similarity-Aware Accelerator for Diffusion Model Inference

Author(s): Junhong Qian, Yangbo Wei, Shaoqiang Lu, Dongge Qin, Qifan Wang, Xiao Shi, Chen Wu, Lei He, Longxing Shi

Presenter: Qian, Junhong, Southeast University

Abstract: Diffusion Transformers (DiTs) deliver state-of-the-art generative performance but suffer from high inference costs due to substantial spatiotemporal redundancy across iterative denoising steps. We observe that DiT features become increasingly predictable over time and that many layers exhibit reduced computational necessity, motivating adaptive computation. Leveraging these insights, we introduce DiT-OPU, a hardware software co-designed FPGA accelerator that integrates algorithmic sparsity with architectural specialization. At the algorithmic level, our Similarity-aware Multi-Mode Dynamic Inference (SMDI) framework employs a lightweight predictor, DELTASEER, to guide dynamic switching among prediction reuse (reuse), partial computation (reduce), and full computation (compute). At the hardware level, an asymmetric multi-core architecture enables tightly pipelined execution that hides prediction overhead and maximizes utilization. Evaluated on leading DiT models for text-to-image/video generation, DiT-OPU achieves $1.3\times$ – $9.2\times$ end-to-end speedup and $4.8\times$ – $35.9\times$ energy efficiency improvements over GPU baselines, while fully preserving generation fidelity.

11:40-12:00 | Paper ID: 134

Title: Anti-AdvTamp: Exposing Adversarial Evasive Tampering Attacks in Network-on-Chips with a Multi-Scale Attribute Fusion Detection

Author(s): Shengkai Hu, Boojoong Kang, Basel Halak

Presenter: Hu, Shengkai, University of Southampton

Abstract: Network-on-chip (NoC) architectures are increasingly becoming a widely adopted system-on-chip (SoC) architecture due to their scalable and modular communication structure, which matches the requirements of modern multiprocessor system-on-chip (MPSoC) and deep learning (DL) accelerator architectures, yet they are vulnerable to routing tampering attacks (e.g., traffic diversion) that induce Denial-of-Service (DoS). Therefore, machine learning (ML) have been increasingly applied to tampering detection in NoC; however, existing ML-based detectors struggle to cope with more covert adversarial tampering attacks. To address this challenge, this paper proposes SEAA, a symmetry-exploiting framework to construct two adversarial tampering attack models (AdvTamp1 and AdvTamp2), and evaluates their impact on ML-based detection. Furthermore, this research proposes a novel ML detection framework based on multi-scale attribute fusion, which integrates coarse-grained and fine-grained NoC performance metrics to improve detection accuracy. Experiments under PARSEC and synthetic Traffic Pattern benchmarks on various NoC topologies show that traditional ML models achieve detection accuracies of only 57.4% and 49.4% under AdvTamp1 and AdvTamp2 attacks, respectively. In contrast, our proposed multi-scale attribute fusion method significantly improves detection performance, achieving accuracies of 96.19% and 98% under AdvTamp1 and AdvTamp2, respectively. These results demonstrate the effectiveness of the proposed method in improving detection under adversarial tampering and highlight its potential for enhancing the security of NoC.



POSTER SESSION



15:30-16:00, May 8



Peony Ballroom

Poster No.: P01

Paper ID: 21

Paper Title: MGS: A Multi-Granularity Scheduling Framework for DNN Inference on Capacity-Constrained ReRAM-Based CIM

Author(s): Qinghang Zhao, Tao Li, Zhuohui Xu, Jinjian Wu

Presenter: Xu, Zhuohui, Xidian University

Poster No.: P02

Paper ID: 23

Paper Title: Boundary-Only Layout Encoding for Lithography Hotspot Detection

Author(s): Yufeng Zhang, Zhili Xu, Yuqian Zang, Jiale Li, Silin Chen, Zhaoyang Wang, Chihwei Chiang, Ningmu Zou

Presenter: Zhang, Yufeng, Nanjing University

Poster No.: P03

Paper ID: 28

Paper Title: High-Performance Configurable Floating/Fixed-Point Processing Element for Multiple-Precision Deep Learning

Author(s): Zichen Yan, Chenghua Wang, Fei Lyu, Yijun Cui, Weiqiang Liu

Presenter: Yan, Zichen, Nanjing University of Aeronautics and Astronautics

Poster No.: P04

Paper ID: 34

Paper Title: A FeFET-Based Compute-in-Memory Ising Annealing Processor for Irregular Combinatorial Optimization Problems

