72nd International Electron Devices Meeting

2026 Technical Program

Devices at the Heart of the Intelligence Revolution

2026 International Electron Devices Meeting

December 12-16, 2026 | Hilton San Francisco Union Square San Francisco, CA
www.ieee-iedm.org



IEEE Electron Devices Society

As Of November 26, 2026

A DISTINGUISHED WELCOME TO IEDM 2026



GENERAL CHAIR
Meng-Fan (Marvin) Chang
National Tsing Hua University & TSMC



TECHNICAL PROGRAM CHAIR
Jungwoo Joh
Texas Instruments



TECHNICAL PROGRAM VICE CHAIR
Rihito Kuroda
Tohoku University

On behalf of the IEEE International Electron Devices Meeting (IEDM) Committee, it is my distinct honor to welcome you to the 72nd Annual IEDM, taking place in San Francisco, December 12–16, 2026.

For more than seven decades, IEDM has served as the world's premier forum for presenting breakthrough advances in semiconductor and electronic device technologies. Each year, the meeting brings together leading researchers, industry pioneers, academic innovators, and emerging talent from around the world to share discoveries, exchange ideas, and explore the technologies that will define the future of electronics.

The theme of IEDM 2026, **"Devices at the Heart of the Intelligence Revolution,"** reflects a defining moment for our field. The rapid expansion of artificial intelligence and intelligent systems is driving unprecedented demands for computing performance, energy efficiency, memory, connectivity, sensing, and integration. Meeting these demands will depend fundamentally on continued innovation in semiconductor devices and technologies.

IEDM 2026 captures the breadth and momentum of this transformation. Advances in logic and memory, new materials and device architectures, heterogeneous and 3D integration, power and RF electronics, photonics, sensors, quantum and emerging computing technologies are increasingly interconnected. Innovation now extends across traditional boundaries—from materials and devices to circuits, architectures, systems, and applications. IEDM provides a unique global forum where these technologies and communities converge.

Across five days, IEDM 2026 will showcase breakthrough research, emerging technologies, and forward-looking perspectives through a comprehensive program spanning technical presentations, educational programs, discussions, and exhibitions. Together, they provide a broad view of both the latest scientific and technological advances and the directions shaping the future of the semiconductor industry. Program information and paper summaries are available at the IEDM website, with conference materials also available electronically through IEEE Xplore and IEDM OnDemand.

Beyond its technical program, IEDM is a global community built on the exchange of ideas and collaboration among colleagues from industry and academia, established leaders and the next generation of researchers. These interactions have long been an essential part of IEDM and continue to strengthen our community and inspire future innovation.

I extend my sincere appreciation to all authors and speakers whose outstanding contributions form the foundation of IEDM 2026. I also thank the members of the IEDM Executive Committee and Technical Subcommittees for their dedication and expertise in creating this year's conference. IEDM is proudly sponsored by the IEEE Electron Devices Society (EDS), whose continued support helps advance electron-device science and technology worldwide.

Welcome to the 72nd IEDM. We look forward to five exciting days of discovery, exchange, and collaboration as we advance devices at the heart of the intelligence revolution and shape the future of electronics.

Meng-Fan (Marvin) Chang – General Chair

Executive Committee **4**
Introduction **4**
Digital & Social Media 4
Topics of Interest 5
Schedule-at-a-Glance 8
Program **10**
Saturday, December 12, 2026 ..... **10**
Tutorials 10
Tutorial 1: Evolution of the Transistor from Planar to Nanosheet and Beyond, ... 10
1:30 PM– 2:50 PM 10
Tutorial 2: Chiplet, Heterogeneous Integration, 3D Stacking Technologies Principles and State of the Art 10
Tutorial 3: Memory Technologies and System Co-Optimization in the AI Era: Past, Present and Future 11
Tutorial 4: AI in Physics-based Semiconductor Modeling: From PDE-constrained Learning to Agentic TCAD/DTCO Workflows 11
Tutorial 5: Diverse Memory Technologies for Adaptive Computing: From Device Physics to Cooperative Architectures 11
Tutorial 6: Advanced Optical Devices and Wafer-Scale Assembly for Next-Generation CPO and Optical I/O 12
Sunday, December 13, 2026 ..... **12**
Short Course 1: Logic and System Innovations Powering the AI Era 12
Short Course 2: Powering Next-Generation AI and HPC with Advanced Memory Technologies 13
Monday, December 14, 2026 ..... **15**
Keynote Session 15
Session 1: Welcome and Awards Presentations 15
1-1 AI Driven Semiconductor Paradigm Shift: Optimizing the Semiconductor Industry from Device Scaling to System Integration (Invited) 16
1-2 Driving AI Memory Innovation: From Device Technology to AI Memory Platform (Invited) 16
1-3 Technology Trends in the IDM Semiconductor Business (Invited) 16
Technical Sessions 17
Session 2: MT | Ferroelectric Memories 17
Session 3: ALT | Platform and DTCO 19
Session 4: SMB | Optical, Radiation, Quantum Sensing 21
Session 5: NC | Neural Network Inference with Emerging Technologies 23
Session 6: EDT | 2D Module Engineering and Heterogenous Integration 25
Session 7: MS | Advanced Device Physics & Compact Modeling 28
Session 8: ODI | Integrated Photonics and Single Photon Devices 30
Session 9: PMA | Advanced (Ultra-) Wide-Bandgap Power Devices and System Integration 33
Session 10: RSD | Oxide-Semiconductor FET Reliability and Degradation Physics 36
Tuesday, December 15, 2026 ..... **39**
Session 11: MT | NAND & DRAM 39
Session 12: ALT | Oxide Semiconductor and 2D Channel Device 41
Session 13: SMB | Physical, Chemical, and Multimodal Integrated Sensors 45
Session 14: MS | Memory, Device and System-Technology Co-Optimization 47
Session 15: NC | Probablistics and In-Memory Computing for Robotics and Optimization 49
Session 16: EDT | Novel Computing Paradigms 52
Session 17: FS | Semiconductors for Quantum Computing 55
Session 18: RSD | Reliability of Ferroelectric Materials and Devices 57
Session 19: PMA | Advances in RF, mmWave, and Sub-THz III-V and Wide-Bandgap Transistor Technologies for 6G and Beyond 60
Session 20: MT | Nonvolatile Memories 63
Session 21: ALT | BEOL, Thermal and Packaging 65
Session 22: SMB | MEMS Resonators and Transducers 68

Session 23: MS | Atomic Simulation: Devices, Circuits, and Process 70
Session 24: NC | LLM Acceleration Using In-Memory Compute 72
Session 25: EDT | Oxide-Channel Devices 75
Session 26: ODI | Emerging Photonic Devices 78
Session 27: RSD | Component-to-System Level Reliability and Thermal Management 81
Session 28: PMA | Advanced GaN Power Devices: Power Delivery, Bidirectional and Low Voltage 83
Session 29: Evening Panel | Rewriting the Rules of Semiconductor Manufacturing: Why the Old Manufacturing Model May Not Be Enough 86
Wednesday, December 16, 2026 86
Session 30: PMA | High-Performance RF FETs and Switches 86
Session 31: NC | Edge Intelligence and On-Chip Learning 89
Session 32: EDT | Ferroelectric Devices 92
Session 33: ALT | Advanced Device and Process Innovations 95
Session 34: MS | Multiscale, AI & Advanced Simulation Methodologies 98
Session 35: RSD | Memory Technologies and Compute-in-Memory 100
Session 36: ODI | Advanced Image Sensors 103
Session 37: MT | Oxide Semiconductor Memories 106
Session 38: EDT | Emerging Memories 108
Session 39: FS | System-Level Optimization Combining Logic, Memory and Packaging 111
Session 40: MS | Emerging Devices, Ferroelectrics & Nonvolatile Memory 113
Session 41: RSD | Reliability of Next-Generation Transistors 115
Session 42: SMB | Smart Sensors and Bio-Integrated Interfaces 117
IEDM Committee List 120

Executive Committee



Seated Left to Right
First Row: Uygar Avci, Courses Co-Chair (Tutorials); Huimei Zhou, SubCommittee Chair - RSD; John Paul Strachan, Publicity Chair; Rihito Kuroda, Technical Program Vice Chair; Meng-Fan (Marvin) Chang, General Chair; Jungwoo Joh, Technical Program Chair; Srabanti Chowdhury, Publications Chair; Gaudenzio Meneghesso, Publications Co-Chair; Kang-ill Seo, Courses Chair;

Second Row: Dongyean Oh, SubCommittee Chair - MS; Geunsook Park, SubCommittee Chair - ODI; Vita Pi-Ho Hu, Asian Arrangements Chair; Sangbum Kim, Asian Arrangements Co-Chair; Chung-Hsun Lin, SubCommittee Chair - ALT; Sandy Liao, Publicity Co-Chair; Dechao Guo, Focus Session Co-Chair; Olga Spahn, Focus Session Chair; Johannes Mueller, SubCommittee Chair - MT; Shawn Boon, Conference Manager;

Third Row: Geert Eneman, Courses Co-Chair (Short Courses); Seyoung Kim, SubCommittee Chair - NC; Daniele Ielmini, European Arrangements Co-Chair; Gail Sparks-Riegel, Conference Manager; Elena Gnani, Virtual Arrangements Chair; Roozbeh Tabrizian, SubCommittee Chair - SMB; Zhikai Tang, SubCommittee Chair - PMA; Michael Waltl, European Arrangements Chair;

Not pictured: Maud Vinet, SubCommittee Chair - EDT; Jan Hoentschel, Treasurer and Polly Hocking, Conference Manager

Introduction

IEEE International Electron Devices Meeting (IEDM) is the world's preeminent forum for reporting technological breakthroughs in the areas of semiconductor and electronic device technology, design, manufacturing, physics, and modeling. IEDM is the flagship conference for nanometer-scale CMOS transistor technology, advanced memory, displays, sensors, MEMS devices, novel quantum and nano-scale devices and phenomenology, optoelectronics, devices for power and energy harvesting, high-speed devices, as well as process technology and device modeling and simulation.

Digital & Social Media

- LinkedIn: https://www.linkedin.com/groups/7475096/
- Twitter: https://twitter.com/ieee_iedm
- Facebook: https://www.facebook.com/IEEE.IEDM
- YouTube: https://www.youtube.com/@iedminternationalelectrond4318
- Wikipedia: https://en.wikipedia.org/wiki/International_Electron_Devices_Meeting

Topics of Interest

ADVANCED LOGIC TECHNOLOGY (ALT)

Papers are solicited in the areas of CMOS platform technologies and applications (e.g., HPC, LOP, mobile, automotive, low-temperature CMOS, etc.), logic devices and circuits, process integration schemes for advanced nodes, innovations in materials, process, and metrology techniques, design technology co-optimization (DTCO) and system technology co-optimization (STCO), especially to meet demands in the AI era. Platform technologies include state-of-the-art Si and beyond-Si channel devices, gate-all-around devices, stacked devices with different polarity transistors, advanced interconnect, novel power distribution integration schemes, heterogeneous 2.5D/3D integration schemes, and BEOL-compatible transistors. Device architecture, device design and analysis, process integration, module advancements in process and patterning, metrology, physical layout effects, techniques for reduced variability, yield, thermal management, methodologies, and solutions for DTCO/STCO in the solicited areas are of high interest.

EMERGING DEVICE and COMPUTE TECHNOLOGY (EDT)

Papers are solicited on emerging nano-electronic devices and physics. This includes devices based on novel transport and control mechanisms such as tunnel FET, negative capacitance FET, cold-source FET, cryogenic devices, topological materials and devices, phase transitions, ferroelectrics and quantum effects. Devices based on low-dimensional systems including 2D materials, CNTs, nanowires, single electron transistors and quantum dots are welcomed. Exploratory devices with novel device functions and/or novel materials for neuromorphic compute, approximate, probabilistic and analog compute, and non-charge-based compute such as spintronics are key topics. Furthermore, emerging state machines and time dynamical compute systems are also of interest. Qubit devices as well as devices and systems designed to enable quantum computing, quantum simulation and quantum annealing are of high interest. Papers in EDT focus primarily on device physics and novel elaboration concepts.

MEMORY TECHNOLOGY (MT)

Papers are solicited across various domains in memory and storage technologies. This encompasses advancements and scaling in established technologies like SRAM, DRAM, and Flash, as well as breakthroughs in emerging innovations such as MRAM, PCM, FRAM, RRAM, ECM, SOM Cross-Point memory, organic memory, and NEMS-based memory. These include research pertaining to materials and devices for memory and selector in electrostatic and atomistic switching mechanisms, as well as the design and implementation of memory cells and arrays in 3D constructs, including stacking and tiering, alongside read/write access mechanisms. Additionally, submissions of novel approaches to homogenous or heterogenous integration and manufacturing techniques of memory in semiconductor fabrication or packaging assembly are welcome. Papers not mainly focused on improvements of memory FoMs but are disclosing novel computing architecture or algorithms such as CiM may be transferred to other technical sections such as EDT or NC at the discretion of the committee.

MODELING and SIMULATION (MS)

Papers are solicited on theoretical or computational studies of electronic devices, including logic, memory, optical components, interconnects, power related devices, display, sensors, MEMS, and bioelectronics. The approaches used in the study may include analytical, numerical, statistical, and machine-learning/AI-based methods applied to structures ranging from atomistic to device dimensions, and up to full-chip dimensions, including physics-based compact modeling. Central to submissions is the innovation of devices, whether through predictive insight into novel device concepts, predictive analysis demonstrating significant device improvements, breakthroughs in theoretical understanding of device operation, advancements in knowledge of device processing facilitating enhanced device performance, novel insights into variability, reliability, and yield issues, or breakthroughs in device optimization based on DTCO and STCO. Topics also include ab-initio/atomistic materials modeling, neuromorphic computing modeling, quantum computing and quantum-resilient hardware modeling, spintronics, low-dimensional devices, ferroelectrics, thermal modeling, optoelectronics, displays and imaging systems modeling, power devices, millimeter-wave and analog technologies, 3D/heterogeneous integration and advanced packaging modeling, electro-chemical/mechanical devices, bio-nano sensors for brain-computer interfaces and innovation in compact modeling. Encouragement is given for comparison with experimental data, model calibration, and utilization of multi-scale simulation chains.

NEUROMORPHIC and NOVEL COMPUTING (NC)

Papers are solicited on advancements in semiconductor memory and logic devices, as well as the circuits and algorithms that leverage them, for new and unconventional compute paradigms. Specific areas of interest span

neuromorphic computing and advanced artificial intelligence (AI) acceleration techniques, broadly encompassing analog/mixed-signal computing, compute-in-memory (CIM) devices and circuits for advanced AI workloads, probabilistic computing, and bio-inspired computing such as spiking neural networks, reservoir computing, combinatorial optimization, and content-addressable memory. Use cases across the compute continuum—from the datacenter to the edge (e.g., in-sensor computing)—are highly contemplated. Submissions demonstrating novel device concepts that fundamentally improve computational efficiency, full hardware integration, explicit device-algorithm co-optimization to mitigate non-ideal device properties, and real-world applications are of highest interest. Papers focusing primarily on near-memory computing may be transferred to the MT subcommittee, while papers centered on optical or quantum computing may be transferred to EDT, at the discretion of the committee.

OPTOELECTRONICS, DISPLAYS, and IMAGING SYSTEMS (ODI)

Papers are solicited on optoelectronics, displays, and imaging systems. This includes novel devices, structures, and integration for image sensors, displays, light sources, photonic devices, and high-speed photodetectors and modulators. New technologies on heterogeneous integration of optoelectronics devices as well as on photonic-electronic integration for optical interconnects, on-chip networks and sensing are welcomed. Papers on quantum photonics, neuromorphic photonics, and plasmonics for quantum computation, sensing and encryption are also of interest. Furthermore, ODI includes CMOS imagers, high-speed and high-time resolution imagers, tacked imagers sensors, singlephoton sensitivity, Time-Of-Flight and non-visible image sensors. In addition, papers on displays of all types, for augmented or virtual reality, holography, TFTs for photonics applications, flexible, stretchable, and/or printed electronics, in-display sensors are encouraged. Papers on displays or light emitting devices with novel materials such as perovskites or quantum dots are also of interest.

POWER, MICROWAVE/MM-WAVE and ANALOG DEVICES/SYSTEMS(PMA)

Contributions are sought on novel circuit topologies, manufacturing processes, supporting modeling (TCAD and compact models), device physics, reliability, and materials (Si, III-Vs, SiC, (Al)GaN, Ga2O3, AlScN, LiNbO3, diamond, LTO, BN, etc.) along with fundamental studies on doping, deep-level traps, interface states, and device reliability for power switching and high frequency devices. Papers are solicited on discrete and integrated power and high frequency (micro, mm-wave and THz) devices and physics, modules and systems. Topics of interest include devices (diodes, BJTs, FETs, super-junction devices, heterostructures, IGBTs, HEMTs, HBTs, light-triggered structures for galvanic isolation and faster switching, bi-directional switches, vertical geometry devices, and device/package/circuit interactions, including thermal management. A wide variety of applications are also within the scope of PMA: power conversion, supply, regulation and conditioning for computers and data centers, motor drives, transportation, solar, wind, smart grid applications, wireless power harvesting/transfer, beam formers, power amplifiers, tunable passives, antenna arrays.

RELIABILITY OF SYSTEMS and DEVICES (RSD)

Papers are solicited that focus on component-level FEOL/MOL/BEOL characterization and reliability modeling, reliability evaluation, both experimental and modeling, of devices dedicated to analog, logic, and memory applications, interconnects, circuits, and systems. In addition to Si-based technologies, authors are encouraged to submit their recent achievements made employing other material systems, such as SiGe, IGZO, ferroelectric materials, 2D materials, etc. The reliability topics include, for FEOL, transistor degradation due to hot carriers, bias temperature instabilities, random telegraph noise, dielectric SILC and breakdown as well as modeling the aging and wearout behavior. For MOL/BEOL, topics include the breakdown of MOL spacers and BEOL dielectrics, electromigration, and stress migration failures of contacts and interconnects. For product, system, and circuit reliability, topics include latch-up, ESD, soft error mechanisms, variability-aware design, and design for reliability, robustness, and security of electronic circuits and systems. Of particular interest are investigations of degradation mechanisms for devices, circuits, and systems in the following areas: conventional and emerging memories; beyond CMOS devices; 3D IC package reliability; more-than-Moore applications; biomedical devices and systems; automotive and aerospace.

SENSORS, MEMS, and BIOELECTRONICS (SMB)

Papers are solicited in the areas of sensors, MEMS/NEMS, microfluidics and lab-on-chip, BioMEMS, and bioelectronic devices and systems, with emphasis on new device concepts, integrated CMOS implementations, embedded intelligence, and multifunctional microsystems for applications in health, communication, mobility, energy, defense, and extreme environments. Sensors include chemical, biological, acoustic, electrical, electrochemical, magnetic, and mechanical modalities. Topics of interest include physical and biochemical sensors, actuators, resonators, inertial microsystems, RF MEMS, SAW/BAW and acoustic devices, optomechanical

devices, energy-harvesting and micropower devices, and microsystems that integrate sensing, actuation, readout, control, and packaging. Bioelectronics covers hybrid organic/inorganic devices, point-of-care and implantable systems, neural and bioelectronic interfaces, flexible and soft devices, multimodal biomedical sensing, and nanosensors for brain-computer interfaces.

Schedule-at-a-Glance

SATURDAY - TUTORIALS (Separate Registration Required)				
12:00 PM	-	**6:00 PM**	**East Lounge**	**Registration** (12-3:30 pm: Tutorial registrants only, Open to all at 3:30 pm)
1:30 PM	-	2:50 PM	Cont 1-5	T1 | Evolution of the Transistor from Planar to Nanosheet and Beyond
1:30 PM	-	2:50 PM	Cont 6-9	T2 | Chiplet, Heterogeneous Integration, 3D Stacking Technologies Principles and State of the Art
3:05 PM	-	4:25 PM	Cont 1-5	T3 | Memory Technologies and System Co-Optimization in the AI Era: Past, Present and Future
3:05 PM	-	4:25 PM	Cont 6-9	T4 | AI in Physics-based Semiconductor Modeling: From PDE-constrained Learning to Agentic TCAD/DTCO Workflows
4:40 PM	-	6:00 PM	Cont 1-5	T5 | Diverse Memory Technologies for Adaptive Computing: From Device Physics to Cooperative Architectures
4:40 PM	-	6:00 PM	Cont 6-9	T6 | Advanced Optical Devices and Wafer-Scale Assembly for next-generation CPO and Optical I/O
SUNDAY - SHORT COURSES (Separate Registration Required)				
8:00 AM	-	**5:00 PM**	**East Lounge**	**Registration**
9:00 AM	-	6:00 PM	Cont 1-5	SC1 | Short Course 1: Logic and System Innovations Powering the AI Era
9:00 AM	-	6:00 PM	Cont 6-9	SC2 | Short Course 2: Powering Next-Generation AI and HPC with Advanced Memory Technologies
MONDAY				
8:00 AM	-	**5:00 PM**	**East Lounge**	**Registration**
8:30 AM	-	**9:15 AM**	**Grand A**	**Coffee Break**
9:00 AM	-	**12:00 PM**	**Grand B**	**1 | Keynote**
12:00 PM	-	**5:30 PM**	**Yosemite**	**Exhibits Open**
1:45 PM	-	**5:35 PM**	**Technical Sessions**	
1:45 PM	-	4:45 PM	Grand A	2 | MT | Ferroelectric Memories
1:45 PM	-	4:20 PM	Grand B	3 | ALT | Platform and DTCO
1:45 PM	-	4:45 PM	Cont 1-3	4 | SMB | Optical, Radiation, Quantum Sensing
1:45 PM	-	4:45 PM	Cont 4	5 | NC | Neural Network Inference with Emerging Technologies
1:45 PM	-	5:35 PM	Cont 5	6 | EDT | 2D Module Engineering and Heterogenous Integration
1:45 PM	-	4:45 PM	Cont 6	7 | MS | Advanced Device Physics & Compact Modeling
1:45 PM	-	5:10 PM	Cont 7-9	8 | ODI | Integrated Photonics and Single Photon Devices
1:45 PM	-	5:35 PM	Imperial A	9 | PMA | Advanced (Ultra-) Wide-Bandgap Power Devices and System Integration
1:45 PM	-	5:35 PM	Imperial B	10 | RSD | Oxide-Semiconductor FET Reliability and Degradation Physics
3:00 PM	-	**4:15 PM**	**Yosemite**	**Coffee Break with Exhibitors**
6:30 PM	-	**8:00 PM**	**Grand B**	**Monday Evening Reception**
TUESDAY				
8:30 AM	-	**5:00 PM**	**East Lounge**	**Registration**
8:30 AM	-	**5:00 PM**	**Yosemite**	**Exhibits Open**
9:00 AM	-	**12:50 PM**	**Technical Sessions**	
9:00 AM	-	12:50 PM	Grand A	11 | MT | NAND & DRAM
9:00 AM	-	12:50 PM	Grand B	12 | ALT | Oxide Semiconductor and 2D Channel Device
9:00 AM	-	12:00 PM	Cont 1-3	13 | SMB | Physical, Chemical, and Multimodal Integrated Sensors
9:00 AM	-	12:00 PM	Cont 4	14 | MS | Memory, Device and System-Technology Co-Optimization
9:00 AM	-	12:25 PM	Cont 5	15 | NC | Probablistics and In-Memory Computing for Robotics and Optimization
9:00 AM	-	12:50 PM	Cont 6	16 | EDT | Novel Computing Paradigms
TUESDAY *(continued)*				