Author(s): Wei Hu, Sihao Sheng, Zhihao Li, Di Wu, Shuo Yang, Yu Liu, Wenjuan Lu, Ming Zhu, Xiulong Wu, Chunyu Peng

Presenter: Hu, Wei, Anhui University

Poster No.: P05

Paper ID: 45

Paper Title: GRAIN: A Design-Intent-Driven Analog Layout Migration Framework

Author(s): Bingyang Liu, Haoning Jiang, Haoyi Zhang, Xiaohan Gao, Zichen Kong, Xiyuan Tang, David Z. Pan, Yibo Lin

Presenter: Zhang, Haoyi, Peking University

Poster No.: P06

Paper ID: 48

Paper Title: RAMP: An RRAM-Based Compute-in-memory Architecture for mmWave Radar Point Cloud Generation

Author(s): Shulin Zhu, Xuliang Yu, Yi Zhou, Wei-Han Yu, Liang Zhao

Presenter: Zhu, Shulin, Zhejiang University



Poster No.: P07

Paper ID: 50

Paper Title: RefEvo: Agentic Design with Co-Evolutionary Verification for Agile Reference Model Generation

Author(s): Yifan Zhang, Jianmin Ye, Jiahao Yang, Xi Wang

Presenter: Zhang, Yifan, Southeast University

Poster No.: P08

Paper ID: 51

Paper Title: PhysSAR: Feasibility-Frontier SAR Sizing with Physics Constraints and LLM Guidance

Author(s): Chang Li, Tianze Wu, Zhidi Yang, Shan Zhang, Dongdong Li, Zhichao Tan, Liang Zhao

Presenter: Li, Chang, Zhejiang University

Poster No.: P09

Paper ID: 58

Paper Title: XFS-MMoe: Cross-Task Feature Stitching with Multi-Gate Mixture-of-Experts for the Prediction of Drop-Impact Reliability

Author(s): Longfei Li, Xinwei Chen, Min Tang, Junfa Mao

Presenter: Li, Longfei, Shanghai Jiao Tong University

Poster No.: P10

Paper ID: 63

Paper Title: A Physics-Aware Deep Learning Framework for Closed-Loop FinFET DTCO

Author(s): Hanshuo Geng, Guangxin Guo, Yifan Tai, Yuqing Han, Hailong You

Presenter: Geng, Hanshuo, Xidian University

Poster No.: P11

Paper ID: 84

Paper Title: A Controller-Wrapper Decoupling Architecture Utilizing Reusable Atomic Test Algorithm on Memory Testing

Author(s): Liangyong Chen, Xiaoyan Zhang, Zonggang Xiong, Zhaoji Wei, Weiye Ren, Jun Xia, Fan Yang

Presenter: Chen, Liangyong, Shenzhen Funxin Technology Corporation

Poster No.: P12

Paper ID: 85

Paper Title: A Fault Diagnosis Method of SerDes Circuit

Author(s): Yujie Li, Xiaole Cui, Juncheng Pu

Presenter: Li, Yujie, Peking University Shenzhen Graduate School

Poster No.: P13

Paper ID: 101

Paper Title: LLM-AlsAgent: A Text-Editing Proximal Policy Optimization Framework for Approximate Logic Synthesis

Author(s): Kuo Liao, Weichuan Zuo, Jienan Chen

Presenter: Chen, Jienan, University of Electronic Science and Technology of China

Poster No.: P14

Paper ID: 106

Paper Title: Efficient Majority-Based Resubstitution via Two-Step Filtering

Author(s): Yulin Chen, Zhang Hu, Chengyu Ma, Yinshui Xia, Lunyao Wang, Zhufei Chu

Presenter: Chen, Yulin, Ningbo University



Poster No.: P15

Paper ID: 114

Paper Title: PDNDataset: A Machine Learning-Based Dataset for Multi-Load Copper-Plane DC IR-Drop Analysis

Author(s): Yuxi Ji, Yubin Song, Rui Chen, Jienan Chen

Presenter: **Chen, Jienan**, University of Electronic Science and Technology of China

Poster No.: P16

Paper ID: 116

Paper Title: ECOS Studio: An RTL-to-Chip Silicon Design Solution with Open-Source EDA, IP, and PDK

Author(s): Xueyan Zhao, Yuchi Miao, Tong Liu, Zihao Yu, Zengrong Huang, Zhipeng Huang, Zhisheng Zeng, Hao Wang, Weiguo Li, Yihang Qiu, Qiming Chu, Tianji Liu, Yipei Xu, Ruihao Gao, Yujia Peng, Lida Xu, Yuyang Miao, Chunan Zhuang, Hanzhang Liu, Biwei Xie, Yungang Bao

Presenter: **Zhao, Xueyan**, Institute of Computing Technology, Chinese Academy of Sciences

Poster No.: P17

Paper ID: 139

Paper Title: IC-SELECT: Agentic RAG-Based Component Selection for System-Level Analog IC Design

Author(s): Yubo Zhang, Bijian Lan, Qing Peng, Jing Wan

Presenter: **Zhang, Yubo**, Fudan University

Poster No.: P18

Paper ID: 152

Paper Title: Finding the Physics-Aware Winning Ticket: A Hardware-Efficient Capacitance Extraction Framework via Feature Derivation, Pruning and Quantization

Author(s): Jian Zhang, Boyang Yan, Zifei Cai, Xiangshui Miao, Xingsheng Wang

Presenter: **Yan, Boyang**, Huazhong University of Science and Technology

Poster No.: P19

Paper ID: 167

Paper Title: 1FeFET-1FeCap: Non-volatile Charge-Domain CiM Design with Robust Sensing and MLC Capability

Author(s): Zhongtian Yu, Zhixuan Zheng, Hengbin Gao, Yuxiao Jiang, Xunzhao Yin

Presenter: **Yu, Zhongtian**, Zhejiang University

Poster No.: P20

Paper ID: 192

Paper Title: A Systematic Fault Injection Framework for Deep Neural Networks in Space Applications and Reliability Analysis

Author(s): Fenlin Zou, Bi Wang, Yulong Qiu, Chao Wang, Zhaohao Wang

Presenter: **Zou, Fenlin**, Beihang University

Poster No.: P21

Paper ID: 205

Paper Title: DiffHLS: Differential Learning for High-Level Synthesis QoR Prediction with GNNs and LLM Code Embeddings

Author(s): Zedong Peng, Zeju Li, Qiang Xu, Jieru Zhao

Presenter: **Peng, Zedong**, Shanghai Jiao Tong University

Poster No.: P22

Paper ID: 207



Paper Title: STP-Based Combinational Equivalence Checking for Ternary Logic Synthesis

Author(s): Yanzhen Wang, Zhang Hu, Jingning Yu, Chengyu Ma, Yinshui Xia, Lunyao Wang, Zhufei Chu

Presenter: Wang, Yanzhen, Ningbo University

Poster No.: P23

Paper ID: 225

Paper Title: A Bottom-Up Two-Level Compilation Optimization Framework for RTL Simulation Acceleration

Author(s): Jiepeng Zheng, Minghua Shen

Presenter: Zheng, Jiepeng, Sun Yat-sen University

Poster No.: P24

Paper ID: 229

Paper Title: OptoKD: Transferring Vision Model Capabilities to Compact, Efficient and Robust Opto-Electronic Hybrid Models via Distillation

Author(s): Mingyuan Li, Ziyang Huang, Xing Mou, Xueyan Shan, Hongwei Hu, Qisheng Yang

Presenter: Li, Mingyuan, Hunan University

Poster No.: P25

Paper ID: 253

Paper Title: HiMoE-NoC: A Hierarchical Network-on-Chip Design Framework for MoE Inference on 3D-Stacked DRAM Accelerators

Author(s): Xiping Dong, Yihan Yin, Guangyu Sun

Presenter: Dong, Xiping, Peking University

Poster No.: P26

Paper ID: 268

Paper Title: A Novel Superposition Method for Chip-Scale Thermal Stress Simulation of 3D ICs

Author(s): Zhongkun Guo, Linbo Zhang, Tao Cui

Presenter: Guo, Zhongkun, Academy of Mathematics and Systems Science, Chinese Academy of Sciences; University of Chinese Academy of Sciences



Shanghai Peifeng Tunan Semiconductor Co., Ltd. is a leading provider of multi-scale and multi-physics simulation software for the semiconductor industry. After nearly two decades of development and over ten major version iterations, PFTN is proud to introduce Mozz-TCAD, the next-generation process and device TCAD simulator. Equipped with the most advanced physical models, Mozz-TCAD is serving leading-edge logic and memory IDMs and foundries in technology development and path-finding.