9:00 AM	-	12:00 PM	Cont 7-9	17 | FS | Semiconductors for Quantum Computing
9:00 AM	-	12:50 PM	Imperial A	18 | RSD | Reliability of Ferroelectric Materials and Devices
9:00 AM	-	12:25 PM	Imperial B	19 | PMA | Advances in RF, mmWave, and Sub-THz III-V and Wide-Bandgap Transistor Technologies for 6G and Beyond
10:15 AM	**-**	**11:15 AM**	**Yosemite**	**Coffee Break with Exhibitors**
2:00 PM	**-**	**5:50 PM**	**Technical Sessions**	
2:00 PM	-	5:25 PM	Grand A	20 | MT | Nonvolatile Memories
2:00 PM	-	5:50 PM	Grand B	21 | ALT | BEOL, Thermal and Packaging
2:00 PM	-	5:00 PM	Cont 1-3	22 | SMB | MEMS Resonators and Transducers
2:00 PM	-	5:00 PM	Cont 4	23 | MS | Atomic Simulation: Devices, Circuits, and Process
2:00 PM	-	5:25 PM	Cont 5	24 | NC | LLM Acceleration Using In-Memory Compute
2:00 PM	-	5:50 PM	Cont 6	25 | EDT | Oxide-Channel Devices
2:00 PM	-	5:25 PM	Cont 7-9	26 | ODI | Emerging Photonic Devices
2:00 PM	-	5:25 PM	Imperial A	27 | RSD | Component-to-System Level Reliability and Thermal Management
2:00 PM	-	5:50 PM	Imperial B	28 | PMA | Advanced GaN Power Devices: Power Delivery, Bidirectional and Low Voltage
3:15 PM	**-**	**4:15 PM**	**Yosemite**	**Coffee Break with Exhibitors**
8:00 PM	**-**	**10:00 PM**	**Cont 4-5**	**EP | Rewriting the Rules of Semiconductor Manufacturing: Why the Old Manufacturing Model May Not Be Enough**
WEDNESDAY				
8:30 AM	**-**	**3:00 PM**	**East Lounge**	**Registration**
9:00 AM	**-**	**11:30 AM**	**Yosemite**	**Exhibits Open**
9:00 AM	**-**	**12:50 PM**	**Technical Sessions**	
9:00 AM	-	12:50 PM	Cont 1-3	30 | PMA | High-Performance RF FETs and Switches
9:00 AM	-	12:25 PM	Cont 4	31 | NC | Edge Intelligence and On-Chip Learning
9:00 AM	-	12:25 PM	Cont 5	32 | EDT | Ferroelectric Devices
9:00 AM	-	12:50 PM	Cont 6	33 | ALT | Advanced Device and Process Innovations
9:00 AM	-	11:35 AM	Cont 7-9	34 | MS | Multiscale, AI & Advanced Simulation Methodologies
9:00 AM	-	12:50 PM	Imperial A	35 | RSD | Memory Technologies and Compute-in-Memory
9:00 AM	-	12:25 PM	Imperial B	36 | ODI | Advanced Image Sensors
10:15 AM	**-**	**11:15 AM**	**Yosemite**	**Coffee Break with Exhibitors**
2:00 PM	**-**	**5:25 PM**	**Technical Sessions**	
2:00 PM	-	5:25 PM	Cont 4	37 | MT | Oxide Semiconductor Memories
2:00 PM	-	5:25 PM	Cont 5	38 | EDT | Emerging Memories
2:00 PM	-	5:00 PM	Cont 6	39 | FS | System-Level Optimization Combining Logic, Memory and Packaging
2:00 PM	-	5:00 PM	Cont 7-9	40 | MS | Emerging Devices, Ferroelectrics & Nonvolatile Memory
2:00 PM	-	5:00 PM	Imperial A	41 | RSD | Reliability of Next-Generation Transistors
2:00 PM	-	5:00 PM	Imperial B	42 | SMB | Smart Sensors and Bio-Integrated Interfaces
3:15 PM	**-**	**4:15 PM**	**East Lounge**	**Coffee Break**

Program

Saturday, December 12, 2026

Tutorials

The tutorials are in their sixteenth year and are standalone presentations on specialized topics taught by world-class experts. These tutorials will provide a brief introduction to their respective fields and facilitate understanding of the technical sessions. In contrast, the traditional short courses are focused on a single technical topic.

Co-Chairs: Uygar Avci, Intel and Kang-ill Seo, IBM
1:30 PM - 6:00 PM

Tutorial 1: Evolution of the Transistor from Planar to Nanosheet and Beyond

Larry Bair, AMD
1:30 PM– 2:50 PM
Continental 1-5
Chair: Uygar Avci, Intel

This talk will trace the evolution of the logic transistor over four decades of CMOS scaling, from the planar MOSFET era shaped by Dennard scaling, through strained silicon and high-k/metal gate innovations, to the FinFET transition that restored electrostatic control at deeply scaled dimensions, and onward to today's gate-all-around nanosheet architecture. The presentation will examine the device-physics drivers behind each architectural shift, including short-channel control, parasitic resistance and capacitance, variability, power density, and manufacturability, and will discuss how these factors reshaped both transistor design and process integration. Building on that foundation, the tutorial will explore the next stages of scaling, including fork sheet concepts, vertically stacked complementary FETs (CFETs), backside power delivery, and the growing role of design-technology co-optimization. It will conclude with a forward-looking perspective on the opportunities and challenges of post-silicon channel option.

Tutorial 2: Chiplet, Heterogeneous Integration, 3D Stacking Technologies Principles and State of the Art

Oliver Noblanc, STMicroelectronics
1:30 PM– 2:50 PM
Continental 6-9
Chair: Uygar Avci, Intel

With CMOS technology scaling reaching some physical and economic limits the high-performance computing industry has introduced new technologies aiming to keep on offering improved performances to the product design community leveraging on new front-end process and assembly capabilities modifying the frontiers between the 2 words. This trend intending to "de-integrate" is a real technological breakthrough after more than 60 years leveraging on CMOS integration driving to system on chip product architectures. Showing some practical hardware developments highlighting spectacular process capabilities and product performances serving the field of chiplet product architecture. On the other hand, derivative technologies such as CMOS image sensors and silicon photonics, 3D-stacking has entered high-volume manufacturing since several years with technology features that can be extended to other application domains. The rational of the move from 2D to 3D will be explained and technology principles and examples will be given mainly leveraging on STMicroelectronics own experience. On top of process features, CAD capabilities have also to be developed and offered by EDA vendors enabling 2.5 and 3D product designs. Here again some developments with EDA vendors will be covered covering a set of tailored 3D capabilities.

Tutorial 3: Memory Technologies and System Co-Optimization in the AI Era: Past, Present and Future
Fatih Hamzaoglu, Intel
3:05 PM– 4:25 PM
Continental 1-5
Chair: Uygar Avci, Intel

Memory Wall has been a popular term lately in the AI Era, due to criticality of Data Storage and movement in Training and Inference workloads. Low latency, high Bandwidth, Low Power and Cost are all crucial to scale AI workloads to achieve the optimal Tokens/sec/$. SRAM and DRAM/HBM have been the main workhorses to provide the memory needs, but new solutions such as High Bandwidth Flash (HBF) are being introduced. This tutorial will go through the Memory subsystem, system level integration, scaling from past to present and how future memory technology candidates are being proposed to push the boundaries of Memory Wall in AI.

Tutorial 4: AI in Physics-based Semiconductor Modeling: From PDE-constrained Learning to Agentic TCAD/DTCO Workflows
Pratheep Balasingam, Synopsys
3:05 PM– 4:25 PM
Continental 6-9
Chair: Uygar Avci, Intel

This tutorial presents a physics-grounded perspective on the application of artificial intelligence (AI) and machine learning (ML) across semiconductor modeling workflows, ranging from feature-scale process and device simulation to block-scale design technology co-optimization (DTCO). Anchored in governing PDEs and numerical methods, the tutorial highlights how ML augments, rather than replaces, physics-based solvers through surrogate modeling, operator learning, and solver acceleration. Applications of ML models to (a) device and process optimization, (b) compact model construction and extraction, and (c) stochastic variability are discussed with emphasis on physical consistency and robustness. The integration of these techniques into DTCO and design closure flows, for faster exploration of design-technology tradeoffs, is then examined. Finally, emerging agentic AI approaches for orchestrating end-to-end TCAD/EDA workflows are outlined.

Tutorial 5: Diverse Memory Technologies for Adaptive Computing: From Device Physics to Cooperative Architectures
Elisa Vianello, CEA-Leti
4:40 PM– 6:00 PM
Continental 1-5
Chair: Uygar Avci, Intel

Large-scale AI systems have reinforced a memory roadmap centered on DRAM and HBM scaling for workloads dominated by capacity and bandwidth. However, adaptive computing applications impose additional constraints: persistent local states across idle phases, energy-efficient retraining, low-standby inference, probabilistic exploration and low-latency transfer between data with different lifetimes. In these regimes, neither non-volatility nor analog computing is needed everywhere, but both become valuable when matched to the right data and algorithmic roles. This tutorial presents a device-level perspective on diverse memory technologies for adaptive in-memory computing. Rather than seeking a universal memory, we discuss how heterogeneous memories can cooperate, each operating where its physical properties are most useful. The focus is on the principles that make such cooperation meaningful: separating update and inference roles, exploiting device physics as computing primitives, reducing unnecessary data conversion, and turning memory-to-memory interfaces into mechanisms for transfer, adaptation, sampling or annealing. The tutorial will position this approach with respect to DRAM/HBM-centric systems, review the opportunities and limits of analog in-memory computing, and discuss integration paths from 2.5D packaging to hybrid bonding and monolithic 3D. Selected experimental examples will illustrate how cooperative memories can support adaptive learning, probabilistic computing and optimization, while highlighting practical challenges in endurance, variability, precision, tiling, calibration and system-level mapping.

Tutorial 6: Advanced Optical Devices and Wafer-Scale Assembly for Next-Generation CPO and Optical I/O

Joris Vancampenhout, imec
4:40 PM– 6:00 PM
Continental 6-9
Chair: Uygar Avci, Intel

We will discuss recent R&D advances in silicon-integrated modulators and photodetectors, targeting both upcoming 400G/lane "Fast-and-narrow" Pluggable Optics and Co-Packaged Optics. Next, we will cover active optical devices designed for future "Wide-and-Slow" approaches enabling high-density optical I/O with ultra-low power consumption, eventually targeting chip-to-chip interconnects. We will elaborate on advanced EIC-PIC assembly, packaging and co-design and outline our vision for ubiquitous, sub-pJ/bit wafer-level optical interconnects.

Sunday, December 13, 2026

Short Courses

IEEE IEDM will offer two short courses with in-depth coverage of highly relevant topics from world experts.

Co-Chairs: Geert Eneman, imec and Kang-ill Seo, IBM
1:30 PM - 6:00 PM

Short Course 1: Logic and System Innovations Powering the AI Era

9:00 AM, Continental 1-5
Organizer: Marko Radosavljević, Notre Dame University

9:00 AM - 10:15 AM
SC1-1 | Distributed Network Design for AI: Evolution, Scaling, and System Implications, Brad Beckmann, AMD
As AI systems scale from a single die to packages, boards, racks, and eventually warehouse-scale deployments, the dominant challenge increasingly shifts from compute alone to efficient data movement across every level of the system. This talk begins with a retrospective on how data-movement bottlenecks have evolved within the SoC and package, and how AMD's architecture, SoC design, and advanced packaging strategies have adapted in response to rising bandwidth, latency, and energy demands. It then examines how modern AI workloads continue to expand these requirements, motivating a broader shift from a traditional component-centric view toward a system-level perspective in which interconnect, packaging, memory, and communication fabrics must be co-optimized. Building on that foundation, the talk frames the key design spaces of modern scale-in, scale-up, and scale-out systems, including the growing role of offload data-mover engines in sustaining performance and efficiency at scale. Finally, it discusses the opportunities and tradeoffs of extending optical communication more deeply into AI platforms, highlighting why device, packaging, and interconnect innovations are becoming central to overall system capability.

10:30 AM – 11:45 AM
SC1-2 | Advanced Logic Transistor Scaling and 3D Architectures in the Angstrom Era, Seung Hoon Sung, Intel
The relentless demand for energy-efficient computing, driven by AI and high-performance applications, continues to push logic technology beyond conventional scaling limits. The transition from planar FETs to FinFETs, and now to gate-all-around (GAA) transistors, has enabled substantial improvements in performance, power, area, and cost. However, extending this trajectory into the Angstrom era requires a holistic rethinking of device architecture, process integration, materials innovation, and design-technology co-optimization. This course reviews key enablers for advanced logic scaling, including RibbonFET (GAA) architecture, backside power delivery networks (BS-PDN), gate stack and contact integration, strain engineering, advanced patterning, and variability and reliability challenges. It further discusses emerging pathways toward 3D logic architectures, including complementary FET (CFET), and highlights the critical process-integration challenges required to translate these concepts into manufacturable technologies.

12:45 PM - 2:00 PM

SC1-3 | Edge Device Performance/Power Enhancement in Mobile SOC Platform Through Device to DTCO to STCO, Jun Yuan, Qualcomm

With the rapid growth of AI, the rising demand for real-time, intelligent applications has made edge AI computing a critical enabler of next-generation user experiences. This discussion reviews how process technology evolution—through innovative transistor architectures (from planar to FinFET to 2nm GAA) and continuous scaling—has historically driven improvements in mobile SoC PPAC (power, performance, area, and cost). As Moore's Law faces increasing challenges & slowing down, we examine the role of design-technology co-optimization (DTCO) in sustaining its momentum and explore future opportunities with system-technology co-optimization (STCO) to further enhance PPAC in meeting future AI/Computing requirements.

2:00 PM – 3:15 PM

SC1-4 | Atomic Layer Processing: Applications in Advanced Semiconductor Devices, Thorsten Lill, LAM

Monolithic 3D integration is one of the vectors to advance device density of logic and memory integrated circuits. 3D integration requires vertical and lateral processing. In addition to deposition of tall stacks with alternating layers and high aspect ratio (HAR) etching, manufacturing of these structures will require a combination of lateral deposition and etching to achieve what can be called lateral film stacking. Lateral film stacking can be achieved by area selective ALD or alternating ALD and selective isotropic etching processes. Among the isotropic etching processes, ALE stands out as a method which enables etching of traditionally "hard to etch" materials with good etch control and top to bottom etch rate uniformity. We will discuss the mechanisms and performance of radical etching, continuous thermal etching and thermal ALE. One of the important figures of merit of the combined lateral deposition and etching processes is 3D edge-placement-error. We will introduce application examples for lateral stacking in advanced logic and memory devices and discuss the challenges the industry is facing.

3:30 PM – 4:45 PM

SC1-5 | Reliability Challenges in Advanced Packaging and System-Level Interaction, Yao-Chun Chuang, TSMC

Advanced packaging enables the high-performance computing essential for next-generation AI, but it also increases system complexity. As form factors expand and power densities rise, new system-level reliability challenges emerge—specifically regarding thermal management and thermomechanical stress. This tutorial discusses the various stages of package qualification and illustrates how TSMC's component-level expertise can be leveraged to mitigate these system-level risks.

4:45 PM - 6:00 PM

SC1-6 | Cryogenic Electronics for Quantum Computing, Loïck Le Guevel, Google

Stable, accurate, and precise signal delivery is at the heart of state-of-the-art superconducting quantum processors with tens of quantum bits. In today's flagship architectures, the experimental setup—comprising cryogenics, wiring, filtering, and room-temperature electronics—is as critical as the design of the processor itself. Small disturbances or distortions of the control signal lead to catastrophic errors when executing quantum algorithms. As quantum processors grow, targeting more than 100,000 qubits in the next decade, scaling becomes a central challenge. The experimental setup increases in complexity and size, leading to larger signal distortions and poorer delivery. Moving electronics from room temperature to cryogenic environments has the potential of reducing wiring, thereby mitigating distortions and shrinking system size. However, new challenges arise, including stringent power constraints due to watt-level cooling capacities and unmodeled cryogenic physics, such as mobility shifts and more pronounced thermal effects.

Short Course 2: Powering Next-Generation AI and HPC with Advanced Memory Technologies

9:00 AM, Continental 6-9

Co-Chairs: Nanbo Gong, IBM

9:00 AM - 10:15 AM

SC2-1 | Memory-Centric Computing for the Generative AI Era, Damien Querlioz, Universite Paris-Saclay

Generative AI is changing the meaning of in-memory computing. For more than a decade, the field has been motivated by the mapping of neural-network weights to dense matrix-vector multiplications in memory arrays. This

remains essential, but it is no longer sufficient. Modern AI systems are dominated by several memory objects with very different lifetimes: persistent model weights, transient activations, session-dependent KV caches, retrieval memories. This short course will build a memory-centric view of modern AI, contrasting cloud inference, where bandwidth, HBM capacity, batching and context management dominate, with edge inference, where energy, latency, privacy and non-volatile storage matter more. We will then examine how new model directions (long-context transformers, KV-cache compression, grouped or latent attention, state-space models, mixture-of-experts, and hybrid architectures) reshape the opportunity for volatile and non-volatile memories. The course will identify where classical analog and digital in-memory computing is powerful, where near-memory computation is more realistic, and where device innovation is needed. The goal is to define a roadmap toward energy-efficient generative AI in which memory is no longer only where data resides, but where intelligence is executed, stored, adapted and remembered.

10:30 AM – 11:45 AM

SC2-2 | HBM Design and Technology Challenges for AI Generation, HyunBae Lee, SK Hynix

The rapid advancement of AI has driven an exponential surge in data usage, placing unprecedented demands on memory subsystems. While High Bandwidth Memory (HBM) was engineered for high-throughput data processing, it faces escalating development challenges. From a design perspective, these hurdles stem not only from transistor performance limitations but also from interconnect constraints across block-to-block, chip-to-chip, and system-to-system levels. Consequently, traditional design methodologies focused solely on performance are shifting toward approaches that simultaneously optimize power efficiency and performance. To address this, HBM development has begun incorporating advanced logic processes for base dies. This short course will examine the design challenges HBM faces in this evolving landscape, discuss the limitations of current advanced logic processes for HBM applications, and explore future technical directions.

12:45 PM - 2:00 PM

SC2-3 | Oxide-Semiconductor Channel Transistor DRAM (OCTRAM): High-Density and Low-Power Consumption DRAM Technology for Sub-10nm Nodes and Beyond, Keiji Ikeda, Kioxia

In recent years, the scaling of conventional DRAM beyond the sub-10nm node has encountered significant barriers, primarily due to challenges in process integration and electrical performance. To address these issues, DRAM technology leveraging the oxide semiconductor Indium Gallium Zinc Oxide (InGaZnO) has garnered considerable attention. This innovative material exhibits ultra-low off-state leakage current characteristics, which are superior to those of traditional silicon, contributing to a reduction in power consumption, especially in future large-capacity DRAM. Moreover, the structural transformation of DRAM cell transistors into a three-dimensional architecture such as 4F2 vertical channel transistors and 3D DRAM is achievable using InGaZnO as the channel material, thanks to its depositable nature. These advancements offer substantial opportunities for progress in DRAM technology, focusing on achieving higher bit density and lower power consumption through architectural modifications driven by cell transistor innovations. In this presentation, we will explore the potential for technological breakthroughs in DRAM by implementing three-dimensional structuring of cell transistors using InGaZnO. This approach leads to the development of a novel DRAM technology termed Oxide-Semiconductor Channel Transistor DRAM (OCTRAM), aimed at achieving future high-density and low-power consumption DRAM technology.

2:00 PM – 3:15 PM

SC2-4 | The Evolving Role of NAND: Technology and System Scaling from Traditional Storage to Memory-Storage Convergence, Kishore Muchherla and Akira Goda, Micron

As artificial intelligence (AI) reshapes computing and datacenter architectures, the demands placed on storage systems are evolving well beyond traditional storage system specifications. NAND flash, long the cornerstone of the storage hierarchy, now faces new system-level requirements driven by AI applications and cost-performance scaling pressure for traditional applications. At the same time, the boundary between memory and storage is blurring, calling for a fresh look at how NAND expands into the broader system hierarchy. In this talk, we will begin by revisiting how traditional NAND architecture characteristics map to vital storage system metrics. Next, we will examine how traditional storage system scaling (density, performance, and power) interacts with NAND technology scaling; explore how AI workloads are reshaping storage requirements in AI servers and discuss how these emerging demands interact with underlying storage technologies. Finally, we will highlight the expanding role of NAND beyond

traditional bit storage — evolving from passive storage medium towards an active participant in the converging memory-storage hierarchy.

3:30 PM – 4:45 PM
SC2-5 | Spintronics for MRAM and Beyond, Shunsuke Fukami, Tohoku University
Spintronics is a field that harnesses both the charge and spin degrees of freedom of electrons to realize novel physical phenomena and advanced device functionalities. While long established in applications such as magnetic sensors, spintronics has more recently evolved into a cornerstone of next-generation memory technology, most notably through the advancement of Spin-Transfer Torque Magnetoresistive Random Access Memory (STT-MRAM). This short course will provide an overview of recent breakthroughs in spintronics, with a focus on their integration into MRAM and emerging computational paradigms. Scaling challenges and device innovations in Spin-Transfer Torque MRAM (STT-MRAM) are first discussed. The session then transitions to Spin-Orbit Torque (SOT) mechanisms, examining the materials science and switching principles required to realize high-speed SOT-MRAM. Finally, the horizon of "beyond-memory" applications is explored, specifically focusing on probabilistic computing. By utilizing a probabilistic bit (p-bit) with a low-stability magnetic tunnel junction, spintronics offers a unique hardware platform to generate high-quality random streams, providing a promising path toward addressing complex computational problems in the AI era with high efficiency. Through this overview, the current R&D landscape is summarized, and a strategic outlook on the future of spintronic technologies in the semiconductor industry is provided.

4:45 PM - 6:00 PM
SC2-6 | Chiplet & Advanced Packaging Technologies for AI & HPC Systems, John Knickerbocker, IBM
Energy efficient AI data centers to support high performance computing, large language models and 500 billion parameter model algorithms require scaling high bandwidth density electrical and optical interconnections. The solutions take advantage of next generations of heterogeneous integration chiplet technology solutions using advanced packaging technologies. The need to manage large amounts of data for AI training and AI inference applications requires innovations for more efficient architectures, software and hardware solutions to support AI workloads and data. New architectures, digital twin models and simulations along with co-design for chiplet, 3D stacking and co-packaged interconnections are needed to support both scale out and scale up improvements in bandwidth density, energy efficiency and distance between chips to overcome bottle necks in compute, memory, and network communication links. This chiplet and advanced packaging technology review will address elements of importance including aspects of (1) Architecture / EDA / co-design ; (2) chiplets, packaging and components fabrication and test; (3) Chiplet and advanced package assembly, (4) Module level and rack level integration, power delivery, integration compatibility, thermal solutions and optical / electrical integration and (5) Reliability, serviceability and characterization.