PFTN offers a full suite of manufacturing EDA software for design-process co-optimization (DTCO), including optical proximity effect correction (OPC), process and device simulation (TCAD), topography emulation, SPICE modeling, circuit simulation and parasitic parameter extraction. PFTN also offers semiconductor radiation effect simulation software and radiation-tolerant IP products have served dozens of world-wide customers in the aerospace industry.

Product Matrix

Physics-based Simulators	Interactive Environment	Design Connectivity
Mozz Device Device Simulation	Mozz Workbench Workflow & DoE	Mozz RCEX 3D RC Extraction
Mozz Process Process Simulation	Mozz Visual 2D/3D Visualization	Mozz Model Builder TCAD-to-SPICE
GK-Photonics Multi-physics Simulation	Mozz Structure Structure Editor	Mozz PEX Builder TCAD-to-PEX
Mozz ProcEmu Fast Process Emulation	Mozz ProcFlow Process Flow & Recipe	Mozz Tech builder APR Tech Files
	Mozz Intelligence Calibration & Optimization	

Featured Technologies

Rigorous and proven 3D TCAD simulator engine

- Traditional polyhedral geometry engine
- Next-generation voxel-based geometry engine

Complete suite advanced physical models

- Rich material database with API for extension and customization
- k.p-based band structure and quantum correction models
- Multi-valley stress-dependent mobility models
- Ballistic transport models
- Crystal monte-carlo implantation models

Support for modern high-performance computing platforms

- Linux and Windows operating system
- X86, ARM CPU and Nvidia GPU
- Massive parallel computation

Semitronix, A Leading Provider of EDA Software and Wafer-Level Electrical Test Equipment

Semitronix Corporation is a leading supplier of characterization and yield improvement solutions encompassing software, hardware, and services for the semiconductor industry, which has numerous successful cases covering multiple integrated circuit process nodes.

Service Value & Advantages:

				
Accelerate design-to-market cycles	Ensure superior quality and yield	Speed up R&D and process stability	Lower costs and boost efficiency	Drive optimization through data insights

"One-Stop" Chip Yield Enhancement Solutions

As IC products demand higher performance and lower costs, continuous process iteration makes yield enhancement critical to success.

With 20 years of expertise, Semitronix provides a proprietary, full-cycle yield enhancement solution. Our one-stop services—integrating test chip design, electrical testing, and data analysis—accelerate R&D cycles and maximize production yield for new process nodes.

• EDA for Yield Enhancement

Test chip EDA tools (SmtCell/TCMagic/ATCompiler/Dense Array/ICSpider) to optimize manufacturing processes and identify yield-limiting factors.



• DFT & DFM EDA Tools

QuanTest (DFT) Tools: Support diverse applications and chip scales for Design-for-Test requirements.

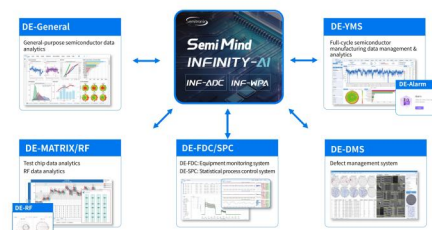
DFM Tools: Multi-dimensionally detect and predict layout manufacturing risks to ensure design success.

QuanTest DFT Tools	DFM Tools
• QuanTest Scan • QuanTest ATPG • QuanTest JTAG • QuanTest IJTAG • QuanTest MBIST	• CMPEXP • PatternScan • LayoutVision • LayoutInsight • VirtualYield • CviDGS • CviDGM • CviDGC • AI-TPG •

• Big Data Analytics System

A semiconductor data intelligence and monitoring platform featuring

DE - G, DE - YMS, DE - DMS, DE - Alarm, DE - iCASE, DE - FDC, DE - SPC, INF - ADC, DE-DN, YAD, INF - AI, SemiMind, DE-FBM, INF-TPC, QuickRoot, AutoInk.



• Wafer-Level Electrical Test Equipment

Supports high-parallel R&D testing, high-efficiency WAT mass production, and WLR reliability testing.

• Wafer-Level Burn-In (WLBI) Systems

Wafer-level burn-in systems specifically designed for SiC and GaN wafer-level burn-in testing.



Linking Chip Design & Manufacture Innovative EDA Solutions

Primarius is a global competitive EDA company
Committed to delivering Innovative DTCO EDA solutions



Empyrean Technology, founded in 2009, is an EDA and services provider to the global semiconductor industry. The company strives to be the world's leading EDA provider, delivering a comprehensive front-to-back design flow across all industries.

In the EDA domain, Empyrean Technology offers EDA solutions for full custom design, digital SoC design solutions, foundry EDA solutions, advanced packaging design solutions and 3DIC design solutions. Additionally, EDA solutions for full custom design include complete solutions for analog design, complete solutions for memory IC design, complete solutions for RF IC design, and complete solutions for flat panel display design. The company provides EDA-related services, including foundry design enablement services. EDA and services find applications mainly in the fields of IC design, foundry, and packaging.

Empyrean is headquartered in Beijing, with major R&D centers in Nanjing, Chengdu, Shenzhen, Shanghai, Hongkong, Xi'an, Wuhan, Chongqing, Xiamen and Suzhou in China.



 www.emyrean-tech.com

 info@emyrean-tech.com

 Beijing · Nanjing · Chengdu · Shenzhen · Shanghai · HongKong · Xi'an · Wuhan · Chongqing · Xiamen · Suzhou

2.5D/3D Advanced Packaging

EDA⁺ STCO : A New Paradigm



3Sheng[®] Integration Platform 2.5D/3D IC EDA

3Sheng Zenith[®] Architecture Design



System-Level Modeling	Chiplet Partitioning
Design Space Exploration	IP Implementation
μbump/TSV Planning	Standardized Chiplet Modeling
Multi-objective Planning	Early Architecture Analysis for Chiplets

Chiplet splitting needs to start from the top-level architecture division Chiplet partition, IP implementation, μbump plan from D2D interconnection

3Sheng Ranger[®] Physical Implementation

Physical Implementation

A complete physical design solution for 2.5D/3D chiplet partitioning, including layout and routing



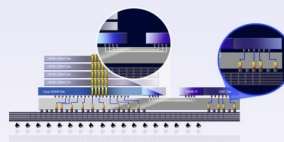
Floorplanning	Routing
Placement	

Chiplet IP Customization	D2D Interfaces	TSV/μbump I/O Co-optimization
3D Global Optimization	Compatibility with Multiple Si-interposer Process	Thermal Co-design Floorplanning

3Sheng Ocean[®] Multi-Die DFT

Multi-Die DFT

Multi-die DFT & Fault Tolerance



DFT
Fault Tolerance

Supports Single-die Scan Testing	On-chip Circuitry with Low Overhead and High Repair Rates	Multi-die DFT Testing
Supports Detection of Multiple TSV Defects	Complies with IEEE 1838 3D IC DFT Standard	IEEE Std 1149.1 and IEEE Std 1500

3Sheng Volcano[®] Co-Simulation

Co-Simulation



Unified physical simulation platform for collaborative analysis of multiple physical fields

SI	PI	Thermal
----	----	---------

S, Y, Z Parameter Extraction	DC IR Drop Analysis	Propagation Coefficient Analysis
Warpage Model Analysis	Current & Power Density Distribution	Electrothermal Coupling Simulation

3Sheng Stratify[®] Multi-Die Physical Verification

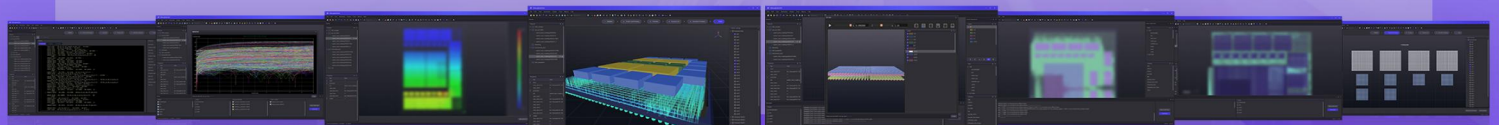
Multi-Die Physical Verification



System-level connectivity check, Multi-die DRC runset support

Multi-die DRC	Multi-die LVS
---------------	---------------

System Connectivity Check	SPICE Model Analysis	Signal Electrical Integrity
μbump Micro Rule	RDL Open/Short Check	Cross-domain Clock Distribution