Monday, December 14, 2026

Keynote Session

Session 1: Welcome and Awards Presentations

9:00 AM, Grand Ballroom B
Co-Chairs: Meng-Fan (Marvin) Chang, National Tsing Hua University and TSMC and Jungwoo Joh, Texas Instruments

Keynote Papers

Technical Program Chair: Jungwoo Joh, Texas Instruments

9:45 AM, Grand Ballroom B

1-1 AI Driven Semiconductor Paradigm Shift: Optimizing the Semiconductor Industry from Device Scaling to System Integration (Invited)

Naga Chandrasekaran, Executive Vice President – Chief Technology and Operations Officer – General Manager, Intel Foundry

Artificial Super Intelligence (ASI) demands unprecedented technology gains in performance, bandwidth, energy efficiency, and scale. This paper presents a full-stack semiconductor roadmap spanning advanced transistor architectures, backside power delivery, next-generation interconnects, 3D integration, and optical connectivity. It shifts the focus from isolated component optimization to workload-aware co-optimization of compute, memory, power, thermal management, manufacturing, and reliability. This approach supports scalable and sustainable AI platforms by improving performance per watt, reducing data movement, and accelerating manufacturable innovation.

10:30 AM, Grand Ballroom B

1-2 Driving AI Memory Innovation: From Device Technology to AI Memory Platform (Invited)

Yunik Son, Vice President of Device Technology Solution, SK Hynix

As AI workloads continue to scale, overall system performance is increasingly determined by the interplay between compute and memory performance. This drives demand for high bandwidth, power efficiency, and reliability. Limits in cell scaling have accelerated the transition from 2D to 3D memory architectures, with AI memory competitiveness increasingly driven by Peripheral (Peri) transistor technology. Wafer bonding technology is reshaping memory architecture by decoupling Cell and Peripheral wafer fabrication. This separation enables performance improvement and independent optimization of Peri technology. However, this shift alone does not guarantee successful innovation. Memory systems remain constrained by cost, high-voltage operation, pitch-limited design rules, leakage control, and reliability. The increasing complexity of HBM and future 3D architectures also introduces thermal management challenges. Consequently, the industry is shifting from adapting logic technologies to developing memory-specific HKMG, RMG, and FinFET solutions tailored to these constraints. Advances in memory-optimized design methodologies are equally imperative. Because memory technology and design are developed concurrently, accurate pre-layout modeling is essential for rapid design exploration. Memory design also continues to evolve from full-custom toward hybrid custom-digital methodologies. Building on this foundation, memory DTCO and future system-technology co-optimization (STCO) frameworks can extend optimization across device, circuit, package, and system levels. The convergence of Peri device innovation and memory-native design methodologies establishes a new AI Memory Platform paradigm. This enables workload-specific memory solutions built upon a common technology foundation, transforming memory from a commodity into a strategic enabler of AI systems.

11:15 AM, Grand Ballroom B

1-3 Technology Trends in the IDM Semiconductor Business (Invited)

Fabio Gualandris, President, Quality, Manufacturing & Technology, STMicroelectronics

Digitalization and advanced packaging have been reshaping integrated device manufacturer (IDM) processes in recent years. Digitalization connects front-end modeling, device and circuit simulation, manufacturing analytics, and system-level validation in a continuous flow, accelerating learning and optimization. In parallel, advanced packaging, including panel-level packaging and wafer-to-wafer hybrid bonding, enables heterogeneous integration by combining the most suitable technologies to maximize system performance. Building on these enablers, ST is pursuing three complementary technology trends: 1) silicon photonics for AI server infrastructure; 2) computing, from AI-enabled microcontrollers (MCUs) to quantum systems; and 3) sensing and actuation, based on imaging, MEMS, and power technologies, to connect the physical and digital worlds. Full automation and robotics in manufacturing complement a new approach to R&D conducted in a fully virtual environment and supported by ST's diversified technology portfolio. The IDM structure enables design- technology co-optimization and greater system-level flexibility.

Session 2: MT | Ferroelectric Memories

1:45 PM, Grand Ballroom A

Co-Chairs: Kyunghwan Lee, Samsung and Sunghil Lee, Tokyo Electron

The Ferroelectric Memories session highlights key breakthroughs in high-density integration, reliability engineering, and low-power operations for next-generation computing architectures. The first paper is demonstrating a chip-level 176-layer 4D Poly-Si FeNAND that solves block erase limitations while achieving a 32% energy reduction for AI-oriented applications. The second paper is introducing a 5-bit ferroelectric–ionic hybrid FET that utilizes polarization barrier engineering and time-scale separation to eliminate pass-disturb and secure >1E7 second retention. The third paper is presenting a hidden domain-state pre-switching technique that suppresses pass-disturb by up to 87% to enable robust TLC operation in 3D Fe-NAND strings. The fourth paper is unveiling a measurement-driven inverse reconstruction method to resolve hidden internal state trajectories and switching mis-alignments in MFMIS FeFETs. The fifth paper is presenting a BEOL-compatible monolithic 3D 1T-1FeFET technology that bridges SRAM-class latency with DRAM-class density for high-endurance on-chip weight buffers. The last paper is demonstrating an ultra-high electric field cycling method that induces low-thermal-budget orthorhombic crystallization in sub-4 nm HZO films for sub-0.5 V operation. Together, these six presentations bridge fundamental thin-film physics with advanced array-level architectures, offering vital insights for researchers and engineers developing low-power, high-performance non-volatile memories.

1:50 PM

2-1 Chip-Level Demonstration of 176-Layer 4D Ferroelectric NAND Towards High-Bandwidth and Ultra-Low-Power Operations

Taehyung Park, Sangchul Oh, Hyunjeong Kim, Jongil Kim, Gap Sok Do, Hyera Shin, Wonjun Lee, Woohyung Choi, Hyeongjin Lee, Dae Hyun Kim, Woocheol Lee, Inseon Yeo, Youngjae Kwon, Myoungsub Kim, Seungjun Lee, Hamin Sung, So Young Yim, Jeongsoo Kim, Jong-Young Lee, Min Yong Lee, Kyoungsu Choi, Chihyun Park, Su Jin Chae, Won-Tae Koo, Hyung Dong Lee, Keonsoo Shim, Hyunkyu Kim, Sungyong Lim, Sungchul Shin, Ju Yeab Lee, Sanghyun Oh, Jum Soo Kim, Seho Lee, Soo Gil Kim, Seiyon Kim, Jaeyun Yi, Youngmann Cho, Seonyong Cha

SK Hynix

This paper demonstrates the chip-level operation of 176-layer 4D Ferroelectric NAND (FeNAND) designed for high-speed and ultra-low-power consumption. While previous Oxide Semiconductor (OSC) channel-based FeNAND studies proposed power reduction through Pass Voltage (V_{PASS}) decrease, they faced limitations in Block Erase efficiency due to low hole mobility. We propose a Poly-Si channel-based FeNAND structure compatible with conventional NAND processes, enabling efficient Block Erase. Through Gate Stack engineering, specifically optimizing the mid-ferroelectric and channel-side interlayers, we achieved Erase V_{th} < 1V and Memory Window > 6 V. As a result, reducing V_{PASS} from 6 V to 3 V is projected to achieve a 32% reduction in Energy per Bit (EpB), demonstrating the feasibility of Poly-Si FeNAND for low-power AI-oriented NAND applications.

2:15 PM

2-2 A 5-bit Ferroelectric–Ionic Hybrid FET with >10^7 s Retention and Disturb-Free Operation toward 3D V-NAND via Polarization-Engineered Asymmetric Barrier and Time-Scale Selective Switching

Jongseon Seo[1], Laeyong Jung[1], Junghoon Park[2], Hojung Jang[1], Jiho Kim[1], Jangseop Lee[1], Wanki Kim[2], Daewon Ha[2], Hyunsang Hwang[1]

[1]Pohang University of Science and Technology (POSTECH), [2]Samsung Electronics

Pass-disturb and limited retention hinder ionic memories in serial V-NAND strings. We demonstrate a 3D-stackable ferroelectric–ionic hybrid FET that addresses both without array overhead. Electrically injected ions modulate threshold voltage (V_{th}), while an embedded ferroelectric (FE) layer serves two functions. Its polarization asymmetrically modulates the ionic migration barrier by 0.2 eV—lowering it during programming to allow ion injection and raising it afterward to suppress back-diffusion—enabling >10^7 s retention at room temperature. In addition, the ~1000 × time-scale separation between FE switching (~ns) and ionic transport (~μs) allows a short inhibit pulse to reverse polarization only in unselected cells, screening the V_{PASS}-induced electrolyte field. Pass-disturb V_{th} shift is thereby reduced from 680 to 21 mV, consistent with measurements and TCAD. The device also

achieves an ~8.1 V memory window, 5-bit operation, and > 10^7 cycle endurance with < 3% window loss, supporting high-density, disturb-free 3D V-NAND.

2:40 PM

2-3 Hidden Domain-State Engineering for Pass-Disturb Suppression in FeFET-Based 3D NAND Flash Memory

Sheung Hun Kim, Yungjun Kim, Dae Hyun Kang, Youngkeun Park, Dongkyu Lee, Byung Jin Cho
Korea Advanced Institute of Science and Technology (KAIST)

As 3D NAND scales to hundreds of stacked cells, a higher pass voltage (V_{pass}) is required for reliable reads. In ferroelectric-FET (FeFET) based 3D NAND (3D Fe-NAND), the intrinsic vulnerability of FeFETs intensifies unintended threshold voltage (V_{TH}) shift under high V_{pass}, making pass-disturb suppression a critical prerequisite for multi-bit operation. We propose hidden domain-state engineering via a pre-switching operation, which controls internal domain configurations that are not easily apparent from electrical readout. This suppresses disturb in both regimes: (1) in the high-V_{TH} regime, negative ΔV_{TH} from partial polarization switching is suppressed by pre-switching the low-field-switchable domains; and (2) in the low-V_{TH} regime, positive ΔV_{TH} from charge trapping is counterbalanced by partial switching of intentionally preserved switchable domains. As a result, disturb could be reduced by up to 87%, achieving disturb tolerance over 10^7 V_{pass} cycles in triple-level cell (TLC) operation — which has remained challenging for 3D Fe-NAND.

3:30 PM

2-4 Transistor Switching, Ferroelectric Switching, and Charge Injection in MFMIS FeFET: Reconstruction of Internal State Trajectory

Hongkai Ning, Yanjie Shao, Jing Kong, Jesús A. del Alamo
Massachusetts Institute of Technology

The MFMIS FeFET is regarded as a versatile, reliable device technology for both memory and AI computation. To date, its internal switching dynamics remain elusive, largely because the floating-gate node and charge evolution are hidden from terminal measurements. Here, we introduce a measurement-driven inverse reconstruction method that experimentally maps out and cross-validates the internal trajectories of voltage (V), capacitance (C), and charge (Q). Counterintuitively, our result reveals a gross misalignment between ferroelectric switching (FS) and transistor switching (TS): the ferroelectric capacitor can be programmed while the transistor is off and erased while it is on. In addition, charge injection through the ferroelectric is found to play a crucial role in the memory window (MW) establishing a tradeoff between memory performance and reliability.

3:55 PM

2-5 Monolithic 3D 1T-1FeFET Technology Bridging SRAM-Class Latency and DRAM-Class Density for Weight Buffer Memory

Hyeonwoo Park, Sharadindu Gopal Kirtania, Emmanuel Quezada, Kiseok Lee, Ishaan Joshi, Chengyang Zhang, Hyun Jae Lee, Eknath Sarkar, Asif Khan, Shimeng Yu, Suman Datta
Georgia Institute of Technology

This work demonstrates a backend-of-line (BEOL)- compatible, monolithic 3D BiCS-inspired 1T-1FeFET memory architecture, spanning experimental fabrication, device physics, compact modeling, and array-level benchmark. A two-tier amorphous oxide semiconductor (AOS) channel MOSFET and FeFET-based hardware achieves single-shot program/erase (P/E) operations within the 20-ns measurement limit at logic compatible voltages (+1.0/-1.2 V). A circuit-compatible model captures width-dependent abrupt percolative switching by coupling ferroelectric (FE) switching kinetics, self-consistent electrostatics, AOS transport, and a hysteresis-dependent percolation threshold correction. Parasitic-aware six-tier array simulations project 1.5-ns P/E and sub-2-ns read latency. Crucially, we identify array-level erase-disturb and sneakcurrent limits and successfully mitigate them through a combination of co-design and a precharge-and-float sensing scheme. Based on extracted unit-cell dimensions, a 16-tier array reaches 0.575 Gb/mm^2, approaching DRAM density. Combined with > 10^{11}-cycle endurance at 85 °C, this architecture bypasses SRAM capacity scaling walls to enable on-chip weight buffer memory.

4:20 PM

2-6 Ultra-High Electric Field (UHEF) Enhanced Ferroelectric Crystallization of Scaled HZO Film for High P_{sw} (> 50 μC/cm²) and Low Voltage (< 0.5 V)

Minjong Lee[1], Dushyant M. Narayan[1], Soham Shirodkar[1], Jongchan Kim[1], Dan N. Le[1], Thi Thu Huong Chu[1], Doo San Kim[1], Esther H. R. Tsai[2], Yugang Zhang[2], Chang-Yong Nam[2], Pil-Ryung Cha[3], Seungbin Lee[4], Hyeonhong Min[4], Gwanghyeon Jang[4], Youngbae Ahn[5], Seung Wook Ryu[5], Kyeongjae Cho[1], Si Joon Kim[4], Jiyoung Kim[1]
[1]University of Texas at Dallas, [2]Brookhaven National Laboratory, [3]Kookmin University, [4]Kangwon National University, [5]SK Hynix

This study presents a promising method to induce ferroelectric orthorhombic (O)-phase crystallization in scaled HZO films using ultra-high electric field (UHEF) bipolar cycling. The process achieves strong ferroelectricity ($2P_r$ > 50 μC/cm²) in sub-4 nm HZO films with a minimized thermal budget. UHEF cycling promotes direct phase transformation from the amorphous state to the ferroelectric O-phase, and this effect is further boosted by tuning pulse amplitude and substrate temperature during cycling. Joule-heating energy (U_{J-H}) and field-induced energy barrier lowering are identified as key driving forces behind the crystallization process. This approach enables 2.5 nm-equivalent HZO capacitor operation, achieving >50 μC/cm² P_{sw} at sub-0.5 V. These findings establish a thermally controlled and scalable route toward low-power ferroelectric for logic-compatible memory integration.

Session 3: ALT | Platform and DTCO

1:45 PM, Grand Ballroom B

Co-Chairs: Lin Hu, Nvidia and Kwan-Yong Lim, Qualcomm

Continued logic scaling remains the foundational enabler for advances in modern computing platforms. This session features five papers highlighting the latest progress in logic technology and Design-Technology Co-Optimization (DTCO). First, AMD presents scaling challenges and innovations in Si process technology and system level integration. Next, TSMC showcases its 2nd generation nano-sheet A14 technology, demonstrating full-node PPA benefits for High-Performance Computing (HPC)/AI and mobile SoC applications. The third paper, from TSMC, focuses on SRAM DTCO through targeted enhancements in pattern fidelity and variation control on 2nm technology platform, enabling robust stability during high-speed operation to meet stringent requirements of HPC in the AI era. Kookmin University presents a comprehensive study of the PPA and thermal benefits of backside clock and signal routing in GAA technology using a BSPDN process. Lastly, Samsung introduces an EUV-less 8nm technology platform with a native 3.3V I/O transistor and automotive-grade eMRAM, emphasizing its role in cost-sensitive, reliability-critical next-generation SoCs across automotive, AI, and robotics domains. Together, these papers provide a cohesive view of how process, device, and DTCO innovations are advancing logic scaling and enabling the next wave of computing capabilities.

1:50 PM

3-1 The Physics Bill is Due: Power, Thermal, and Mechanical Challenges for the Next Decade of AI

Bill En
AMD

AI capabilities have advanced dramatically — from early chatbots to sophisticated reasoning agents — driven by sustained, year-on-year improvements in silicon performance per watt. Maintaining that trajectory is becoming harder. As AI workloads grow in scale and complexity, the gains required to sustain annual performance improvements must increasingly be found in how power delivery, thermal management, and mechanical constraints are solved at the silicon level. Realizing the full potential requires a fundamental shift in methodology: power, thermals, and mechanical integrity must be treated as first-class inputs — co-optimized from the outset alongside transistor performance, compute architecture, and memory bandwidth.

2:15 PM

3-2 A14 CMOS NanoFlex™ Pro Platform Technology with 2nd Gen. Nano-sheet and Full Node PPA Benefits for HPC/AI and Mobile SoC Applications

Shien-Yang Wu, Chih-Hao Chang

Taiwan Semiconductor Manufacturing Company (TSMC)

A14 NanoFlex™ Pro platform technology features the world's smallest SRAM with a cell size <0.017μm^2 and highest SRAM macro density to date through dimensional ground rule scaling. Coupled with standard cell innovations, it delivers a full-node PPA improvement over the predecessor N2 technology for a 10~15% speed enhancement, a 25~30% power reduction and a chip density increase by ~20% driven by logic and SRAM scaling. Key features, such as TSV, innovative RDL that enables SoIC bond pitch of 4.5μm and high-performance MiM processes, have also been developed to accelerate the system-level integration and scaling for SoIC products. Supported by robust yield and device progress, A14 technology is on track for volume production in 2028.

2:40 PM

3-4 Nanosheet SRAM DTCO for Enhanced Speed, Vmax, and Vmin at 2nm platform technology and beyond

YuKuan Lin, Ping-Wei Wang, Ping-En Cheng, Zi-Yuan Wei, Chia-Hao Pao, Chris Lin, Kian-Long Lim, Sean Huang, Feng-Ming Chang, Alex Lin, Chia-Cheng Chang, George Liu, Alex Chen, Sen-Nien Chung, Tzung-Hua Lin, Jian-Hao Chen, Szu-Ping Lee, Hung-Chi Wu, Tzu-Hua Chiu, Chien-Tai Chan, Hua Tai Lin, Tommy Chen, Tsung-Yung Jonathan Chang, Huiling Shang, Yao Ching Ku, Shyue Shyh Lin, Geoffrey Yeap

Taiwan Semiconductor Manufacturing Company (TSMC)

This paper presents a novel SRAM design and technology co-optimization (DTCO) that achieves more than 100 mV increase in the maximum operable supply voltage (V_{max}), faster SRAM speed and better minimum operable supply voltage (V_{min}). Realization of these enhancements requires overcoming critical challenges in pattern fidelity, process variation and designing the optimized SRAM array. A 7% increase in cell current results in faster SRAM performance by optimizing the effective nanosheet width through process/SRAM-layout optimization. These performance gains are achieved with no compromise in yield or V_{min}. This DTCO has already been integrated into TSMC's mass production 2nm platform technology (N2), enabling robust stability during high-speed operation to meet the stringent requirements of high-performance computing (HPC) in the AI era.

3:30 PM

3-5 Beyond Backside Power: Backside Clock and Signal Routing for PPA Improvement in Angstrom Nodes

Junjong Lee[1], Mingyu Park[2], Sanguk Lee[3], Seungchul Song[4], Ukjin Roh[4], Ming Cai[4], David Greenlaw[4], Yoonjung Lee[1], Sungjin Choi[1], Daehwan Kim[1], Rock-Hyun Baek[3], Sungkyu Lim[2]

[1]Kookmin University, [2]University of Southern California, [3]Pohang University of Science and Technology (POSTECH), [4]Google LLC

For the first time, we develop a fully backside (BS)-routed clock and a novel BS signal routing scheme using direct BS contacts for the Angstrom nodes. The fully BS-routed clock tree achieves lower clock skew and fewer buffer counts compared to prior nano-TSV-based clocking. Combining the fully BS-routed clock with BS intra-cell routing significantly improves worst negative slack (WNS), power consumption, area utilization, and max temperature. The BS signal inter-cell routing further improves WNS by relocating only a small number of nets to the BS. By jointly applying three BS schemes—BS clock and BS signal intra-/inter-cell routing—we achieve a 20.3 % improvement in power-delay product (PDP). Furthermore, these BS routing technologies are successfully applied to 3DIC and provide PPA improvements.

3:55 PM

3-6 Advanced EUV-less 8nm technology platform for cost-effective edge AI applications

Junho Yang, Tae Young Lee, Seeui Hong, Junghun Choi, Seokgyu Lee, Narae Shin, Sicheon Seong, Jong-Ik Lee, Hanjin Im, Dongwook Yang, Young-sik Ham, Kangduck Lee, Sehong Kim, Gyoungkuk Min, Seungteck Lee, Jangwoong Uhm, Juchan Lee, Joonbum Park, Hyunjung Shin, Jiho Kim, Seungju Hwang, Tae-Hyung Kim, SangShin Han, Changhoon Sung, JinJu Kim, Dongjin Han, Yongsang Jeong, Youngmok Kim, Kiseok Suh, Shinhee Han,

Shigenobu Maeda, Kangho Lee
Samsung Electronics
This paper presents the 2nd-generation 8nm technology platform (8LPU+) tailored for cost-effective automotive and robotics applications. 8LPU+ achieves 12% performance enhancement, 15% power reduction, and 8% area shrink over the previous 8nm generation (8LPU), while reducing stand-by leakage through the introduction of ULL-SRAM and cHVT devices. For edge AI applications, where total chip size is governed not only by logic density but also by I/O and eNVM area, native 3.3V planar I/O devices are seamlessly integrated into the 8LPU FinFET platform, delivering additional ~10% I/O area reduction. 8LPU is fully qualified for automotive grade-1 (AG1) and offers AG1 eMRAM with an industry-leading bitcell size of 0.017 μm^2 (~20Mb/mm^2), providing substantial eNVM area benefit. Together, these I/O and eMRAM area competitiveness establishes 8LPU as a highly area-competitive platform for edge AI applications such as AI MCU and automotive ZCU.

Session 4: SMB | Optical, Radiation, Quantum Sensing

1:45 PM, Continental 1-3
Co-Chairs: Taeko Ando, Ritsumeikan University and Ming-Huang Li, National Tsing Hua University
This session brings together emerging optical, radiation, and quantum sensing technologies, ranging from advanced imaging sensors to intelligent and quantum-enabled sensing systems. A Sn-Pb perovskite photodetector array was monolithically integrated on a 640 × 512 CMOS readout chip, enabling broadband 400–1100 nm imaging with a 186.8 dB linear dynamic range. A dual-gate a-Si:H TFT was developed to combine optical and touch sensing in a single pixel using only standard display manufacturing, enabling a flexible 500-ppi, 256 × 256 electronic-skin array. A wafer-to-wafer integrated microsystem was developed to perform signed analog multiply-accumulate and spike encoding within the sensor, achieving 14 fJ per 3 × 3 feature event for neuromorphic vision. A group at Tsinghua University developed a perovskite/CMOS-integrated X-ray image sensor by leveraging large-area perovskite deposition, high-speed response, an ASIC architecture, and BEOP (back-end-of-package) tiling technology. Another X-ray image sensor was realized by bonding a photodiode chip with a 650-μm-thick, high-resistivity n-type Si X-ray-sensitive layer to a readout integrated circuit (ROIC) chip. An intelligent diamond quantum sensor was developed through monolithic integration of an optical filter, a phototransistor, and RRAM, enabling in- and near-sensor processing.

1:50 PM
4-1 A Monolithically Integrated Sn-Pb Perovskite/CMOS Visible-to-Near-Infrared Image Sensor Achieving 17.79 lp/mm Spatial Resolution
Chaoyu Ren, Zhizhong Wang, Shuo Yang, Guoxiang Zhang, Hao Xu, Yuangan Chen, Mingzhu Hu, Yiran Hao, Lijuan Huang, Gangsen Su, Sihui Yang, Hao Li, Yan Shi, Fan Xu, Xunyong Yang, Fang Xiang, Qinghai Tan, Qin Hu
University of Science and Technology of China (USTC)
We demonstrate a novel Sn-Pb hybrid perovskite/CMOS broad-spectrum image sensor covering visible-to-near-infrared (NIR) wavebands (400-1100 nm). The CMOS readout chip integrates a 640×512-pixel array with an active area of 9.6×7.68 mm and a pixel pitch of 15 μm. Monolithic integration of the perovskite photodetectors with the CMOS chip is achieved through integrated low-temperature fabrication processes. The fabricated detector delivers a linear dynamic range (LDR) up to 186.8 dB and a specific detectivity of 3.5×10^{12} Jones. Our perovskite image sensor achieves excellent uniformity and 17.79 lp/mm spatial resolution, showing strong potential for broadband visible/NIR sensing in intelligent sensor microsystems, human-machine interaction, and environmental monitoring.

2:15 PM
4-2 A 500-ppi 256×256 Flexible Active-Matrix Multimodal Sensing Array Fully Compatible with Industry-Standard Dual-Gate TFT Process
Zikang Mei[1,2], Hao Gong[2], Jun Li[2], Zibo Li[2], Xiao Peng[2], Wei Tang[2], Xiaojun Guo[2]
[1]National Key Laboratory of Advanced Micro and Nano Manufacture Technology, Shanghai Jiao Tong University, [2]Shanghai Jiao Tong University
Flexible high-resolution multi-modal sensing arrays are critical for human-machine interaction and embodied intelligence. Though mature large-area thin-film transistor (TFT) manufacturing lines from display industries offer mass-production merits, extra steps and special sensing materials largely offset such advantages. This work proposes a dual-gate (DG) TFT structure fully compatible with industrial a-Si:H technology to integrate optical and

touch sensing, utilizing two physically distinct transduction pathways. A pixel statistical signal processing method is devised to decouple multi-source currents. Based on this, a flexible 500-ppi 256×256 dual-gate a-Si:H TFT sensing array, monolithically integrated with a gate-driver-on-array (GOA) circuit, is fabricated to identify four interaction modes and serve as a robotic electronic-skin prototype, establishing a scalable display-TFT route to high-resolution multimodal interactive surfaces.

2:40 PM

4-3 W2W-Integrated Intelligent Image Sensor with Reconfigurable In-Pixel Signed MAC and Direct Spike Encoding

Seungyeob Kim[1,2], Seungmin Cheon[2], Julian Oh Hyunkwen Yang[2], Sanghun Jeon[2]
[1]EE, Korea Advanced Institute of Science and Technology (KAIST), [2]Korea Advanced Institute of Science and Technology (KAIST)

We demonstrate a wafer-to-wafer-integrated intelligent optical microsystem that combines polarity-configurable photodetection, reconfigurable in-sensor feature extraction, and direct spike encoding. A vertically structured polarity-configurable photodetector generates symmetric positive and negative photocurrents, enabling signed analog computation with a responsivity of 0.25 A/W and a linearity of . The detector array is vertically integrated with a poly-Si complementary TFT-based spike enabler, eliminating intermediate analog-to-digital conversion and off-sensor data transfer. Signed weights are configured through the detector-bias polarity, while their magnitudes are programmed by duty-cycle-controlled charging. The integrated microsystem experimentally performs reconfigurable 3×3 signed analog multiply–accumulate operations using Sobel X, Laplacian, and emboss kernels and directly converts the resulting feature signals into frequency-encoded spikes. A dark-reference cell further compensates for temperature-dependent dark-current drift. The proposed sensing–computing–encoding microsystem achieves a projected device-core energy of 14 fJ per 3×3 feature event, demonstrating an integrated sensor platform with embedded intelligence for energy-efficient neuromorphic vision.

3:30 PM

4-4 A 32.5-cm^2 FOV, 2000-FPS Scalable Perovskite/CMOS Integrated X-Ray Imager Enabled by Material-to-System Co-Optimization

Zi Wang[1], Jia-He Zhang[1], Guan-Hua Dun[1], Hexuan Wang[1], Shuyuan Yang[1], Ruolong Zhou[1], Yida Gao[1], Yuhan Wang[1], Jia-Li Peng[1], Xiang-Shun Geng[1], Yuanyuan Li[1], Zhen-Peng Shi[1], Jia Zhang[2], Yuan-Jie Cao[2], Chong-Shen Song[2], Peiyan Cao[2], Yi Yang[1], Dan Xie[1], Tian-Ling Ren[1]
[1]Tsinghua University, [2]Shenzhen Xpectvision Technology Co., Ltd

We demonstrate a perovskite-on-CMOS X-ray imager that breaks the area–speed trade-off, achieving 2000-FPS full-frame imaging over a 32.5-cm^2 field of view (FOV) through material–device–circuit–system co-optimization (M2S-CO). A water-based, annealing-free process enables room-temperature large-area perovskite deposition, while the device with graded band alignment yields fast 9.70/12.74-μs rise/fall times. A scalable readout ASIC architecture combines in-pixel dark-current compensation and charge digitization with intra- and inter-die parallel readout. Back-end-of-package integration realizes a 2×4 tiled imager with 1152×576 pixels and 3-pixel-pitch physical inter-die gaps. At 2000 FPS, the imager maintains 7.03-lp mm^{-1} resolution and a contrast-to-noise ratio (CNR) of 7, enabling scalable, large-area, high-speed X-ray imaging.

3:55 PM

4-5 Cu-Cu 3D-Stacked Si Direct-Conversion X-ray CMOS Image Sensor for Low-Noise Energy-Resolved Imaging

Kazunobu Ota[1], Takaki Hatsui[2], Hikaru Iwata[1], Ryu Kamibaba[1], Hiroki Tojinbara[1], Katsunori Hiramatsu[1], Masataka Kawasoe[3], Atsushi Suzuki[1], Atsuhiro Ando[1], Sakiko Nishida[1], Yasuyuki Sato[1], Katsunori Tozawa[1], Hiroaki Ishiwata[1], Ryo Hayashibara[3], Kenji Moriyama[1], Yoshiaki Honjo[2], Kyo Nakajima[2], Kyosuke Ozaki[2], Yuji Matsuda[2], Toru Shirakata[1], Shinji Miyazawa[1], Yasushi Tateshita[1]
[1]Sony Semiconductor Solutions Corporation, [2]Riken, [3]Sony Semiconductor Manufacturing Corporation

A Cu-Cu 3D-stacked Si direct-conversion X-ray CMOS image sensor for energy-resolved imaging was developed. Under-pixel readout reduced input capacitance, while small anodes, dark-current separation, charge-collection control, and high-voltage biasing enabled low-dark-current, low-noise X-ray imaging. The sensor demonstrated energy-resolved X-ray imaging with elemental identification and high-definition image acquisition.

4:20 PM

4-6 Monolithic 3D Integration of Optical Filter, Phototransistor and RRAM for In-/Near-Sensor Computing in Diamond NV Quantum Sensing

Zhiyuan Du, Madhav Gupta, Ruibin Mao, Feng Xu, Bofan Yang, Haozheng Wan, Zemeng Lin, Jizong Jiang, Rongchen Qin, Wendi Li, DongKeun Ki, Ngai Wong, Zhiqin Chu, Can Li
The University of Hong Kong

Diamond NV quantum sensing offers unrivalled sensitivity, but conventional systems record fluorescence frame-by-frame and transmit GB-scale raw data off-chip, imposing minute-level latency and a bulky bench-top setup. We monolithically integrate in 3D, an optical filter, a programmable-responsivity phototransistor array for in-sensor spatiotemporal processing of NV signals, and a TaOx RRAM crossbar on a 180-nm CMOS platform for near-sensor in memory classification, completing the sensing-to-classification chain. A tunable pixel implements an in-sensor SG filter that differentiates the ODMR spectrum, accelerating sensing by 14× with minimal precision loss, further demonstrated on dynamic current tracking. Reconfigured as an in-situ Fourier transformer, it further resolves three frequency components from the harmonic-mixed signal. Coupled with an RRAM perceptron, it decodes neural-signal for RNS epilepsy detection, reaching 98.3% seizure-classification accuracy (vs 98.9% software), cutting latency by 14× and energy by > 10^5 versus an EMCCD+GPU pipeline. This work turns diamond-quantum-sensor into an intelligent edge terminal.

Session 5: NC | Neural Network Inference with Emerging Technologies

1:45 PM, Continental 4

Co-Chairs: Ming-Hsiu (Eric) Lee, Macronix and Laura Bégon-Lours, ETH Zurich

This session focuses on advances at the materials and integration level for the acceleration of neural networks. It covers six papers addressing the monolithic integration of different technologies and the development of innovative solutions the inference of neural networks using emerging memories. In one example by Fudan university, magnetic random access memories (MRAM) are used for two workloads within a transformer: static-weight linear projections and dynamic-weight self-attention. Another approach followed by two papers propose the co-integration of different technologies: first, a 4-tiers Monolithic 3D (M3D) integration of Si CMOS logic, RRAM, DRAM and photosensor array is demonstrated. This heterogeneous chip from Tsinghua University supports neural network computation, cache and image input. Second, a 2-tier monolithic 3D ferroelectric thin-film transistor array with 3-bit/cell is demonstrated at Seoul National University, based on an amorphous semiconducting oxide technology. Towards beyond-CMOS circuits, two papers propose transistor-less implementation: on one hand, efforts on electrode materials at KAIST allows ferroelectric tunnel junctions to exhibit large dynamic ranges and rectifying effects for selector-less arrays. On the other hand, bipolar junction transistors are proposed by Politecnico di Milano for the realisation of 1B1R arrays with phase-chang memory resistive weights for high-density and drift-resilient in-memory computing. Finally, Fudan University demonstrates 4-bit/weight analog CiM operation with oxide-based DRAM array on a flexible substrate and under elevated temperature.

1:50 PM

5-1 First Demonstration of VCMA-MRAM-based Attention Engine with Unified Static/Dynamic IMC and Write-Error-Tolerant Spiking Attention

Siyuan Ouyang[1], Zheng Zhu[2], Albert Lee[2], Xinxu Lin[1], Zhicheng Li[1], Yifan Gao[1], Wenkui Zhang[1], Chin-Chung Chen[2], Di Wu[2], Hao Jiang[1], Xumeng Zhang[1], Xianzhe Chen[1], Qi Liu[1], Ming Liu[1]
[1]Fudan University, [2]Suzhou Inston Technology Co., Ltd.

Transformer attention requires nonvolatile storage for fixed projection weights and frequent low-energy updates for input-dependent K/V matrices. Existing in-memory accelerators typically rely on heterogeneous memories or support only one of these workloads. Here, for the first time, we present a VCMA-MRAM in-memory attention engine that unifies static-weight linear projections and dynamic-weight self-attention on a homogeneous nonvolatile substrate. The fabricated macro achieves >10^{12} endurance, 0.55 ns switching, and 30 fJ/bit write energy, enabling frequent K/V updates while retaining static weights without standby power. To avoid the latency and energy overhead of write verification, we co-design a write-error-tolerant spiking attention scheme. The system achieves 75.6% accuracy on CIFAR-100 and 96.2% on DVS128 Gesture, while device-aware SpikeLLM-7B simulation reaches

a perplexity of 12.57 on Wikitext-2. The engine achieves an energy efficiency of 0.27 pJ/SOP and > 30×power consumption reduction than CPU/GPU platforms at comparable latency.

2:15 PM

5-2 Monolithic 3D Integration of 3D-Stacked Dual-Gate IGZO 2T0C DRAM Cache-CIM for Edge-Cloud Collaborative Embodied AI

Kexin Wang[1], Yibei Zhang[1], Jianshi Tang[1], Zhen Zhu[1], Queran Pan[1], Xingchu Li[1], Yanbo Su[1], Zhidong Tang[1], Yijia Fan[1], Yanqing Li[2], Heyi Huang[2], Qingtian Zhang[1], Yuyan Wang[1], Bin Gao[1], He Qian[1], Huaqiang Wu[1]
[1]Tsinghua University, [2]Institute of Microelectronics of Chinese Academy of Sciences

To deploy embodied AI, compact hardware must support for large-model interaction, fast on-chip decisions, and dynamic K/V cache updates. Here, we present M3D-SCIS, a 4-tiers M3D chip integrating Si CMOS logic, RRAM-based CIM, IGZO 2T0C DRAM array-based CIM, and IGZO-FET Sensor array for edge-cloud collaborative embodied AI. Using BEOL-compatible process below 300 ℃, M3D-SCIS integrates control logic, local MLP computation, dynamic weight storage/cache-CIM, and image input on a single chip. In particular, for the first time, we realize a vertically aligned 3D-stacked 2T0C DRAM array in the 3rd-tier, implemented as two-sub-tier of read/write DG IGZO-FETs with L_{ch} = 150 nm. 3D stacking and DG control enhance gate modulation and storage-node capacitance, enabling 5-bit/cell capability, retention >300 s, and DRAM-based MVM. Finally, in the benchmark task with π_0 model , M3D-SCIS enables overlapped local inference, dynamic cache update, and edge-cloud collaborative decision making, improving energy efficiency by 16.6 times compared to GPU.

2:40 PM

5-3 Experimental Demonstration of a 2-Tier Monolithic 3D FeTFT Array with 3-b/cell Analog MAC and Full-Stack Evaluation for CNN and MoE-LLM Inference

Hyun-Jun Kang[1], Soo Hyung Lee[2], Jae-Hyun Lee[1], Seon-Jin Yoo[1], Jiabao Ye[2], Soohyeon Kim[2], Wei D. Lu[2], Joon-Kyu Han[1]
[1]Seoul National University, [2]University of Michigan

We present a monolithic 3D (M3D) analog compute-in-memory (CiM) platform built on BEOL-compatible amorphous oxide semiconductor (AOS) channel ferroelectric TFTs (FeTFTs). Gate-stack and channel co-optimization yields a 3.5 V memory window, eight well-separated V_{th} states for 3-b/cell weight storage, linear potentiation and depression (R^2 > 0.999), 10^5-cycle endurance, and 10-year retention. A 2-tier NOR-type array fabricated at or below 400℃ shows no tier 1 degradation, tight array-level uniformity, disturb-free selective write, and accurate 56-level analog MAC (R^2 > 0.99) in both tiers. A CiM core with two 128×128 3-b/cell tiers with 28 nm periphery is evaluated by full SPICE simulation including parasitics, σV_{th}, and ADC nonidealities. A hybrid 1-b+3-b INT4 mapping halves the cell count and doubles VGG-8 energy efficiency to 88.5k img/J. The 3D core achieves 152 pJ/MVM, 6.1 Mb/mm^2, 4.22 TOPS/mm^2, 646 TOPS/W, a 0.66× footprint compared with 2D, and MoE-LLM energy efficiency ~10^3 beyond an H100 GPU.

3:30 PM

5-4 A Selector-Free 3-Bit FTJ Crossbar with Giant TER (>30,000) and Record-High RR (>5,000) Enabled by Multifunctional MoO_2 Bottom Electrode for In-Memory Inference

Paul Gallegos[1], Geonhyeong Kang[2], Yongsu Kim[2], Sohyeon An[2], Jae-Hyeon Cho[2,3], Seokwon Lee[2], Yubin An[2], Sangmin Park[2], Sanghyun Park[3], Suhwan Lim[3], Kwangyou Seo[3], Wanki Kim[3], Daewon Ha[3], Mincheol Shin[2], Jinho Ahn[4], Sanghun Jeon[2]
[1]School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), [2]Korea Advanced Institute of Science and Technology (KAIST), [3]Samsung Electronics, [4]Hanyang University

We demonstrate a highly stable, self-rectifying multi-level ferroelectric tunnel junction (FTJ) using a multifunctional MoO_2 bottom electrode for selector-free in-memory inference. The MoO_2 electrode simultaneously resolves sneak-current leakage and multi-level instability through dual functionality: (i) its high work function (5.53 eV) induces a large built-in field enabling self-rectification, and (ii) its oxygen-reservoir action passivates interfacial oxygen vacancies, reducing the HfO_{2-x} sub-oxide fraction from 6.7% to 3.4% and stabilizing the ferroelectric orthorhombic phase. The optimized FTJ achieves TER of 30,900, rectifying ratio of 5,651, 8 analog conductance states (3 bits/cell), endurance exceeding 10^8 cycles, and retention beyond 10 years at 85℃. A fabricated 12×12 selector-free crossbar projects read-margin scaling to 6,553 cells. Hardware-based inference simulations using measured 8-level

statistics achieve 95.3% MNIST accuracy, outperforming 3-bit SRAM (82.3%), positioning MoO_2 FTJ as a practical candidate for high-density 3D weight-storage in edge intelligence.

3:55 PM

5-5 Drift-resilient multilevel PCM technology with BJT selector for accurate, energy-efficient, high-density analog in-memory computing

Giandomenico Panettieri[1], Stefano Giangrasso[2], Danlo Bondì[2], Jakub Andrzej Jastrzebski[2], Lorenzo Sali[2,3], Fabio Carletti[2], Andrea Al Muktash[2], Piergiulio Mannocci[2], Matteo Baldo[3], Giulia Samanni[3], Andrea Motta[3], Enrico Gomiero[3], Andrea Giovanni Bonfanti[2], Daniele Ielmini[2]

[1]Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano and IU.NET, Milano, Italy, Politecnico di Milano, [2]Politecnico di Milano, [3]STMicroelectronics

Analog in-memory computing (AIMC) based on phase change memory (PCM) is a promising solution for edge AI acceleration; however, drift and variations severely limit the compute accuracy. Here, we present the characterization of a PCM technology in 28 nm fully-depleted silicon-on-insulator (FD-SOI) with bipolar junction transistor (BJT) selector. Compared to conventional MOS-selected one-transistor/one-resistor (1T1R) array, the one-BJT/one-resistor (1B1R) array displays multilevel cell (MLC) operation with intrinsic drift linearization, a 3.6× reduction in conductance variability, a 2× reduction of cell area and high precision at low read currents. Near-software accuracy is demonstrated for ResNet-9, ResNet-18, ResNet-34 and transformer-based ALBERT models under low energy consumption, establishing the 1B1R cell as a strong solution for high-density, drift-resilient AIMC.

4:20 PM

5-6 Flexible Ln-IZO DRAM-Based Analog Compute-in-Memory Array with Computing Accuracy Robust to Mechanical Strain and Temperature

Tingpan Lan[1], Junyu Hou[1], Mingxun Li[1], Zhiwei Chen[1], Dongzi Yang[1], Xumeng Zhang[1], Xianzhe Chen[1], Min Li[2], Lei Zhou[2], Chenxin Zhu[3], Zhongguang Xu[3], Ming Wang[1], Qi Liu[1], Ming Liu[1]

[1]Fudan University, [2]Guangzhou New Vision Optoelectronic Technology Corporation, [3]University of Science and Technology of China (USTC)

For the first time, we experimentally demonstrate a highly robust flexible 32×32 analog compute-in-memory (CiM) array based on 4T1C lanthanum-doped indium zinc oxide (Ln-IZO) DRAM cells. Using a ratio-coded readout scheme, the CiM array effectively suppresses computation errors induced by intrinsic device variations and mechanical deformation. At the cell level, computation error remains below 0.1% under bending with a 2-mm radius. At the array level, parallel vector-matrix multiplication achieves integral nonlinearities of 1.99, 2.04, and 2.82 LSB flat state and bending radii of 10 and 5 mm, respectively. The CiM array is further deployed to execute electromyography (EMG) recognition for clinical myopathy assessment, achieving a recognition accuracy of 92.1% even under bending with a 5-mm radius. This work establishes an effective approach for mechanically resilient, high-accuracy flexible analog CiM.

Session 6: EDT | 2D Module Engineering and Heterogenous Integration

1:45 PM, Continental 5

Co-Chairs: Albert Cheng, TSMC and Gaurav Thareja, Applied Materials

This session highlights rapid progress in two-dimensional electronics, spanning fundamental device engineering to system-level integration. The first half focuses on breakthrough 2D module technologies, including defect-healed dielectric stacks for near-ideal WSe_2 transistor switching, advanced gate-stack engineering for low-voltage operation, record-low-resistance contacts and high-performance CFETs, and heterogeneous electro-photonic integration using 2D materials. The second half showcases scalable heterogeneous integration strategies, featuring monolithic 3D MoS_2/Si computing platforms, dual-sided CMOS enabled by backside 2D transistor integration, area-selective self-aligned semimetal contacts, and thermally robust BEOL-compatible contact modules. Together, these works demonstrate a compelling pathway toward energy-efficient, highly integrated, beyond-silicon computing systems.

1:50 PM

6-1 Record-Performance Monolayer WSe_2 FETs with ~120 Ω·μm R_C Enabling 28-nm-node CGP 2D CFET

Ni Yang[1], Haofei Zheng[2], Yu-Cheng Lu[3], Chengdong Yao[2], Lin-Yun Huang[4], Yu-Ming Chang[2], Chenyang Li[2], Haomin Liu[2], Baoshan Tang[4], Wanying Li[2], Ji Lu[2], Yi Wan[2], Kah-Wee Ang[2], Vita Pi-Ho Hu[3], Lain-Jong Li[2], Jing-Kai Huang[4]
[1]NEXSTROM PTE. LTD., [2]National University of Singapore, [3]National Taiwan University, [4]Nexstrom Pte. Ltd.

In this work, we systematically optimize CVD-grown monolayer (1L) WSe_2 FETs through process engineering, achieving enhancement-mode operation with a positively shifted threshold voltage (V_{th}) and a record-low contact resistance of ~120 Ω·μm with record high on current (I_{on}) for ~1.25 mA/μm (V_{ds} = -1V). Based on the optimized pFET, we further demonstrate a pitch-scaled 2D CFET featuring vertically stacked 28-nm pFET and nFET with a common gate (CG). The well-matched pFET/nFET pair enables subthreshold logic operation, achieving an ultralow static power consumption of 21 pW at V_{dd} = 0.75 V, a maximum voltage gain of 42.5 V/V, and a high static noise margin (SNM) of ~90% at V_{dd} = 1.0 V. Although the demonstrated performance has not yet reached that of state-of-the-art silicon technologies

2:15 PM

6-2 CVD WSe_2 p-FETs with SS = 60 mV/dec and Record-High $g_{m,max}$ = 882 μS/μm via Defect-Healed HfO_2 Gate Stack

Zhenghang Yu[1], Yihao Zhao[1], Yifan Liu[1], Zhiwei Tian[1], Tiancheng Gu[2], Bei Zhao[2], Zhiyong Zhang[1], Yongfeng Wang[1], Jianfeng Jiang[1], Peng Wu[1]
[1]Peking University, [2]Southeast University

Interface traps caused by defects in high-k dielectric severely limit the gate efficiency and V_{DD} scaling in 2D p-FETs. In this work, we demonstrate high-performance CVD WSe_2 p-FETs using a dielectric healing strategy, in which the oxygen vacancy defects in the HfO_2 gate dielectric are healed by an optimized dual-oxidation PEALD process and a post-deposition SF_6/O_2 plasma treatment. Across optimized devices, we achieve a near-ideal SS ≈ 60 mV/dec, a record-high peak transconductance $g_{m,max}$ = 882 μS/μm and a maximum I_{ON} = 870 μA/μm at V_{DS} = -1 V, while retaining a large I_{ON} = 500 μA/μm at a scaled contacted poly pitch (CPP) of 59 nm. Multi-frequency C-V measurements further yield a minimum interface trap density (D_{it}) well below 10^{12} $cm^{-2}eV^{-1}$ after defect healing.