JOIN US LEAD COMPUTING FUTURE

WITMEM*
FUTURE CREATOR

Recruiting
Master's, PhD, Postdoctoral
worldwide

Research Areas



CIM Architecture &
Algorithms



High-Efficiency
Chip Design



Semiconductor
Devices & Process



AI Model
Implementation

Salary and Benefits

- * Competitive salary and benefits package
- * Chance to work on groundbreaking technologies
- * Strategic R&D roles with real-world impact
- * Clear path for professional advancement



Career
Opportunities



Internship
Programs

Witmem
知存科技

LEADING
COMPUTING-IN-MEMORY
TECHNOLOGY

Bridging Digital and Physical Worlds

Realizing Ideas as Products

UniVista Industrial Software Group

Shanghai UniVista Industrial Software Group Co., Ltd. ("UniVista") is an EDA, IP and design solutions provider committed to serving semiconductor companies worldwide. We aim to become a trusted partner that achieves continuous market success through technology innovation.

▪ Chip-level EDA (Full-Flow Digital IC Verification + Full-Flow DFT Platform + Agentic AI EDA Platform)

• Digital Chip Verification

- UniVista Simulator Plus – UVS+, UniVista Debugger Plus – UVD+, UniVista Verification Productivity System – VPS
- UniVista Unified Verification Hardware System – UVHS/UVHS-2
- UniVista Hyperscale Emulator – UVHP
- PHINE DESIGN Advanced Solo Prototyping Platform – PD-AS
- Virtual Prototyping Toolset – UniVista V-Builder/vSpace

• Digital Chip Implementation

- DFT Platform – UniVista Tespert BSCAN/MBIST/ATPG/DIAG/YIELD

• AI Platform

- Agentic AI EDA Platform for Digital Design – UniVista Design Agent (UDA) 2.0

▪ IP Solutions: High Performance IP and Customized Design Solutions

- Chiplet Interface & IO Die solution
- HiPi IP/VIP solution
- Memory Interface solution: HBM3/E, LPDDR5, DDR5 IP
- Multi-protocol 32G SerDes PHY solution: UniVista 32G MPS IP
- PCIe Gen5 IP solution
- AI computing networking scale-out solution – RDMA IP, scale-up solution – PAXI IP
- Ultra Ethernet IP solution – UniVista UEC MAC IP
- Controller IP for Ethernet, FlexE, Interlaken, etc

▪ System-level EDA (High-Performance Large-Scale PCB Design Platform + Advanced Packaging-Level Full-Scenario Design Platform)

- Design Data Management Platform – UniVista EDMPro
- Electronic System Design Platform – UniVista Archer Schematic
- PCB Design Platform – UniVista Archer PCB
- Co-Design Platform – UniVista Integrator (UVI)

Follow us on WeChat



Offices:

China: Shanghai, Beijing, Nanjing, Xi'an, Wuxi, Hangzhou, Changsha, Chengdu, Xiamen, Shenzhen

Overseas:

Singapore, Japan

Explore more at:

www.univista-isg.com

Contact us:

sales@univista-isg.com

Chips Wherever, **GIGA** Within

Building a leading domestic full-flow digital IC toolchain

GIGA is a high-tech enterprise specializing in Electronic Design Automation (EDA) solutions.

Founded in 2018 and supported by a robust local industrial ecosystem, GIGA has established R&D and technical support centers in Shenzhen, Shanghai, Beijing, and Chengdu. The company is committed to developing a leading domestic full-flow digital IC toolchain.

Leveraging its trusted local R&D capabilities and responsive technical support, GIGA strives to better address the rapidly evolving innovation demands of semiconductor design companies, earning industry recognition through its advanced EDA products and technological expertise.



2018

Establishment of the company



SoC EDA

Building a complete domestic SoC EDA platform



> 88%

Percentage of R&D and technical support personnel



> 80%

Percentage of graduate degree and above

Fantaisie®
Detailed-Route-Centric P&R Architecture

RocSyn®
Layout Driven Logic Synthesis

ChimeTime®
Timing Signoff Solution

HesVesPower®
EMIR Signoff Tool



Shenzhen Giga Design Automation Co., Ltd

Company Website: www.giga-da.com

Official Email: market@giga-da.com



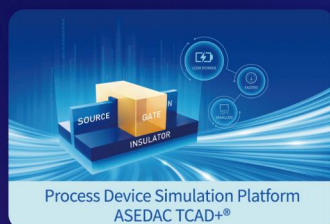
China's Frontrunner in Semiconductor Intelligent Manufacturing EDA

Company Profile

Advanced Manufacturing EDA Co., Ltd. (AMEGADAC) was founded in September 2019 with a registered capital of RMB 360 million. Headquartered in Hefei, the company has established wholly-owned subsidiaries in Shanghai, Beijing, Guangzhou, Jinan, Shenzhen, Wuhan, Hong Kong, and Hangzhou.