2:40 PM

6-3 Reducing Supply Voltage in WSe_2-Channel pFET Devices by Gate Stack Engineering

Tien Dat Ngo[1], Goutham Arutchelvan[2], Daire Cott[1], Benoit Van Troeye[1], Fabian Ducry[1], Kaustuv Banerjee[1], Pierre Morin[1], Giacomo Callori[1], Olivier Richard[1], Kris Paulussen[1], Sui-An Chou[2], Ming-Yang Li[2], Mark van Dal[2], Cesar Javier Lockhart de la Rosa[1], Iuliana P. Radu[2], Bogdan Govoreanu[1]
[1]imec, [2]Taiwan Semiconductor Manufacturing Company (TSMC)

We report on dual-gate (DG) WSe_2 pFETs with improved off- and on-state through gate stack and channel engineering. While most publications focus on increasing maximum current, often at the cost of degraded off current (at V_g=0V), here we develop solution that increases current density at 1V gate voltage overdrive (V_{ov}) without significant impact on off-state. We compare two gate stacks: one based on native oxide (WO_x formed by plasma assisted oxidation of WSe_2) and second one using ALD Al_2O_3 as nucleation layers for HfO_2 high-k. Post-deposition O_2 annealing is found to be crucial for performance with both gate stacks. The gate stack using ALD Al_2O_3 enables widest current range reduction in subthreshold swing. This enables the current density at lowest gate voltage I_{on}=200μA/μm at V_{ov} of 1V, and over 500 μA/μm, at V_{ov} of 2V, for on/off ratio well over 10^5 in 50 nm channel length DG devices.

3:05 PM

6-4 WSe_2/Lithium Niobate Dual-Type Electro-Optic Modulator Array for Bipolar Weight Encoding, Bidirectional Propagation in MRM-Based PNNs

Tiancheng Yu[1,2], Shi Zhao[2], Quanzhen Wan[2], Chenxuan Zeng[2], Jianan Li[2], Haofei Zheng[2], Jie Yang[2], Evgeny Zamburg[2], Aaron Voon-Yew Thean[2]
[1]Department of Electrical and Computer Engineering, National University of Singapore, [2]National University of Singapore

We report a mixed electro-optic modulator array that performs forward propagation, transpose-mode backpropagation, with bipolar weight encoding within a single photonic neural network (PNN). The BEOL-

compatible, low-thermal-budget platform integrates lithium-niobate (LN) micro-ring modulators (MRMs) [1], WSe_2 electro-absorption modulators (EAMs) [2]. With a dual-MRM cell, we propose a new reciprocal optical routing array that allows the same micro-ring matrix to implement both W and W^T weight vectors, eliminating a duplicated transpose path, while differential MRMs support bipolar weight encoding and multiplication. We show forward and reverse paths with matched tuning efficiencies of 8.94 and 8.99 pm/V and 0.12-dB crossing loss. The WSe_2 EAMs operate uniformly from 1540 to 1560 nm with 6.64-mW/V modulation efficiency. System modeling predicts 100.76-TOPS throughput, 7.75-TOPS/mm^2 compute density, and 1.61×/10.4× higher area efficiency than dual-bank MRM/MZI-mesh PNNs. The low-temperature process supports 3D BEOL integration with interposers and chiplets.

3:55 PM

6-5 3-Tier TMD/Si-FinFET Heterogeneous Integration Enabling Memorizable Logic Primitives towards On-Chip Hyperdimensional Computing Solutions

Jin-Kai Lin[1], Li-Hsuan Li[2], I-Min Liu[2], Pin-Yuan Wang[2], Yu-Lin Li[2], Bo-Wei Yang[2], Ying-Hao Chu[2], Chih-Chao Yang[3], Chao-Hui Yeh[2]

[1]Institute of Electronics Engineering, National Tsing Hua University, [2]National Tsing Hua University, [3]Taiwan Semiconductor Research Institute (TSRI)

In this work, we experimentally demonstrate and develop a heterogeneous MoS_2/Si monolithic 3D (M3D) integration platform on 8-inch wafers, realizing complementary FET (CFET) cells and memory-in-logic architecture. MoS_2 transistors on/off current ratios exceeding 10^{10}, and ultra-low standby power of ~10^{-14} W/μm. A scalable three-tier M3D architecture along with transistor-level monolithic inter-tier vias (MIV) is established to orchestrate vertically integrated logic and embedded-memory building blocks, including ring oscillators (ROSC), XOR and full-adder circuits, newly designed 6T static ramdon access memory (SRAM) with compact 3T-cell macrocache architectures, for multifunction, energy-efficiency, and local data movement in modern ICs. This work further introduces a first-ever memorizable XNOR (mXNOR) logic as a similarity-check operation cell to accelerate in-memory updates to hypervector encodings. providing a promising hardware solution for future on-chip hyperdimensional learning and edge AI applications.

4:20 PM

6-6 First Experimental Demonstration of Dual-sided CMOS with Frontside FinFETs and Backside 2D FETs by Backside Monolithic Integration

Yanbang Chu[1], Chenyang Cui[2], Zhentao Xiao[1], Hua Yu[3], Haoyang Cheng[3], Chenxu Wang[2], Jieying Liu[2], Hongyue Du[3], Jianxiang Jin[1], Yalun Sheng[2], Xinyue He[1], Wan Wei[2], Weihai Bu[1], Jin Kang[1], Ming Li[1], Runsheng Wang[1], Guangyu Zhang[2], Ru Huang[1], Heng Wu[1]

[1]Peking University, [2]Institute of Physics, Chinese Academy of Sciences, [3]Songshan Lake Materials Laboratory

Here we report, for the first time, a successful demonstration of the compact dual-sided CMOS platform by backside monolithic 3D integration, uniting frontside (FS) advanced Si FinFETs and BS 2D FETs with ultra-scaled contact pitch (CP) down to 65 nm—in both *dual-sided CFET* (FS pFET + BS nFET) and *dual-sided full CMOS* (FS CMOS + BS CMOS) configurations. Multiple logic gates were realized with inverter voltage gain ~ 23 V/V @ V_{dd} = 1 V, functional V_{dd} down to 0.2 V, **rail-to-rail** response at 100 ns input transition time. A selenium assisted MoS_2 transfer technique was developed on the wafer backside for device enhancement. To mitigate the key challenge of voltage level mismatch across technology nodes and material platforms, a novel *3D level shifter* was also first demonstrated based on the dual-sided CMOS as the adaptive interface of BS M3D integration, successfully converting voltage level with sharp step > 1 V.

4:45 PM

6-7 Area-Selective Sb(01$\bar{1}$2) Contacts on MoS_2 by Elevated-Temperature Electron-Beam Evaporation

Eun-Kyu Lee, Huije Ryu, Minsu Seol, Joung Eun Yoo, Min Seok Yoo, Alum Jung, Eunji Yang, Dongmin Kim, Yongsung Kim, Kyung-Eun Byun, Changhyun Kim

Samsung Advanced Institute of Technology (SAIT)

Self-aligned source/drain (S/D) formation via selective epitaxial growth has been a critical technology in Si CMOS. This approach has not been realized for 2D transistors due to the absence of an area-selective metal deposition process for 2D channels. We report area-selective growth of semi-metallic Sb on MoS_2 by electron-beam evaporation (EBE) at elevated temperature. Sb grows selectively on exposed MoS_2, with no detectable deposition

on adjacent dielectric or resist surfaces. We attribute this selectivity to Sb's high vapor pressure, which drives rapid re-evaporation from surfaces lacking the favorable $(01\bar{1}2)$ registry with MoS_2. The film is $(01\bar{1}2)$-preferential, the orientation known to minimize contact resistance on MoS_2. We form S/D contacts through oxide vias exposing clean MoS_2 and demonstrate back-gated MoS_2 transistors with quasi-Ohmic Sb contacts (R_C ~370 Ω·μm). This EBE-based, lift-off-free route provides a fab-friendly, area-selective contact and points toward self-aligned, all-around contacts for 2D gate-all-around transistors.

5:10 PM

6-8 BEOL-Compatible Nickel-Bismuth Alloy Contacts for High-Performance MoS_2 n-FETs

Shu-Jui Chang[1], Yi-Fan Ho[2], Ke-Hsien Lin[2], Sin-Yue Lee[1], Shin-Yuan Wang[2], Zih-Syuan Huang[3], Wei-Sheng Yun[1], Ying-Mei Yang[4], Tse-Ming Chen[4], Chueh-Cheng Yang[5], Chia-Hsin Wang[5], Wei-Yen Woon[1], Chih-I Wu[3], Wen-Hao Chang[2], Chao-Hsin Chien[2], Tsung-En Lee[2], Edward Yi Chang[2], Chao-Ching Cheng[1], Chenming Hu[2], Iuliana P. Radu[1], Min Cao[1]

[1]Taiwan Semiconductor Manufacturing Company (TSMC), [2]National Yang Ming Chiao Tung University, [3]National Taiwan University, [4]National Cheng Kung University, [5]National Synchrotron Radiation Research Center

This study reports an epitaxial NiBi alloy contact that resolves the BEOL-compatible bottleneck while preserving high ON-current (I_{ON}) in 2D electronics. The NiBi alloy retains low metal resistivity ~50 μΩ·cm up to 500 °C and sustains reaction-free, stable contact interface on MoS_2 channel at 400 °C. The contact resistance (Rc) of the as-deposited NiBi contacts is ~420 W-mm at the carrier density of 1E13 cm^{-2}, which remains as low as ~550 W-mm after 400 °C forming gas annealing for 1 hour. The study here describes for the first time the materials and contact module analysis for NiBi on MoS_2 channel. We conclude that the NiBi alloy has sufficient thermal resilience and low contact resistance to advance readiness for practical applications.

Session 7: MS | Advanced Device Physics & Compact Modeling

1:45 PM, Continental 6

Co-Chairs: Aryan Afzalian, imec and Gianluca Fiori, Univ. Pisa

This session presents advances in the physical modeling, simulation, and characterization of scaled and emerging semiconductor devices. The first paper introduces COSMO, a compact model for ultrathin oxide-semiconductor transistors, that includes quantum confinement and disorder-shaped density-of-states physics. The second paper is an invited paper and presents a comprehensive process-simulation framework for advanced logic, memory, and 3D integration technologies. The third paper introduces MVS-3, a physics-based analytical MOSFET model that unifies drift-diffusion and quasi-ballistic transport. The fourth paper proposes an admittance-based framework for extracting interface metrics from lateral distributed MOS capacitor structures used in low-dimensional technologies such as 2D TMDs and carbon nanotubes. The fifth paper systematically characterizes device and ring-oscillator variability from 10 K to 300 K and builds a BSIM-compatible compact model for cryogenic SRAM design-technology co-optimization. The last paper of this session develops and validates a physical 1/f noise model for CMOS across 4 K to 400 K combining oxide-trap dynamics with band-tail-state trapping.

1:50 PM

7-1 COSMO: A Unified Quantum-Confinement-Aware Compact Model for Oxide Semiconductor Transistors Across 77–300 K: Bridging First-Principles Disorder Physics and VLSI Circuit Design

Ning Ji[1], Bingjie Dang[1], Duy Hieu Trinh[1], Chaoming Wu[1], Yanxiao Austin Wang[1], Yi Tong[1], Leming Jiao[1], Yiyuan Sun[1], Ying Xu[1], Gan Liu[1], Dong Zhang[1], Kaizhen Han[2], Yuye Kang[1], Jingsheng Chen[1], Xiao Gong[1]

[1]National University of Singapore, [2]Huazhong University of Science and Technology

We present the Compact Oxide Semiconductor Model (COSMO), a unified quantum-confinement-aware compact model for aggressively thinned indium-oxide thin-film transistors operating over 77–300 K. COSMO derives charge, transport, and dynamic response from a common disorder-shaped density of states (DOS) around the mobility edge. Temperature-dependent occupation and connectivity span thermally broadened and disorder-limited regimes without switching transport sub-models. First-principles density functional theory resolves the disorder potential and localized-to-extended state spectrum and supplies the thickness-dependent in-plane effective mass; subband quantization determines the confinement-induced mobility-edge shift. These physical inputs enter closed-form, continuously differentiable, and charge-conserving two-end equations for direct Verilog-A implementation.

COSMO reproduces measured I–V and C–V characteristics across channel thicknesses up to 3.5 nm and lengths from 50 nm to 10 μm. Temperature- and frequency-dependent validation captures the weak temperature dependence of subthreshold swing, pronounced low-temperature mobility reduction, and enhanced C–V dispersion from 50 kHz to 1 MHz. SPICE simulation of a three-transistor source-follower detector pixel yields a full-scale reconstruction error below 0.7%, demonstrating COSMO's applicability to VLSI circuit design and analysis from room temperature to cryogenic conditions.

2:15 PM

7-2 High-fidelity Process and Multi-Physics Simulations for Advanced Semiconductor Technologies

Nuo Xu, Kian Chuan Ong, Zhihao Xu, Xuefeng Zeng, Shourav Pednekar, Po-Yi Wu, Yu-Jen Chang, Mike Zhu, Hao Peng, Dongming Li, Yen-Ting Chou, Bo Pang, Yuan-Hao Chang, Xiaofeng Xiang, Kuang-Han Chen, Seungri Song, Xiaoyuan Qi, Ashish Nair, Yiru Jeong, Yuyang Sun, Danping Peng, Juan Rey
Siemens EDA

As semiconductor technologies advance, process complexity has escalated to a degree that renders purely **experimental-based learning and** optimization impractical. In this work, a comprehensive, physics-based process simulation framework **is introduced to** address the full spectrum of challenges across advanced logic, memory, and **3D** integration schemes. High-fidelity process flow simulations—capturing geometry, doping, and stress evolution—are augmented by metrology data (TEM, CD-SEM) to ensure structural and statistical fidelity. Monte Carlo variability analysis further quantifies process windows against yield targets. Rigorous process module simulations, built upon a unified reaction syntax for plasma etch, ion implantation, and **deposition *etc.***, provide fundamental physical insight into process outcomes. **T**he connection to **adjacent** manufacturing simulation domains, including design-for-manufacturing (DFM) and optical proximity correction (OPC), is **also** discussed.

2:40 PM

7-3 MVS-3: A Physics-based Analytical Model Unifying Drift-diffusion and Quasi-ballistic Transport

Dylan Ma, Lan Wei
University of Waterloo

This paper introduces MVS-3, a fully physics-based analytical model for MOSFETs, unifying both drift-diffusion transport where most of the channel is under quasi-thermal-equilibrium and quasi-ballistic transport where most of the channel is non-thermal-equilibrium. With a small parameter set, the model accurately captures the change in underlying physics and describes channel current, charge and white noise in a coherent framework. The model shows good agreement with measurements of transistors with channel lengths of 40 nm to 980 nm. Simulation results also provide important physical insights on the channel charge and noise under different transport regimes, providing a useful path for transistor compact modeling and device physics exploration in the ultra-scaled technology nodes.

3:30 PM

7-4 Generalized Admittance Framework for Efficient Interface Metrics Extraction on Lateral Distributed MOS Capacitor Circuits

Matthias Passlack[1], Carlo Gilardi[1], Dennis Lin[2], Iuliana Radu[1]
[1]Taiwan Semiconductor Manufacturing Company (TSMC), [2]imec

We developed a G_{it}/ω vs. frequency framework for interface metrics extraction on distributed lateral MOS capacitor of advanced low-dimensional technologies with 2D transition metal dichalcogenides (TMD) or 1D carbon nanotube channel. For the first time, all interface metrics including interface state density D_{it}, standard deviation of surface potential fluctuations σ_s, characteristic trap time constant τ_c, trap capture cross section σ_c, and trap energy E_T are self-consistently extracted for a lateral distributed MOS capacitor. Values for σ_s (1.23 kT/q) and a very small trap capture cross section σ_c (4.7×10^{-19} cm^2) are revealed for MOS capacitors with MoS_2 monolayer channel. Interface metrics are essential for correct FET performance prediction and controlled interface state density mitigation. Various tools for modeling and extraction of interface metrics on a distributed capacitor network are implemented in an integrated python environment. This work closes a significant gap in the development of low-dimensional channel FET.

3:55 PM

7-5 Cryogenic Variability in 7-nm FinFETs: From Systematic Characterization and Physical Decomposition to Variation-Aware DTCO

Hongbo Wang[1], Zirui Wang[1], Haoran Lu[1], Yuji Cheng[1], Lang Zeng[2], Runsheng Wang[1]
[1]Peking University, [2]Beihang University

In this work, we present a systematic characterization of variations in 7-nm FinFET devices and ring oscillators operating at cryogenic temperatures for the first time. A novel methodology is proposed which decomposes the overall measured variation into five distinct physical origins. The percolation-induced band-tail fluctuation is recognized as the origin of the strong low-temperature upturn in the normalized SS variation. Building upon this physics-based decomposition, a comprehensive BSIM-compatible compact model is established. This model precisely reproduces the experimental statistical distributions of cryogenic variations by capturing the underlying physical causes. Finally, leveraging this predictive modeling framework, we perform design-technology co-optimization (DTCO) for cryogenic SRAM. The proposed approach successfully bridges the gap between fundamental cryogenic physics and circuit-level evaluation, providing an accurate and actionable framework for optimizing the performance and stability of next-generation cryogenic CMOS technologies.

4:20 PM

7-6 Revealing the Temperature Dependence of 1/f Noise in CMOS from 4K to 400K: Physical Modeling and Experimental Validation

Jiaxuan Weng[1], Yingzhe Sha[1], Peng Wang[2], Zhangcheng Huang[1], Qi Liu[1], Ming Liu[1]
[1]Fudan University, [2]Shanghai Institute of Technical Physics, Chinese Academy of Sciences

For the first time, we present a temperature-dependent physical 1/f noise model with a quantitative formula from 4K to 400K. The proposed model uniquely combines oxide traps with trapping/de-trapping dynamics governed by the nonradiative multiphonon (NMP) model and band-tail states with trapping/de-trapping dynamics governed by the variable-range hopping (VRH) conduction model. In order to validate the model, we developed a low-noise measurement circuit monolithically integrated with the MOSFETs, enabling accurate 1/f noise extraction over the full temperature range. We measured the 1/f noise of nMOS and pMOS from 4K to 400K under various bias conditions, revealing a pronounced enhancement below 150K. The model is experimentally validated with a coefficient of determination R^2>0.9 over the full temperature range. Random telegraph noise (RTN) further confirms the VRH-governed dynamics of band-tail states. This work paves the way for a physical understanding of 1/f noise and low-noise circuit design across the full temperature range.

Session 8: ODI | Integrated Photonics and Single Photon Devices

1:45 PM, Continental 7-9

Co-Chairs: Gauri Karve, Imec and Yoshiaki Tashiro, sony

This session showcases seven papers in the area of single photon counting imagers and advanced integrated photonic devices. The session opens with a presentation by Kwan et al. from TSMC on 10 µm pitch Ge-on-Si single photon device with record low dark current and NEP (noise equivalent power) at 1310 nm. The second paper by Yoon et al. from Yonsei University describes a 300 mm BSI (back side illuminated) silicon meta surface SPAD (single photon avalanche diode) as a polarimeter. The third paper from the Liu et al. from Nanjing University demonstrates a SiC-SPAD for UV wavelengths with suppressed dark count rate achieved by introducing photo carrier acceleration method. Switching gears to integrated photonics, the fourth paper by Yun et al. from Samsung Electronics shows a fully integrated DWDM (dense wavelength-division multiplexing) optical engine for co-packaged optics, realized through heterogeneous integration of 300 mm silicon photonics and 4-nm FinFET CMOS. The fifth paper by Pan et al. from NSTIC discusses a photonic platform, integrating TFLN (thin-film lithium niobate) for 100 GHz modulation together with 110 GHz Ge photodetectors. The sixth paper in the session by Zhou et al. from A*STAR reports on phase-change material integration on a 200 mm silicon photonics platform enabled through backend-of-line processing. In the last paper, Lee et al. from KAIST discuss wafer scale III-V integration on a 220 nm thick SOI platform, realizing a SISCAP-based laser enabling high speed electro-optic tuning.

1:50 PM

8-1 10-μm Ge-on-Si SPAD with Record Low DCR of 0.60 Mcps, NEP 1.61×10^{-15} W-Hz$^{-1/2}$ at 1310nm under Room Temperature with Good Wafer Uniformity

Man Ho Kwan, Jung-I Lin, Chih-Pin Huang, Jheng-Yi Jiang, Chi-Ying Hsieh, Tsung-Chen Tai, Kuan-Chieh Huang, Shih-Min Huang, Chuan-Pu Chou, Ming-Chieh Hsu, Tzu-Jui Wang, Jun-Lin Tsai, Chen-Jong Wang, Dun-Nian Yaung
Taiwan Semiconductor Manufacturing Company (TSMC)

Ge-on-Si single-photon avalanche diodes (SPADs) have long faced the challenge of simultaneously achieving a low dark count rate (DCR) and high photon detection efficiency (PDE) to minimize the noise equivalent power (NEP). In this paper, multi-factor optimization strategies are employed to demonstrate a record-low DCR of 0.60 Mcps and an NEP as low as 1.61×10^{-15} W-Hz$^{-1/2}$ at 300K, and a wafer-scale DCR uniformity of 7.8%.

2:15 PM

8-2 Metasurface-SPAD Array Enabling Compact Polarization-Resolving Quantum Receivers

Hyeyoung Yoon[1,2], Eunsung Park[3], Kibaek Kim[1], Seyoung Yook[3], Doyoon Eom[3], Changhyun Kim[3], Dae-Hwan Ahn[1], Yong-Su Kim[1], Sangwook Han[1], Hyounghan Kwon[1], Myung-Jae Lee[3]
[1]Korea Institute of Science and Technology, [2]Korea University, [3]Yonsei University

Compact and scalable polarization-resolved singlephoton detection is a key enabling technology for emerging quantum communication and sensing systems, yet conventional implementations rely on bulky optical components and sensitive optical alignment. Here, we propose and demonstrate, to the best of our knowledge, the first metasurface-SPAD array (MSA), which integrates metasurface-enabled polarization-to-space conversion with a back-illuminated Si-SPAD array to directly resolve photon polarization at the single-photon-level within a compact architecture. The proposed MSA is experimentally validated through two representative demonstrations: a single-photonlevel polarimeter that reconstructs Stokes vectors with a mean angular error of 0.0687 rad via learning-based processing; and a compact free-space BB84 quantum key distribution receiver achieving quantum bit error rates of 0.6% to 2.6%. These results establish the MSA as a scalable and CMOS-compatible platform for next-generation quantum photonic applications.

2:40 PM

8-3 Achieving Tens-of-Hz-Level Dark Count Rate in SiC UV SPADs at Room and Elevated Temperatures

Ruide Liu, Yan Zhou, Jiawei Cao, Renjie Xu, Mai Zhang, Dong Zhou, Weizong Xu, Fangfang Ren, Feng Zhou, Rong Zhang, Hai Lu
Nanjing University

We demonstrate a SiC ultraviolet single-photon avalanche diode (SPAD) with a record-low dark count rate (DCR) **per unit area** in free-running mode at both room and elevated temperatures, which is superior to all existing semiconductor avalanche photodiodes. By redesigning the electric-field distribution, photocarriers generated in the absorption layer acquire sufficient pre-acceleration before entering the multiplication layer, enabling photon-triggered avalanche multiplication even with a reduced multiplication-layer electric field for dark-count suppression. The device exhibits a room temperature DCR as low as 44 Hz (0.058 Hz/μm^2) with a peak single-photon detection efficiency (SPDE) of 20.6% at 310 nm. Even at a high temperature of 150°C, the DCR remains below 500 Hz with a SPDE over 15%. The spectral detection performance and timing jitter characteristics are also investigated. This work has not only realized a UV SPAD with outstanding performance, but also provided a novel design strategy for semiconductor-based single-photon detectors.

3:30 PM

8-4 A 32-Gb/s/λ, 8-λ Fully Integrated DWDM Optical Engine for Co-Packaged Optics

Seokho Yun[1], Hyundai Park[2], Kimin Jun[2], Hankyu Lee[1], Wan Kim[2], Seok-Hee Hwang[2], Hyunchul Jung[2], Changgyun Shin[1], Jaekeun Lee[2], Hyochul Kim[2], Soohwan Lee[2], Bongkwon Son[1], Yongjae Park[2], SeungHun Lee[2], Enbin Jo[2], Hongkyu Park[1], Yunju Choi[2], Seung-Kuk Kang[2], Minhwa Chang[2], Jaewook Lee[1], Sungbeom Kim[2], Kwansik Cho[2], Hyungu Han[2], Ghoo Kim[1], Jongpil Cho[2], Raeyoung Park[2], Dohyeon Kim[2], Young Ja Kim[1], Yanghoon Lee[2], Seong-Hwan Kim[2], Seok Geun Ahn[2], Inoh Hwang[1], Haojiang Li[2], Young Chul Kim[2], Heejin Choi[1], Jaehyuck Jang[1], Ruiheng Peng[2], Jongwoo Park[2], Changyeon Son[2], Byungsu Jung[1], Siwook Yoo[2], Seungjun Han[2], Chaeryeong Park[2], Yonghan Kim[1], Xingliang Zhao[2], Jung-Tack Yang[2], Jaesun Baek[2], Young-zoon Yoon[1], Young Kim[1], Ghil-Geun Oh[2], Joonhoi

Hur[2], Wonkyoung Choi[2], Youngdae Kim[2], Sangmin Yoo[2], Kangho Lee[2]
[1]Samsung Advanced Institute of Technology (SAIT), [2]Samsung Electronics
We present a fully integrated 32-Gb/s/λ, 8-λ dense wavelength-division multiplexing (DWDM) optical engine for co-packaged optics, realized through an end-to-end in-house technology stack: a photonic IC (PIC) on 300-mm silicon photonics platform, an electronic IC (EIC) on 4-nm FinFET CMOS, thermo-compression-bonded heterogeneous integration, a glass-block-based optical interface, engine assembly, and electrical/optical test. On the O-band PIC, eight cascaded micro-ring modulators (MRMs) on a transmit (TX) bus and eight micro-ring filters (MRFs) with Ge photodetectors on a receive (RX) bus are located on a 200-GHz DWDM grid. The TX forwards a 16-GHz clock on one of the eight lanes, eliminating per-lane clock-and-data recovery at the RX; the remaining seven lanes carry a 224-Gb/s payload. The engine demonstrates optical eye diagrams on all eight channels and a closed 32-Gb/s TX–RX link. Total engine energy is 3.85 pJ per payload bit, including estimated external-laser power and MRM/MRF resonance tracking.

3:55 PM
8-5 Wafer-Scale TFLN-on-Silicon Photonics Heterogeneous Integration Platform Featuring 110- GHz TFLN Modulators and Ge Photodetectors for 400 G/Lane Optical Interconnects
Biwei Pan[1,2], Bo Li[1,2], Roger Tern[1,2], Chaeeun Lee[1,2], Chew Ping Leong[1,2], Yiwen Zhang[1,2], Yuan Gao[3], Xudong Jia[3], Jiawei Sun[3], Xingwen Hong[3], Xuan Sang Nguyen[1,2], Limeng Zhang[1,2], Chen Lau[1,2], Shi Yuan Khor[1,2], Suan Neng Foo[1,2], Wen Qi Tan[1,2], Jae Ok Yoo[1,2], Hazwani Khairy[1,2], Fujino Masahisa[1,2], Wai Kong Yeoh[1,2], Kah Wee Ang[1], Navab Singh[1,2], Leh Woon Lim[1,2], Landobasa Tobing[1,2], Hoong Shing Wong[1,2], Xianshu Luo[1,2], Yee Chia Yeo[1,2]
[1]National semiconductor Translation and Innovation Centre (NSTIC), [2]Institute of Microelectronics, Agency for Science, Technology and Research (A*STAR), [3]National University of Singapore
We report the first 200-mm wafer-scale photonic heterogeneous integration (PHI) platform that co-integrates thinfilm lithium niobate (TFLN) modulators, germanium (Ge) photodetectors, hydrogen-free silicon nitride (SiN) waveguides, and silicon (Si) photonic circuits using a scalable die-to-wafer (D2W) bonding process. By leveraging TFLN for ultrahigh-speed and energy-efficient modulation, Ge for high-speed photodetection, low-loss SiN for optical routing, and Si for wafer-scale integration, the platform combines the strengths of multiple material systems within a manufacturable architecture. The co-integrated TFLN modulators and Ge photodetectors simultaneously achieve bandwidths of 110 GHz, with the modulators exhibiting a $V_\pi L$ of 2.5 ± 0.37 V·cm and the photodetectors delivering a responsivity of 0.98 ± 0.08 A/W and dark current of 9.3 ± 2.29 nA. This demonstration provides all key transmitter and receiver building blocks required for 400 G/lane optical interconnects and establishes a scalable PHI platform for next-generation high-capacity, energy-efficient optical transceivers targeting AI, datacenter, and communication networks.

4:20 PM
8-6 First Demonstration of Wafer-Scale Integration of Phase-Change Material (PCM) with Silicon Photonics for Photonics Functional Diversification
Jingkai Zhou[1], Hancheng Liu[1], Mingjin Dai[1], Haoyang Sun[1], Tian Dong[1], Landobasa Tobing[1], Roger Tern[1], Lehwoon Lim[1], SuanNeng Foo[1], WenQi Tan[1], Jae OK Yoo[1], Navab Singh[1], Hoong Shing Wong[1], Qian Wang[1], Bowei Dong[1], Xianshu Luo[1], Yee Chia Yeo[2]
[1]Agency for Science, Technology and Research (A*STAR), [2]National Semiconductor Translation and Innovation Centre (NSTIC)
Phase-change materials (PCMs) enable post-trimming and non-volatile programming functions in silicon photonics, in a compact and energy-efficient manner. Despite the active exploration of integrating PCMs with silicon photonics at die-scale, the broader adoption of these materials and functions is hindered by the lack of wafer-scale exploration. Here, we demonstrate the first 200-mm wafer-scale integration of GST ($Ge_2Sb_2Te_5$) with silicon photonics using a continuous foundry fabrication flow. A commercial wafer-scale optoelectronic tester is modified to enable automated write-read characterization across the wafer, providing a general framework for wafer-level PCM testing. Wafer-scale thickness and optical attenuation measurements confirm good material and device uniformity. Programming studies show that one-time multilevel non-volatile programming is consistent across the wafer and promising for large-scale deployment, whereas multi-time programming exhibits substantially larger variation and requires further material and device optimization. These results establish a practical foundation for scalable PCM-integrated silicon photonics and identify key directions for optimization.

4:45 PM

8-7 First Demonstration of a SISCAP-Based Vernier Ring Laser for High-Speed Wavelength Tuning

Shin Hyung Lee[1], Donggil Kang[1], Jae-Hoon Han[2], Ju Hyeok Park[1], Yonghwan Han[1], SangHyeon Kim[1]
[1]Korea Advanced Institute of Science and Technology (KAIST), [2]Korea Institute of Science and Technology

High-speed wavelength-tunable integrated lasers are essential for frequency-modulated continuous-wave (FMCW) LiDAR sensing. We demonstrate, to the best of our knowledge, the first InGaAsP–insulator–silicon capacitor (SISCAP)-tuned Vernier-ring external-cavity laser on a 220 nm silicon-on-insulator (SOI) platform. The InGaAsP-based SISCAP provides high-speed capacitive tuning with lower calculated carrier-induced loss than Si p-i-n tuning, while avoiding continuous injection current and associated Joule heating. We developed a shared bonded III–V epitaxial-stack process that co-fabricates the optical-gain and SISCAP tuning sections within a common laser-integration flow, eliminating a separate modulator-material integration step. The proposed SISCAP modulator exhibits rise and fall times of 5.7 and 12.4 ns, respectively. The integrated laser achieves voltage controlled wavelength tuning over 28.56 nm and a maximum on-chip output power of 1.5 mW. These results establish carrier-accumulation SISCAP tuning as a distinct intracavity control approach for nanosecond-scale, wavelength-agile integrated lasers relevant to FMCW LiDAR.

Session 9: PMA | Advanced (Ultra-) Wide-Bandgap Power Devices and System Integration

1:45 PM, Imperial A

Co-Chairs: Yuhao Zhang, The University of Hong Kong and Donald Gajewski, Wolfspeed

This session highlights cutting-edge advancements in wide-bandgap and ultra-wide-bandgap power semiconductor technologies. Showcasing a diverse range of materials—including SiC, GaN, AlGaN, diamond, and IGZO—the presentations explore breakthrough device architectures, packaging innovations, and system-level integration. The session begins with an overview by Infineon Technologies of new SiC technologies tailored for the critical demands of upcoming power infrastructure. Pushing SiC performance limits, a joint team from Zhejiang University and The University of Hong Kong demonstrates a 6.5-kV SiC superjunction merged PiN Schottky diode achieving a record surge robustness exceeding 1000 A. Addressing packaging challenges, researchers from the University of Toronto, HKUST, and Dana Canada Corporation introduce a chip-on-chip SiC power module that co-optimizes parasitic inductance and thermal resistance. Transitioning to ultra-wide-bandgap and emerging oxide materials, Honda R&D and AIST report the first demonstration of a 10 A-class diamond power MOSFET utilizing a two-dimensional hole gas channel. KU Leuven and the Chinese Academy of Sciences then detail a monolithic 3D-integrated interconnect-gated IGZO transistor achieving breakdown voltages over 1000 V above 180-nm CMOS. Highlighting novel GaN architectures, Peking University, Tsinghua University, and HKUST present a GaN dual-channel p-FET technology that enables complementary logic circuitry to operate robustly from 2 K to 475 K. Nagoya University and Cornell University further advance the ultra-wide bandgap field by demonstrating high aluminum-mole-fraction vertical AlGaN diodes utilizing distributed polarization doping. Finally, Chang Gung University and UMC present a 100V graded-barrier p-GaN HEMT featuring a vertical-grid hot-via drain design, optimized for highly reliable 48-V intermediate bus converters in data centers.

1:50 PM

9-1 New solutions based on SiC power semiconductors meeting upcoming demands for infrastructure applications

Peter Friedrichs
Infineon Technologies

Beside classical power semiconductor applications like motor drive inverters, power supplies or renewable energy generation another field of use is gaining attention – power infrastructure. Mainly highly efficient components like SiC and GaN transistors are intended to serve this market. To meet the special demands in critical infrastructure systems adapted and new power semiconductor technologies need to be developed and rolled out. The paper gives an overview about the major use cases, related requirements as well as new device technologies like SiC JFETs. Furthermore, measures to guarantee highest uptime and reliability will be discussed.

2:15 PM

9-2 Overcoming Unipolar Surge Current Limits: A 6.5-kV SiC Superjunction MPS Diode with Over-1000 A Surge Robustness

Hengyu Wang[1], Chi Zhang[1], Binghao He[1], Haoyu Wang[1], Han Wang[1], Changwei Lei[1], Zhixin Chen[1], Xiangbo Huang[1], Haoyuan Cheng[1], Zan Wu[1], Yuhao Zhang[2], Kuang Sheng[1]

[1]Zhejiang University, [2]The University of Hong Kong

High-voltage SiC unipolar devices are limited in high-current and surge robustness by the strong positive temperature coefficient of on-resistance in thick drift regions. We address this limitation using a superjunction (SJ) architecture that expands the internal p–n junction area, enabling heavy drift-region doping and efficient conductivity modulation under high-current injection. We experimentally demonstrate a 6.5-kV trench-refilled SiC SJ merged PiN Schottky (MPS) diode with a Baliga figure-of-merit above 7 GW/cm^2. The fabricated 1-cm^2 device sustains more than 150 A continuously and exceeds 1000 A surge current, a 4× improvement over previously reported 6.5-kV-class devices. This performance is enabled by a widened MPS P+ anode that accelerates minority-carrier injection, together with a low-thermal-resistance microfluidic cooling platform. Switching tests confirm junction-capacitance-dominated unipolar reverse recovery, preserving high-frequency switching performance. These results establish SiC SJ technology as a promising solution for high-voltage, high-power applications.

2:40 PM

9-3 A Chip-on-Chip SiC Power Module with Ultra-Low Power Loop Inductance and Thermal Resistance

Jingyuan Liang[1], Xiaoyun Zhang[1], ChiJyun Ko[1], Xiangyu Teng[2], Matthew Birkett[3], Andrei Catuneanu[3], Wei Jia Zhang[2], Wai Tung Ng[1]

[1]University of Toronto, [2]Hong Kong University of Science and Technology (HKUST), [3]Dana Canada Corporation

A chip-on-chip, double-sided-cooled SiC power module is demonstrated to address the tradeoff between power loop parasitic and die-to-coolant thermal resistance in WBG power modules. The proposed module has three FR4 PCBs with embedded Aluminum Nitride (AlN) and copper inlays. The AlN inlays on the top and bottom boards provide thermally conductive high-voltage isolations for copper heat exchanger. The center copper inlays provide vertical electrical connection between the high-side and low-side SiC dies. The SiC dies are attached using high-temperature SAC305 solder. The assembled boards are attached to the copper heat exchangers using low-temperature Sn42/Bi58 solder. The chip-on-chip structure shortens the half-bridge current path while enabling double-sided heat extraction. The proposed SiC power module has a small power-loop inductance of 1.14 nH and low die-to-coolant thermal resistance of 0.16 K/W.

3:05 PM

9-4 10A-class hydrogen-terminated diamond MOSFET

Toshimitsu Kobori[1], Keita Takaesu[1], Daisuke Nakagawa[1], Yusuke Yamasaki[1], Yasuhisa Takahashi[1], Masaru Oku[1], Hidetoshi Arai[2], Masahiko Ogura[2], Toshiharu Makino[2], Hitoshi Umezawa[2], Daisuke Takeuchi[2], Sadaharu Matsumura[2]

[1]Honda R&D Co., Ltd., [2]National Institute of Advanced Industrial Science and Technology (AIST)

Hydrogen-terminated diamond MOSFETs fabricated on a half-inch wafer achieved 10 A DC drive capability and 3.4 A switching operation. By selecting functional devices and interconnecting them in parallel, threshold-voltage variation was reduced and high-current operation was realized. The packaged device exhibited an on-resistance of 2.8 Ω, switching times of 12 ns (fall) and 15 ns (rise), and low switching energy losses of 0.38 μJ (turn-off) and 0.13 μJ (turn-on). These results represent the first demonstration of a 10 A-class diamond power MOSFET and highlight the potential of diamond devices for future high-efficiency power electronics.

3:30 PM

9-5 First Demonstration of BEOL-Integrated High-Voltage IGZO Transistors on 180nm Si CMOS with Ultra-high Breakdown Voltage of >1000 V and Record-low Loading of <400 fA by Novel Gate-input Architecture

Eduardo Bemelmans[1], Menggan Liu[2], Xuhui Li[2], Sven Baerten[3], Yari Nowicki[3], Guanhua Yang[2], Ling Li[2], Kris Myny[3], Ming Liu[2]

[1]Department of Electrical Engineering (ESAT), KU Leuven, [2]Institute of Microelectronics of Chinese Academy of Sciences, [3]KU Leuven
This work reports the first monolithic 3D-integrated (M3D) interconnect-gated HV IGZO transistor above CMOS, enabling ultra-low-loading HV-to-LV sensing. The proposed interconnect-gated transistor (IGT) reuses the foundry CMOS metal/interlayer-dielectric (ILD) stack as a buried HV input gate and gate dielectric, thereby field-coupling the HV signal to the IGZO channel without a direct conduction path to the sensed node. This novel gate-input architecture demonstrates ultra-high breakdown voltages exceeding 1 kV. An additional BEOL top gate provides threshold-voltage tuning for CMOS-compatible programmable signal processing, as well as excellent bias-temperature stability under high-voltage gate stress. Hybrid M3D IGZO/Si circuits are demonstrated, in which a BEOL IGZO common-source stage converts the HV input into an analog intermediate voltage that is restored by a FEOL CMOS inverter. The circuit achieves programmable HV sensing, record-low sub-400 fA static input loading, and CMOS readout operation. System-level evaluation in a BMIC case study demonstrates a >10^4 lower input loading.

4:20 PM

9-6 2 K to 475 K GaN Complementary Logic Circuitry Using Dual-Channel p-FET Technology

Teng Li[1], Jingjing Yu[1], Yunhong Lao[1], Sihang Liu[1], Jialin Duan[1], Junjie Yang[1], Han Yang[1], Lin Yao[1], Xuelin Yang[1], Maojun Wang[1], Xiaosen Liu[2], Kevin J. Chen[3], Bo Shen[1], Jin Wei[1]
[1]Peking University, [2]Tsinghua University, [3]Hong Kong University of Science and Technology (HKUST)
This work reports GaN complementary logic (CL) circuitry realized by GaN dual-channel p-FET technology. The upper channel of the p-FET provides a strong 2DHG. At the gate region, the upper channel is recessed, and the lower channel is depleted at zero gate bias, resulting in E-mode operation. This work achieves several breakthroughs: (1) The dual-channel p-FET exhibits a superior V_{th}-I_{max} trade-off: V_{th} = −1 V and I_{max} = 38 mA/mm. This current density far exceeds those previously reported E-mode GaN p-FETs in CL platforms. (2) The GaN CL ROs achieve a record short τ_{pd} of 4.6 ns at 300 K. (3) The p-FETs are highly immune to the freeze-out effect at cryogenic temperatures. The CL circuitry exhibits superior performance from 2 K to 475 K. (4) The dual-channel structure enables an etch-stop process. Therefore, an improved fabrication uniformity is achieved, and a GaN CL arithmetic unit for subtraction is demonstrated.

4:45 PM

9-7 Compositionally Graded Polarization Doped AlGaN Power Diodes with a p^--Drift Layer

Zhiyu Xu[1,2], Jia Wang[1], Yingying Lin[1], Yoshio Honda[1], Debdeep Jena[2], Huili Grace Xing[2], Hiroshi Amano[1]
[1]Nagoya University, [2]Cornell University
High Al-mole-fraction vertical $Al_xGa_{1-x}N$ (x > 0.7) p^--n diodes and p-type Schottky barrier diodes (SBDs) are fabricated on an aluminum nitride (AlN) substrate. Polarization doping was employed for both p-type and n-type layers of the p^--n junction. Capacitance-voltage measurements and calculations based on SIMS-measured Al composition gradient indicate the polarization-induced charge concentration of ~2×10^{17} cm^{-3} in the p^--drift layer. The fabricated diode devices showed a low built-in potential of 5.45 V and turn-on voltage of 6.2 V, with a low differential specific ON-resistance of 2 mΩ·cm^2. Furthermore, high breakdown electric field was achieved at 7.44 MV/cm for the p^--n diode device. The p-SBD was analyzed, showing a thermally activated hole injection across the metal/p-AlGaN barrier. At high current density, the differential ON-resistance becomes nearly temperature-independent, revealing a transition from barrier-limited injection to series-resistance-limited transport.

5:10 PM

9-8 Low-Voltage AlGaN Graded Barrier p-GaN HEMT with Vertical Grid Hot-Via Drain Design: High Frequency and High Gate Reliability GaN Power Device for Data Center Intermediate Bus Converter Application

Hsien-Chin Chiu[1], Chong-Rong Huang[1], Yue-Ming Hsin[2], Jen-Inn Chyi[2], Chang-Ching Tu[2], Li-Yi Peng[3], Hao-Chung Kuo[3], Jia-Fu Lin[4], Ming-Che Lai[5]
[1]Chang Gung University, [2]National Central University, [3]National Yang Ming Chiao Tung University, [4]Wavetek Microelectronics Corporation, [5]United Microelectronics Corporation (UMC)
In this work, we report the first demonstration of a 100V graded-barrier p-GaN HEMT featuring a vertical-grid hot-via drain back-end-of-line architecture, enabling high-frequency and high-reliability operation for 48-V intermediate bus converters in data-center power supply applications. Compared with conventional Cu pillar flip-chip packaging,

the proposed hot-via drain routing through a 100-µm-thick Si substrate provides superior heat dissipation and significantly reduces drain/source routing parasitics. By introducing diluted polarization charges through the graded-barrier design, the GB HEMT achieves a 1.2-V increase in threshold voltage (V_{TH}) compared with the standard (STD) HEMT. Furthermore, the polarization-graded electric-field distribution reduces the V_{DS} dependence of the device capacitances, thereby mitigating switching losses and Miller-induced delay variation during high-speed switching operation. Double-pulse-test (DPT) results show that the gate-voltage (V_G) and drain-voltage (V_D) overshoot ratios are 44% and 48%, respectively, for the STD HEMT, whereas they are reduced to 16% and 32%, respectively, for the proposed GB HEMT.

Session 10: RSD | Oxide-Semiconductor FET Reliability and Degradation Physics

1:45 PM, Imperial B

Co-Chairs: Christine Hau-Riege, Marvell and David Chu, TSMC

In this session, the reliability of advanced oxide-semiconductor transistors is investigated through studies of electrothermal behavior, bias-temperature instability (BTI), hydrogen-related degradation, and long-term device stability. Collectively, these papers provide new insights into reliability mechanisms, mitigation strategies, and scaling limits for next-generation oxide-semiconductor devices. The first two papers examine self-heating in IGZO transistors and experimentally identify positive-feedback electrothermal instability in gate-all-around oxide-semiconductor FETs, establishing key reliability challenges associated with device scaling. The next three papers focus on BTI, demonstrating robust resistance to contact-mediated degradation, overcoming the mobility-reliability tradeoff in crystalline InOx TFTs, and enabling active stabilization of sub-40-nm ITO FETs through workload engineering. Next, the session explores the role of hydrogen in oxide-semiconductor reliability. One paper investigates suppression of hydrogen-induced PBTI using nitrogen doping in indium tungsten oxide transistors, while another reveals universal hydrogen-defect kinetics responsible for non-Arrhenius PBTI behavior across oxide-semiconductor technologies. The final paper demonstrates a record 10-year PBTI lifetime under 2.1-V stress at 100°C in stackable oxide-semiconductor FETs, highlighting the maturity and robustness of this emerging technology.