AMEGADAC is a high-tech enterprise specializing in R&D and delivery of professional software and services for integrated smart manufacturing solutions. With industrial software as its core driver, we actively build collaborative partnerships across the entire industrial chain, facilitating the deep integration of fragmented segments into a fully connected ecosystem. By leveraging cutting-edge technologies such as intelligent systems, we have established a distinctive and comprehensive portfolio covering the full lifecycle of industrial manufacturing software, delivering pivotal breakthroughs in the field of smart manufacturing.

Product Description



Contact Information

☎ 0551-67126702

✉ marketing@amedac.com

🌐 www.amedac.com



Scan to follow us

POWER ANALYSIS

Power Analysis

- Static Power Analysis
- Dynamic Power Analysis



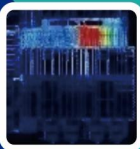
Voltage Drop (IR) Analysis

- Static/Dynamic Voltage Drop Analysis
- ECO Editor

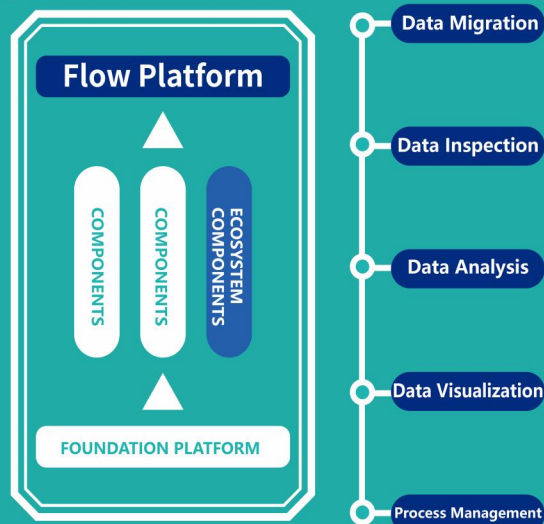


Electromigration (EM) Analysis

- Power/Signal EM Analysis



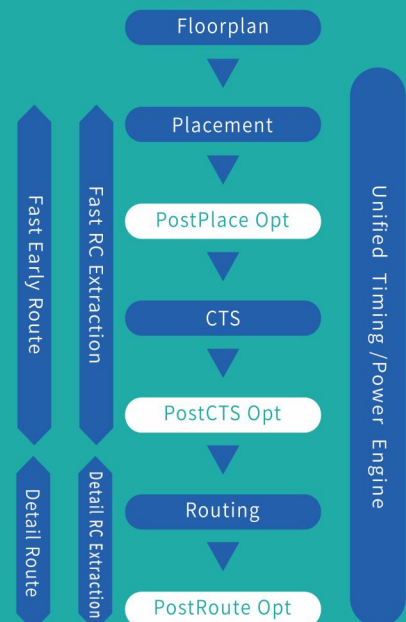
EDA FOUNDATION PLATFORM AND ECOSYSTEM COMPONENTS



PLACEMENT AND ROUTING

Optimus Prime

DIGITAL CHIP NETLIST-TO-GDSII FULL SOLUTION



CHIP DESIGN SERVICE CENTER

DFT ARTIST



Chip Design
Service
Center



Advanced AI
Core Expert
System



Industry-Lead-
ing Automa-
tion System



Exclusively
Product
Validated



Experienced
and Strong
Technical
Team



Flexible and
Diversified
Service-
Modes



Headquarters Address:
Room 2101, 21st Floor, Building B,
Corona Borealis, Phase 6 of IPark,
Xinwu District, Wuxi City, Jiangsu Province

ISED A 2026 Contact Us

Conference Secretary: Joyce Zhong

Email: iseda@eda2.com



[ISED A 2026 Website](https://www.iseda2026.com)

ADVISORS

IEEE/CEDA
ACM/SIGDA

ORGANIZER

EDA Ecosystem Development Accelerator (EDA²)

CO-ORGANIZERS

Singapore Semiconductor Industry Association
Siliconexus Pte Ltd