1:50 PM

10-1 First Direct Demonstration of Self-Heating in ALD IGZO TFTs and Scaled Vertical IGZO Transistors and Its Implications for Electrothermal Scaling and Reliability

Chun-An Shih[1], Sumi Lee[1], Mir Md Fahimul Islam[1], Jian-Yu Lin[1], Yi-Chieh Chen[1], Jeongwoo Park[2], Minji Hong[2], Wonsok Lee[2], Min Hee Cho[2], Pilsang Yun[2], Daewon Ha[2], Ai Shakouri[1], Peide Ye[1], Muhammad Ashraful Alam[1]
[1]Purdue University, [2]Samsung Electronics

Backend-integrated oxide transistors are emerging for next-generation electronics. Here, for the first time, we directly characterize self-heating (SH) in IGZO transistors, including an industry-grade vertical-channel transistor, using high-resolution thermoreflectance imaging across architectures, geometries, compositions, and thermal environments. Thermal dissipation is dictated by the effective heat-spreading area (Aeff) and surrounding thermal environment. For In-rich top-gated planar IGZO TFTs, scaling Lch from 1 to 0.15 µm increases thermal resistance (Rth) from 23 to 54 K/mW. The 70-nm-Lch IGZO VCT exhibits Rth= 260.7 K/mW, and 22.5% lower effective thermal conductivity than its TGP counterpart, indicating stronger thermal confinement. In composition modulates contact resistivity, channel sheet resistance, and Joule heating distribution. In-rich channels have improved Rth at long Lch, whereas lower-In channels exhibit lower Rth after short-channel scaling since contact-dominated heating extends Aeff and weakens its Lch dependence. Temperature-dependent reliability characterization establishes a safe operating power window to avoid SH-induced degradation in IGZO VCT.

2:15 PM

10-2 First Experimental Identification of Positive-Feedback Electrothermal Instability in GAA OS-FETs: Mechanism and Stability Limits

Fabia Farlin Athena[1], Jack C. Evans[2], Jimmy Qin[2], Nathaniel Safron[3], Mary Vaughan Petty[2], Karna Nagalla[2], Amy Siobhan McKeown-Green[2], Prasanna Sarkar[2], Tony Chutong Cui[2], Eda Deniz Demirel[2], Jaden Eric Cramlet[2], Greg Pitner[3], Jennifer A. Dionne[2], Yi Cui[2], Iuliana P. Radu[3], Srabanti Chowdhury[2], Alberto Salleo[2], Krishna Saraswat[2], H.-S. Philip Wong[2]

[1]Electrical Engineering, Stanford University, [2]Stanford University, [3]Taiwan Semiconductor Manufacturing Company (TSMC)
We report the first experimental observation of an abrupt positive-feedback electrothermal instability in gate-all-around (GAA) oxide semiconductor (OS) FETs. An abrupt on-current (I_{on}) rise and hysteresis are observed below 2 V and can lead to breakdown. Temperature-dependent DC I-V (25 °C - 225 °C), pulsed I-V measurements, and atomistic simulations identify the thermal current-amplification mechanism, while time-domain thermoreflectance (TDTR) measurements and electrothermal simulations quantify thermal confinement. Owing to their distinct physics, GAA OS-FETs exhibit a > 300% I_{on} surge, whereas GAA carbon nanotube (CNT) FETs exhibit no such instability. We derive an instability criterion based on electrothermal feedback gain (G_{inst}>1) and establish a safe operating area (SOA) framework. Channel and geometry optimization can reduce the peak temperature (T_{peak}) by 31% and expand the SOA by 45%, enabling future reliable ultra-dense monolithic 3D (M3D).

2:40 PM

10-3 Scaled Oxide Semiconductor Access Transistors with Robust Resistance to Contact-Mediated BTI

Hyun Jae Lee[1], Chengyang Zhang[1], Md Abdullah Al Mamun[2], Jaewon Shin[1], Eknath Sarkar[1], Hwan Kim[3], Wonsok Lee[3], Pilsang Yun[3], Daewon Ha[3], Kyeong Jae Cho[2], Asif Islam Khan[1], Shimeng Yu[1], Suman Datta[1]
[1]Georgia Institute of Technology, [2]University of Texas at Dallas, [3]Samsung Electronics
Bias-temperature instability (BTI) analyses overlook channel length (L_{Ch}) dependence in amorphous oxide semiconductor (AOS) transistors. In this work, we demonstrate that contact-mediated BTI imposes a critical, previously unrecognized bottleneck on AOS FET scaling. Through systematic evaluation of Ni and Pd contacts down to the nanoscale (L_{Ch} = 1 μm to 50 nm), we use on-the-fly DC/AC BTI characterization to demonstrate a major finding: Pd contacts suppress both PBTI and NBTI while mitigating L_{Ch}-scaling penalties. While both contact metals share an identical time-dependent power-law exponent (n»0.3), Pd achieves a lower degradation prefactor, confirming a reduction in active hydrogen-related traps. This mechanism is verified by DFT calculations, which predict a reduced equilibrium oxygen vacancy concentration at the Pd interface and a low hydrogen migration barrier. These findings establish Pd as an effective H-reservoir metal, proving that contact engineering is a vital path toward achieving reliable, scaled AOS access transistors for next-generation memory.

3:05 PM

10-4 Crystalline InO_x TFTs Breaking the Mobility–Reliability Tradeoff with 2.5× Mobility Enhancement, 15 mV PBS Shift after 3600s, Robust H Immunity, and 10× Lower Low-Frequency Noise

Ching-Shuan Huang[1], Shou-Ting Kao[2], Yao-Yu Lee[2], Ying-Hsuan Chen[2], Tzu-Pu Hung[3], Chien-Wei Chen[3], Ju-Huang Hsu[4], Chao-Hsin Chien[2]
[1]Institute of Electronics, National Yang Ming Chiao Tung University, [2]National Yang Ming Chiao Tung University, [3]National Institutes of Applied Research, [4]Integrated Service Technology
We demonstrate a rapid-thermal-oxidation (RTO)-enabled ultrathin InO_x thin-film transistor (TFT) that simultaneously achieves high mobility, excellent bias-stress stability, hydrogen immunity, and reduced low-frequency noise (LFN). Unlike conventional reliability-improvement approaches that often sacrifice carrier transport, RTO crystallizes the InO_x channel, enlarges the effective grain size, and suppresses localized-state-limited percolation. As a result, the RTO-treated TFT exhibits enhancement-mode operation with a 2.5× field-effect mobility (μ_{FE}) enhancement at room temperature and a 6× enhancement at low temperature. In addition, the PBS-induced $|\Delta V_{th}|$ is reduced to 15 mV at 25 °C and 42.1 mV at 85 °C after 3600 s, supported by reduced oxygen vacancy (V_O)- and OH-related defects. Under dynamic AC stress, the drain-current degradation is reduced to only 0.03× that of the as-deposited device. The RTO-treated stack further withstands 400 °C forming-gas annealing (FGA) by suppressing hydrogen penetration. Finally, a 10× reduction in LFN confirms a more uniform conduction network.

3:55 PM

10-5 From Passive BTI Characterization to Active Control: Workload-Engineered Stabilization of Sub-40-nm ITO FETs

Xuanqi Chen[1], Gan Liu[1], Yuxuan Wang[1], Rui Shao[1], Leming Jiao[2], Xiao Gong[2]
[1]National University of Singapore, [2]Agency for Science, Technology and Research (A*STAR)
For the first time, we demonstrate active, workload-aware stabilization of bias-temperature instability (BTI) in sub-40-nm ITO FETs by engineering the low-gate phase of operating waveforms. The control strategy is enabled by a

previously unrecognized drain-bias-activated hydrogen redistribution that competes with electron trapping. At $V_{DS,stress}$ = 2 V, the PBTI response of a 38-nm device reverses non-monotonically, accompanied by $G_{m,max}$ degradation, increased series resistance, and source/drain asymmetry, revealing a spatially nonuniform and recoverable H-related component. Under practical workloads, the high- and low-gate phases drive competing degradation components, allowing waveforms with identical high-phase exposure to produce opposite ΔV_{TH} polarities. Leveraging this phase-resolved competition, we identify an engineerable neutral low-gate voltage, that balances electron trapping and H redistribution. A kinetics-assisted feedforward and closed-loop controller then adaptively updates V_{low}, confining workload-induced ΔV_{TH} within ±120 mV over 15000 s at 1 MV/cm. This work advances OSFET reliability from passive degradation characterization toward active, workload-aware stabilization.

4:20 PM

10-6 Insights into Suppression of Hydrogen-Induced PBTI in Nitrogen-Doped Indium Tungsten Oxide Transistors

Tanvir Haider Pantha, Kharanshu Bhojak, Md Abdullah Al Mamun, Soham Shirodkar, Minjong Lee, Jiyoung Kim, Kyeongjae Cho, Sourav Dutta
University of Texas at Dallas

PBTI in Amorphous Oxide Semiconductor transistors is dominated by hydrogen-induced donor generation and subsequent hydrogen overflow, causing large -$\boldsymbol{\Delta V_T}$. We report nitrogen doping in Indium-Tungsten Oxide(IWO:N) as an effective pathway to suppress this degradation mechanism. Across wide range of electric fields, duty factors, and frequencies, IWO:N outperforms IWO under both DC and AC PBTI. At 358K, AC-stressed IWO:N devices show $\boldsymbol{\Delta V_T = 65mV}$ after 1000s at 4 MV/cm(1 kHz, 50% duty factor)-13x improvement over IWO, with no mobility penalty. XPS and pre-stress multi-frequency C-V confirm the origin to be nitrogen-passivated oxygen-vacancy defects and suppression of sub-gap trap states near the conduction band minimum. Intermittent-fast C-V during PBTI stress directly captures the defect evolution, revealing suppression of hydrogen-related donor generation and elimination of hydrogen overflow in IWO:N. DFT calculations further confirm that N-H bond formation deepens proton trapping potential, raising the hydrogen diffusion barrier and reducing diffusivity by orders of magnitude.

4:45 PM

10-7 Non-Arrhenius Crossover in PBTI of Oxide Semiconductor FETs Reveals Universal Hydrogen Defect Kinetics

Md Sazzdur Rahman[1], Dylan Matthews[2], Bingyu Zhang[2], Hyun Jae Lee[3], Yu Hsin Kuo[3], Xianduo Zhao[2], Arijit Sarkar[2], Aiden Suganuma[2], Austin Harrison[2], Victoria M. Ravel[2], Sarah R. Evans[2], Aaron D. Franklin[2], Gaurav Thareja[4], Asif Khan[3], Suman Datta[3], Tania Roy[2]
[1]Department of Electrical and Computer Engineering, Duke University, [2]Duke University, [3]Georgia Institute of Technology, [4]Applied Materials

High-temperature positive bias temperature instability (HT-PBTI) in amorphous oxide semiconductor (AOS) FETs exhibits an anomalous, non-Arrhenius crossover that fundamentally differs from conventional silicon devices. While the threshold voltage shift is positive at room temperature, it vanishes at an intermediate crossover temperature and becomes strongly negative at elevated temperatures. Through systematic characterization of identical dual-gated devices, we demonstrate that this crossover is a universal phenomenon across RF-sputtered indium tin oxide(ITO), atomic layer deposited(ALD) indium gallium oxide(IGO), and indium tungsten oxide(IWO). The behavior originates from kinetic competition between dielectric electron trapping and stress-induced hydrogen donor(H^+) doping within the channel. We establish the Zero Temperature Coefficient(ZTC), the temperature where ΔV_{th} = 0, as a new reliability metric. A lower ZTC serves as an early indicator of severe PBTI degradation, which manifests as catastrophic negative shift in V_{th} in IGO at 125°C, whereas reducing channel thickness increases ZTC, suppresses H^+-doping, and improves device reliability.

5:10 PM

10-8 Reliable 10-year PBTI Lifetime at Record 2.1V Stress and 100℃ for Oxide Semiconductor Stackable FETs: Immunity to Undesired −ΔV_{TH} from H-doping and Beyond

Gangping Yan[1], Hao Shi[2], Lu Tai[1], Erwei He[1], Chuqiao Niu[1], Zhiyu Tian[1], Tongshang Su[2], Xinlv Duan[1], Yuting Chen[1], Fengbin Tian[1], Huan Liu[2], Qiqi Zhang[1], Zhao Liu[1], Ziheng Bai[1], Zhaoqiang Bai[1], Liguo Chai[1], Xiangli Liu[2], Xiangsheng Wang[1], Yuke Li[1], Jianqi Chen[1], Xueli Ma[1], Jinjuan Xiang[1], Weihan Fan[1], Bryan Kang[1], Jiangliu Shi[1], Jin Dai[1], Hongbo

Sun[1], Guilei Wang[1], Hao Meng[2], Mingyuan Liu[2], Kanyu Cao[2], Chao Zhao[1], Yiming Zhu[2]
[1]Beijing Superstring Academy of Memory Technology, [2]ChangXin Memory Technologies, Inc. (CXMT)
In this work, we present the five-tier-stacked oxide semiconductor field effect transistors (OS FETs) with robust immunity to the negative threshold voltage shift ($-\Delta V_{TH}$) and state-of-the-art $+\Delta V_{TH}$ metrics under positive bias temperature instability (PBTI). It is motivated by unraveling and mitigating significant $-\Delta V_{TH}$ contributors beyond gate oxide hydrogen (H) doping, including shallow oxygen vacancy (V_O) ionization in the OS channel (responsible for >50% of $-\Delta V_{TH}$) and the gate-governed H-doping effect. With mild suppression of V_O and adoption of a H-block gate material with H reduction processes, our devices deliver an ultra-low $+\Delta V_{TH}$ of 77 mV with single power-law fitting at 3 MV/cm, 100 ℃ for 100 ks. Subsequent tailored electron trapping optimization enables a 10-year PBTI lifetime at record-high 2.1-V stress and 100 ℃, validated by both voltage and temperature acceleration for the first time.

Tuesday, December 15, 2026

Session 11: MT | NAND & DRAM

9:00 AM, Grand Ballroom A
Co-Chairs: Tae Kyun Kim, SK Hynix and Yanli Zhang, Infineon
This NAND & DRAM session features eight papers highlighting the latest industry and academic advancements in 3D NAND and DRAM process technology, integration, and architectural innovation. In this session, IMEC will demonstrate a selective deposition enabled 3D NAND architecture with a fully separated charge-trap layer and integrated air-gap structure, targeting improved scalability and performance. SK Hynix will present its progress in integrating Metal-Induced Lateral Crystallization (MILC) technology into 176-word-line 3D NAND, enabling enhanced channel characteristics and process optimization. Addressing the industry's push toward higher capacity storage, Samsung will showcase a 2Tb QLC 3D NAND memory featuring multiple innovations, including advanced material engineering, ONO stack redesign for Mo gate integration, flexible ON-pitch scaling, HARC optimization, and a novel seeded poly-channel architecture. These innovations collectively enable TLC-like endurance and retention performance in a high-density QLC product. In addition, Solidigm will introduce a timed Selective Slow Programming Convergence (SSPC) scheme designed to enable reliable operation of 240-layer PLC NAND, providing a pathway toward further increases in storage density and cost efficiency. DRAM scaling recently advances beyond conventional 2D 1T-1C architectures. In this session, SK Hynix will present capacitor-less $4F^2$ vertical-gate floating-body cell to achieve 170-s retention at 85°C through an asymmetrical double gate structure, while Samsung will share a fabricated 16-tier vertically stacked 3D DRAM to demonstrate key solutions for array design, GAA access transistors, and Peri-on-cell integration where retention constraints are improved. At the system level, IMEC will cover cell-periphery disaggregation with sub-2-μm hybrid bonding to enable higher-density, energy-efficient HBM using n Array dies + 1 CMOS logic die configuration, while Tokyo University will deal with organic-interposer-based orthogonal stacking to demonstrate a 48-die memory stack without TSVs or hybrid bonding on memory dies, enabling an 84-GB, 3.1-TB/s, 1.10-pJ/b memory stack.

9:05 AM
11-1 Demonstration of $4F^2$ Vertical Gate Capacitor-less DRAM with 170-s Retention Time at 85℃
Seiyon Kim, Seonghun Lee, Sujee Kim, Seung Wan Chu, Hyunwoo Sim, Sangmyung Lee, Hyejung Choi, Kyungdo Kim, Kyoungchul Jang, Seungbum Kim, Joodong Park, Dongyean Oh, Jaeyun Yi, Sunghoon Lee, Youngmann Cho, Seonyong Cha
SK Hynix
Capacitor-less DRAM based on the Floating Body Cell (FBC) is demonstrated utilizing $4F^2$ vertical gate (VG) pillar structures. A record retention time of 170 s at 85 ℃ and 8000 s at 25 ℃ was achieved through an asymmetrical double gate structure, a scaled back gate oxide for strong coupling, and a GIDL-suppressed underlap junction. Crucially, the process flow is fully compatible with conventional 1T-1C VG DRAM fabrication, excluding the capacitor formation step, thereby offering a cost-effective path for extending DRAM scaling beyond the current technology limits.

9:30 AM

11-2 Design Strategies for Vertically Stacked DRAM Toward the Next DRAM Generation

Taekyong Kim, Yujin Kim, Jin-Woo Han, Jae-Hyun Park, Sang Hyun Sung, Soohyun Bae, Hyeoncheol Kim, Seung Uk Han, Gyuhwan Oh, Sang-Il Han, Sungmin Park, Juhyun Kim, Joongchan Shin, Taegyu Kang, Taejin Bae, Jeongsu Kim, Hongrak Choi, Kyoseon Choi, Min-Kyu Kim, Minsoo Kim, Incheol Nam, Sangmin Hwang, Siwoo Lee, Se Guen Park, Jaihyuk Song
Samsung Electronics

Vertically stacked DRAM (VS-DRAM) is a promising path to extend DRAM scaling beyond the 2D limit, but the transition from the conventional architecture introduces new design obstacles that must be identified and managed. This paper presents design strategies for the three major changes from conventional DRAM. For the cell array, capacitor structure, word-line pad scheme, and bit-line architecture are examined for their impact on net die, and BL shielding is shown to suppress the coupling noise inherent to the facing BL geometry. For the floating-body gate-all-around cell access transistor, low threshold voltage, GIDL, and floating-body effect are identified as retention constraints that must all be improved, and the corresponding engineering items are demonstrated to improve all four retention modes. For Peri-on-cell bonding, overlay, via-induced I_{off} degradation, and well isolation are addressed. These results provide practical design guidelines that position VS-DRAM as the path to the next DRAM generation.

9:55 AM

11-3 Peripheral disaggregation enabling (n)Array+(1) CMOS stack towards low-power HBM evolution

Aakash Patel, Prathamesh Dhakras, Gayle Murdoch, Nilotpal Sarma, Stefano Traferro, Khakim Akhunov, Joeri De Vos, Yongbin Jeong, Nouha Labyedh, Nicolas Pantano, Andy Miller, Marie Garcia Bardon, Julien Ryckaert, Gouri Sankar Kar, Zsolt Tokei, Dwaipayan Biswas
imec

High Bandwidth Memory (HBM) architecture suffers from substantial die area overhead due to dense peripheral logic. We propose a nA+1C (n Array dies + 1 CMOS logic die) configuration, disaggregating peripheral logic to an advanced FinFET node optimized for DRAM. This is enabled by sub-2µm pitch hybrid bonding and ultra-low parasitic (40% lower capacitance) Through Dielectric Vias (TDVs), validated by fabricated and bonded test-vehicle. Transient netlist simulation of a 20HI stack yields 1.42x increase in bit density, 3.66x and 3.62x improvement in energy and bandwidth efficiency respectively compared to *next-generation* HBM (1D DRAM node), highlighting the feasibility of cell-peri-bonding towards low-power scalable HBM evolution.

10:20 AM

11-4 A 48-Die TSV-Free Organic Interposer-Based Orthogonal Stacking Technology for 3D Memory-on-GPU Integration

Yuki Mitarai, Wonhyung Kuk, Yutaka Nishizawa, Mototsugu Hamada, Atsutake Kosuge
The University of Tokyo

An organic interposer-based massive orthogonal stacking assembly of IC (IntMOSAIC) technology is pro- posed for 3D memory-on-GPU integration, enabling an 84- GB, 3.1-TB/s, 1.10-pJ/b memory stack. The newly developed interposer-based assembly eliminates the TSVs, the hybrid bonding, and the sidewall processing on the memory dies, reducing the assembly cost. The first 48-die orthogonal stack is demonstrated at the 130-µm bump pitch, with the connectivity confirmed at all measurement points and the eye openings measured at up to 6.4 Gb/s.

11:10 AM

11-5 Self-aligned Charge-Trap-Layer Confinement by Area-Selective-Deposition and Airgap co-Integration for Improved Retention and Cell Interference in 30 nm Z-pitch 3D-NAND

Sana Rachidi[1], Tathagata Paul[1], Devin Verreck[1,2], Jana Loyo[1], Ivan Erofeev[1], Yoshitomo Hashimoto[3], Daigo Yamaguchi[3], Hui Qu[1], Silvia Armini[1], Christophe Lorant[1], Geert Van den Bosch[1], Maarten Rosmeulen[1,2]
[1]imec, [2]KU Leuven, [3]Kokusai Electric

In this work, we demonstrate a novel integration scheme enabling full SiN charge trap layer (CTL) confinement (CTC) using area-selective deposition (ASD). The approach relies on surface differentiation strategies driven by integration solutions. This approach enables self-aligned CTL selective growth directly from within the memory holes. This concept further enables confined CTL thickness tunability and the co-integration of local airgaps (AGs) between

separated CTL regions. While, as expected, the baseline confined CTL architecture exhibits degraded programming performance relative to continuous CTL devices, the introduction of a wraparound high-k (Hk) liner recovers programming efficiency while preserving complete CTL confinement. In addition, this CTC architecture significantly improves retention while reducing cell interference. Further reduction in neighbor-induced barrier lowering (NIBL) is achieved by AG co-integration with CTC.

11:35 AM

11-6 Advanced MILC Channel Integration for Enhanced Cell Current and Reliable Operation in High-Layer (>176-WL) 3D NAND Flash Memory

Byeongil Han, Sangheon Lee, Sejun Kwon, Heungyeul Lee, Yoonho Choi, Yoocheol Shin, Myungshik Lee, Inseob Noh, Chonga Hong, Eunwoo Baek, Keonsoo Shim, Dongheon Lee, Jaehyeon Ko, Hyun Sagong, Jihoon Park, Sunkak Hwang, Heedo Na, Minsik Jang, Eul Han, Hosung Yu, Ju Yeab Lee, Sanghyun Oh, Jum Soo Kim, Youngmann Cho, Seonyong Cha

SK Hynix

Multi-seed MILC enables large-grain Si channels for 176-WL, scalable to >500-WL 3D NAND, boosting current 3×, SS 30%, and Vt uniformity.

12:00 PM

11-7 Achieving TLC-like Endurance and Retention in 9th-Generation 2Tb QLC 3D NAND Flash Memory

Sejun Park, Changhyeon Lee, Hyun-Ki Choi, Joon Kim, Joonsung Lim, Woojae Jang, Junghoon Lee, Junho Seo, Suk-eun Kang, DoBin Kim, Chaehyeon Lim, Dongho Yoo, Soyeong Gwak, Gang-in Shin, Taein Lee, Woosung Yang, Sungjun Kim, Seungmin Lee, WoongSeop Lee, Sangyong Yoon, Jiho Cho, Byoungil Lee, Kyungyoon Noh, Seungwan Hong, Jin-yub Lee

Samsung Electronics

Rapidly growing AI industry requires higher-capacity, highly reliable, and cost-effective non-volatile memory. In this work, we present an industry-leading 9th-generation 2Tb QLC 3D NAND flash memory with TLC-like endurance and retention characteristics. Trap-assisted electron tunneling during retention is significantly suppressed by reducing shallow trap sites between the tunnel oxide and the trap nitride. Novel material engineering leads to electrical robustness against anode-hole injection damage during program/erase cycles. Based on a fluorine-free molybdenum gate, the ONO layers are redesigned to restore the overall cell reliability. To compensate for reliability variations of WLs, flexible cell pitches are extensively adopted. The channel hole profile distortion is suppressed by HARC etch process optimization. Finally, we demonstrate that the overall reliability of QLC can approach that of TLC in most conventional applications.

12:25 PM

11-8 Innovations in Floating-Gate PLC NAND Flash

Amol Joshi[1], Richard Fastow[1], Siddarth Biyani[1], Tzuning Fang[1], Golnaz Karbasian[1], Minsoo Lee[1], Haiming Li[2], Bhavya Lokasani[1], Ahsanur Rahman[1], Cindy Sun[1], Sagar Upadhyay[1], Robert Ursua[1], Ravi Velpula[1], Jungdal Choi[1]

[1]Solidigm, [2]SK Hynix

We present the next-generation floating-gate (FG) penta-level cell (PLC) NAND platform, scaling from 192 to 240 layers with a 2D pillar shrink that raises areal bit density to 30.9 Gb/mm^2. Structural gains are achieved through 3.7% vertical-pitch and 5.1% pillar-pitch scaling. To recover the reliability and performance margin that higher bit/cell operation erodes, we introduce a timed Selective Slow Programming Convergence (SSPC) scheme that replaces the analog SSPC bias with a digital timing ratio, narrowing programmed-state width. Measured distributions show that the timed algorithm narrows programmed states in both QLC and PLC modes. The recovered margin can be traded for lower raw bit error rate or for a larger ISPP step and faster programming.

Session 12: ALT | Oxide Semiconductor and 2D Channel Device

9:00 AM, Grand Ballroom B

Co-Chairs: Mocuta Anda, Micron and Naoto Horiguchi, IMEC

In this session we present advances in oxide semiconductor and 2D channel devices. We have eight presentations demonstrating enhanced device properties in these material systems, availability of CMOS devices, as well as

system level applications in power gating or as BEOL voltage regulators. Researchers from Samsung Electronics present high-performance MoS2 GAAFETs with interlayer-free HfO2 integration by substituting chalcogenide with nitrogen, which preserves the pristine van der Waals interface during gate dielectric integration. Researchers from TSMC present CMOS nanosheet gate stack study with MoS2 n-channel and WSe2 p-channel. AlN epitaxial interlayer prior to HfO2 deposition suppresses dielectric trap-induced hysteresis. Ultrathin 1-nm AlN interlayer covers NS channel conformally without NS channel damage. Researchers from NUS present low thermal budget BEOL-compatible n-IGZO/p-Te CMOS, featuring high performance and record-low Vt instability. Researchers from Peking University and Southeast University present backside on-chip power delivery architecture on FinFET with metal oxide TFTs and 8:1 integrated voltage regulator. This is the first experimental demonstration of on-chip power delivery integrated with FinFETs. Researchers from NUS demonstrate a double gate In2O3 device using p-NiO layer as charge compensation layer, achieving highly improved device characteristics, like enhancement-mode, high Ion and good short channel effects. Researchers from MIT have produced an ultra-scaled In2O3 device achieving enhancement mode device with good SCE and high performance. Next, researchers from Purdue University propose IGZO devices with Ti seeded contacts, demonstrating high performance and enhancement mode simultaneously. Our last paper, by researchers at Institute of Microelectronics, China, proposes an IGZO based power gating scheme achieving ultra-low static power and low IR drop.

9:05 AM

12-1 Record-High Performance 2D GAAFETs via Interlayer-Free Dielectric Integration

Van Luan Nguyen[1], Taejin Choi[1], Yoonhoo Ha[1], Dongmin Kim[1], Joonyun Kim[1], Min Seok Yoo[1], Sejin Kim[2], Joohoon Kang[2], Junyoung Kwon[1], Minsu Seol[1], Taehoon Kim[1], Yongsung Kim[1], Kyung-Eun Byun[1], Changhyun Kim[1]
[1]Samsung Advanced Institute of Technology (SAIT), [2]Yonsei University

Interlayer formation at the dielectric/2D channel interface severely limits electrostatic control and device performance in 2D gate-all-around FET (GAAFETs). Here, we demonstrate a CMOS-compatible, interlayer-free HfO_2 gate stack by substituting chalcogenide with nitrogen that preserves the pristine van der Walls interface during dielectric integration. This approach enables excellent gate controllability and record-high 2D GAAFET performance. The fabricated devices achieve an on-current of 846 μA/μm while delivering 262 μA/μm at V_{dd} = 1V, alongside a near ideal subthreshold swing (SS) of 61 mV/dec. Furthermore, the drain-induced barrier lowering (DIBL) is minimized to 4 mV/V at a 50 nm gate length, indicating negligible short-channel effects. These results establish interlayer-free dielectric integration as a viable technology for high-performance, CMOS-compatible 2D GAAFETs

9:30 AM

12-2 Nanosheet FETs with MoS2 n channel and WSe2 p channel: gate stack study

Yun-Yan Chung[1], Shu-Jui Chang[1,1], Han-Yi Huang[1], Wei-Sheng Yun[1], Jie-Yin Liu[1], Shin-Yuan Wang[2], Hui-Ting Liu[2], Li-Pu Shen[2], Ang-Sheng Chou[1], Yu-Wei Hsu[1], Dawei Heh[1], Shuer Ku[1], Zheng-Jie Wang[1], Chen-Feng Hsu[1], Pin-Syuan He[1], Goutham Arutchelvan[1], Ming-Yang Li[3], Chia Chin Cheng[1], Shao-Ming Yu[1], Bo-Jhih Chou[2], Yun-Cheng Liu[4], Ying-Mei Yang[5], Bo-Heng Liu[6], Chi-Chung Kei[6], Cheng-Dian Yan[2], Zih-Syun Shen[2], Lain-Jong Li[7], Yu-Chuan Lin[2], Chun-Jung Su[2], Chih-I Wu[4], Wen-Hao Chang[2], Tse-Ming Chen[5], Chao-Hsin Chien[2], Wei-Yen Woon[3], Chenming Hu[2], Chao-Ching Cheng[1], Iuliana P. Radu[1], Min Cao[1]
[1]Taiwan Semiconductor Manufacturing Company (TSMC), [2]National Yang Ming Chiao Tung University, [3]Corporate Research, Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu, Taiwan, [4]National Taiwan University, [5]National Cheng Kung University, [6]National Center for Instrumentation Research, [7]National University of Singapore

Study of dielectric interlayers for GAA FET with 2D materials channel and apply it to both N-MoS_2 and P-WSe_2 channels. Comparing an epitaxial dielectric interlayer-AlN with an amorphous high-k (HfO_X) interlayer, we find that hysteresis can be reduced with the epitaxial dielectric. Both approaches enable I_{OFF} < 10nA/um at V_G = 0V and low-gate-leakage providing ability to further scale EOT. By scaling the HfO_X/AlN stack to an EOT of 1.8 nm, GAA-FET with MoS_2 channels achieve a I_{ON} current of ~ 535 uA/um at V_{DS} = 1 V, I_{OFF} of ~ 8 nA/um at Vg = 0V, and median S.S. of ~ 112 mV/dec. Furthermore, it is validated that the AlN interlayer prior to HfO_X deposition effectively suppresses dielectric trap-induced hysteresis as typically observed with low-temperature HfO_X IL. The electrical improvements can be attributed to near 100% conformal coverage at an ultrathin 1-nm AlN IL with minimal damage to the NS channels

9:55 AM

12-3 Low-Thermal Budget BEOL-compatible n-IGZO/p-Te CMOS Featuring Te pFET with Record-High Performance and Record-Low Vt Instability (Ion,max=198 μA/μm@Vds=-1.6V)

Yue Li, Chaeseon Hong, Zihang Fang, Terry Chien-Jen Yang, Yizhou Jiang, Sonu Hooda, Evgeny Zamburg, Aaron Voon-Yew Thean

National University of Singapore

We present here a low-thermal-budget tellurium (Te)-based pFET enabled by channel, gate dielectric, and surface passivation engineering. The optimized Te pFETs achieved an ultra-low hysteresis of 0.20 V in long-channel devices, while aggressively scaled 60-nm short-channel devices delivered a record-high Ion of 198 μA/μm at Vds = 1.6V. To demonstrate circuit-level feasibility, oxide CMOS circuits were implemented using the optimized Te pFET and IGZO nFETs. CMOS inverters exhibited stable switching with high voltage gain, while multi-stage ring oscillators achieved a minimum stage delay of 25.2 ns in a 7-stage oscillator at VDD = 5 V. A fully functional 6T SRAM was also demonstrated, confirming reliable static memory operation. These results establish a practical BEOL-compatible complementary oxide CMOS platform spanning devices, logic, and memory, providing a viable pathway toward monolithic three-dimensional integrated computing systems.

10:20 AM

12-4 First Experimental Demonstration of Backside On-Chip Power Delivery Architecture on Advanced FinFET: Backside Metal Oxide TFTs, 8:1 Integrated Voltage Regulator, and System-Level Assessment

Feiyu Teng[1], Wenting Xu[2], Xiang Li[1], Ziyang Zhang[2], Yanbang Chu[1], Mingmin Lin[1], Runxiao Shi[2], Shen Xu[2], Weifeng Sun[2], Weihai Bu[1], Jin Kang[1], Wangran Wu[2], Ming Li[1], Runsheng Wang[1], Heng Wu[1], Ru Huang[1]

[1]Peking University, [2]Southeast University

For the first time, the BS on-chip IVR for BSPDN was successfully developed on advanced FinFET platform, where the FS IVR also validated for FSPDN. Thanks to the low thermal budget process of TFT, the IVRs are fully BEOL-compatible, composed of the E-mode IGZO/D-mode ITO TFT and capacitor. The three-stage power converter demonstrated power conversion from 6 to 0.72 V. For the first time, the on-chip power converters were integrated on advanced FinFET and directly powered the BS ROs, oscillating stably at 135 kHz. PI was examined on 3D systems with 2nm-class designs. As power management architecture evolves from MBVR to in-package IVR and finally to on-chip IVR, the IR drop is reduced by 6.8% and 20.2% in the 3D-IC with FSPDN, and more by 32.8% and 56.8% in the one with BSPDN, respectively. This work validated the backside on-chip IVR as an optimal power delivery solution for future 3D-ICs.

11:10 AM

12-5 A Scalable Stacked Double-Gate Charge-Balance In_2O_3 FETs: Achieving Enhancement-Mode Operation, High Drive Current, and Suppressed Short-Channel Effects

Weibing Hao[1], Jiawei Xie[1], Jingguang Lu[1], Leming Jiao[1], Huangbai Liu[1], Chen Sun[1], Kai Ni[2], Gengchiau Liang[3], Xiao Gong[1]

[1]National University of Singapore, [2]University of Notre Dame, [3]National Yang Ming Chiao Tung University

In this work, we demonstrate the first charge-balance junction-engineered stacked double-gate n-In_2O_3/p-NiO/n-In_2O_3 AOSFETs. An ultrathin p-NiO layer depletes the adjacent In_2O_3 channels at equilibrium, enabling enhancement-mode operation, while the stacked In_2O_3 channels maintain efficient on-state conduction. Furthermore, NiO-induced depletion region acts as an electrostatic buffer to screen drain-field penetration and suppress source-barrier lowering, thereby reducing DIBL and V_T roll-off. Through TCAD-guided optimization, our double-gate AOSFET with a 50-nm L_{CH} achieves outstanding electrical characteristics, including E-mode operation with V_T of 0.2 V (3.4 V positive V_T shift), I_{ON} of 1.15 mA/μm, a $G_{m,max}$ of 510 μS/μm at V_{DS} of 1 V, DIBL of 30 mV/V, V_T roll-off of 28 mV, and V_T temperature coefficient of 3.5 mV/°C.

11:35 AM

12-6 Ultra-Scaled Enhancement-Mode BEOL Oxide-Channel FETs

Yanjie Shao[1,2], Marco Pala[3], Hyungeun Choi[1], John C.-C. Huang[1], Hongkai Ning[1], Dimitri A. Antoniadis[1], David Esseni[3], Jesús A. del Alamo[1]

[1]Massachusetts Institute of Technology, [2]Stanford University, [3]University of Udine

We experimentally investigate deep lateral and vertical scaling of back-end-of-the-line (BEOL)-compatible amorphous oxide semiconductor (AOS) field-effect transistors (FETs) and achieve record performance. We show that aggressive gate dielectric thickness scaling down to equivalent oxide thickness (EOT) of 0.28 nm enables deep channel length (L_{ch}) scaling with strongly suppressed short-channel effects (SCEs) and greatly improved on-state performance. Quantum simulations based on Non-Equilibrium Green's Function (NEGF) formalism are in good agreement with charge-control device characterization. Ultra-scaled devices featuring L_{ch} = 8 nm and EOT = 0.28 nm exhibit favorable enhancement-mode (E-mode) operation with a maximum drive current (I_{max}) of 2.65 mA/μm, a peak transconductance ($g_{m,peak}$) of 2 mS/μm, and a close-to-thermal-limit average room-temperature subthreshold swing (S_{avg}) of 70 mV/dec, all at a drain-to-source voltage (V_{ds}) of 1 V. These results establish a pathway toward aggressively-scaled, high-performance BEOL electronics.

12:00 PM

12-7 Breaking V_T – R_C Trade-Off in InGaZnO FETs with Ultra-Scaled Contact Length Through Ti-Seeded Contacts

Jian-Yu Lin[1], Sumi Lee[2], Zehao Lin[2], Chang Niu[2], Chang Liu[2], Juanjuan Lu[2], Kisoo Nam[2], Jeongwoo Park[3], Wonsok Lee[3], Min Hee Cho[3], Pilsang Yun[3], Daewon Ha[3], Haiyan Wang[2], Peide D Ye[2]

[1]Elmore Family School of Electrical and Computer Engineering, Purdue University, [2]Purdue University, [3]Samsung Electronics

Oxide-semiconductor (OS) field-effect-transistors (FETs) with enhancement-mode (E-mode) operation are desirable for monolithic-three-dimensional-integration applications. However, InGaZnO (IGZO) FETs with positive threshold voltage (V_T) typically exhibit high contact-resistance (R_C), resulting in the long-standing V_T – R_C trade-off. In this work, Ti-seeded contacts, formed by inserting a 0.5-nm-Ti-layer between Ni and IGZO, are proposed to overcome this challenge. The Ti-seeded contacts provide sufficient oxygen-scavenging to induce n^+ doping in contact regions while suppressing excessive TiO_X interfacial layer formation, reducing R_C by 88 – 90 % across a wide range of In:Ga:Zn ratios. Leveraging Ti-seeded contacts, the IGZO FETs with ultra-scaled contact length (L_C) = 40 nm achieve a record-low R_C of 169 Ω·μm, record-high peak transconductance of 977 μS/μm, and record-high on-current of 616 μA/μm among reported OS-FETs with L_C = 40 nm and V_T > 0 V. These results demonstrate an effective contact engineering approach for E-mode OS FETs including IGZO.

12:25 PM

12-8 First Demonstration of BEOL-integrated IGZO Power Gates on Foundry CMOS Circuits Achieving Record-low Static Power of 1.59 fW and 48.4% IR-Drop Reduction with Zero Area Overhead

Wendong Lu, Yuenan Han, Menggan Liu, Fuxi Liao, Guanhua Yang, Ling Li

Institute of Microelectronics of Chinese Academy of Sciences

For the first time, a BEOL IGZO-based power gate (PG) architecture is proposed and experimentally demonstrated on foundry Si CMOS process. Contrary to traditional FEOL siliconbased PG techniques, this IGZO PG scheme is directly BEOLintegrated on CMOS circuits, significantly shortening the power delivery path and eliminating FEOL area overhead issue. To realize this architecture, a short-channel dual-gate IGZO PG transistor with LCH =45 nm is fabricated in the BEOL, exhibiting a low on-resistance of 1.51 Ω·mm, an ultra-low leakage of 8.86×10-16 A/μm, as well as stable BTI characteristics at 85C, satisfying the strict PG requirements. The optimized IGZO PG devices are further evaluated on representative CMOS building blocks, including inverters, half-adders, SRAM arrays, and ring oscillators. Experimental results demonstrate a remarkable sixorder reduction of standby leakage power in all CMOS circuits when the BEOL IGZO PG is turn OFF, which achieves a recordlow

Session 13: SMB | Physical, Chemical, and Multimodal Integrated Sensors

9:00 AM, Continental 1-3

Co-Chairs: Sophie Giroud, CEA-Leti and Savannah Eisner, Columbia University

This session spans a range of sensing modalities, including multimodal sensors (temperature, humidity, flow), chemical/gas sensors, motion sensors, tactile/physical sensors, and flexible sensor platforms. The session explores what sensor applications become possible when built on CMOS and thin film transistor platforms, and highlights in-sensor and in-memory hardware. The first topic of this session presented by Jonghyeon Yun and co-authors, will introduce a self-powered tactile sensor integrating a novel floating-gate-embedded triboelectric nanogenerator. The second talk, delivered by Kyeongho Eom and colleagues, will showcase a flexible CMOS-SOI platform, along with a XeF_2 backside etching technique, to realize a single-chip flexible sensing electronics system. Pursuing the sensors application using CMOS technology, Xinbo Jiang and colleagues will demonstrate a fully integrated multimodal sensing CMOS-MEMS platform combining temperature, humidity, and flow sensors. Hunhee Shin and colleagues present monolithically integrated complementary floating gate field-effect transistor gas sensors, amplifiers, and nonvolatile flash memory arrays co-designed for an efficient in-memory artificial olfactory system. Shiqian Chen and colleagues report the experimental demonstration of dual-mode, ferroelectric-based processing-in-sensor hardware within a 2T1C architecture for energy-efficient motion detection. Finally, Jiayi Mao and team demonstrate ultrathin atomic layer deposited indium oxide thin-film transistors and piezoresistive pressure sensors in an active matrix tactile sensor array.

9:05 AM

13-1 Floating-Gate-Embedded Triboelectric Nanogenerator with Non-Volatile Internal Charge Trapping for Self-Powered Synaptic Tactile Sensing

Jonghyeon Yun, Geunchul Kim, Jihyo Lim, Wookyeong Kang, Inkyum Kim, Yuseob Lee, Daewon Kim

Kyung Hee University

To address IoT energy constraints, a floating-gate embedded triboelectric nanogenerator (FG-TENG) emulating synaptic plasticity via internal charge trapping is proposed using a microstructured floating-gate structure with a 240 nm-thick amorphous silicon layer. Under cyclic contact, the FG-TENG stores triboelectric charges non-volatilely, enabling power-free, stepwise output evolution. Bipolar sweeps and 300-cycle endurance confirm robust charge retention, while thermally assisted detrapping under 150 ◦C stress demon strates potential for synaptic reset or volatile relaxation. The device exhibits a sensitivity of 0.320 V kPa−1 below 100 kPa (R2 = 0.985) and a peak power density of 0.188 W/m2 at 20 MΩ. Utilizing these features, the FG-TENG classifies tap, press, and walk with 94.4% average accuracy as a self-powered neuromorphic tactile platform.

9:30 AM

13-2 FlexCMOS: Single-die Flexible Electronics Enabled by XeF_2-etched SOI CMOS

Kyeongho Eom, Hyeong-Seok Oh, Sarika Madhvapathy, Michael Pedowitz, Jun-Chau Chien

University of California, Berkeley

This work presents a flexible electronics platform achieved by removing the silicon substrate via die-level XeF_2 backside etching of a commercially available SOI CMOS chip fabricated using the GlobalFoundries 45SPCLO process. By leveraging the buried oxide (BOX) layer as an etch stop, the bulk silicon is completely eliminated without the need for mechanical grinding, enabling the post-foundry process to be applied at a single-chip level while preserving the high integration density of advanced CMOS technology. The mechanical flexibility of the platform was evaluated under varying radii of curvature, demonstrating structural integrity down to 1.5 mm. Furthermore, we quantified transistor current-voltage (I-V) characteristics, the transfer characteristics of a capacitive-feedback transimpedance amplifier (C-TIA), and optical responses of integrated photonic devices under different bending conditions. This development offers a practical and scalable pathway for next-generation flexible bioelectronics.

9:55 AM

13-3 A Monolithic Vertically Integrated CMOS-MEMS Temperature-Humidity-Flow Sensor with Pulsed Time-Domain Multiplexed Readout

Xinbo Jiang[1], Daishan Yue[1], Ziqi Liu[1], Shuixiang Yin[1], Cheng Wan[1], Xiaoya Zhu[1], Hengzhang Yang[2], Wenbiao Zhou[1], Feng Gao[3], Huikai Xie[1], Bo Zhou[1], Bo Wang[4], Xiaoyi Wang[1]
[1]Beijing Institute of Technology, [2]The Chinese University of Hong Kong, [3]Zhejiang University, [4]Hamad Bin Khalifa University

This paper presents the first vertically integrated CMOS-MEMS temperature-humidity-flow sensor with monolithically integrated interface circuits on a single CMOS die. A CMOS-compatible post-CMOS process stacks a graphene-oxide-coated capacitive humidity sensor above a calorimetric flow sensor, with a temperature sensor and dedicated interface circuits underneath, reducing footprint and sensor-to-circuit interconnect parasitic. The shared microheater is operated in a pulsed time-division multiplexing scheme, enabling flow sensing and humidity desorption during the heater-on phase, while humidity and temperature are measured during the heater-off phase to mitigate thermal interference. The proposed sensor achieves sensitivities of 271.4 mV/(m/s), 0.35 pF/%RH, and 1 mV/°C for flow, humidity, and temperature sensing, respectively. Moreover, pulsed heater operation reduces the maximum humidity hysteresis from 26.1% to 7.1%. Compared with previously reported multimodal MEMS sensors, the proposed architecture uniquely combines vertically stacked multimodal sensing, monolithic CMOS-MEMS sensor-ASIC integration, and shared-heater pulsed time-division operation within a compact single-chip platform.

10:45 AM

13-4 Complementary FET Gas Sensors for Efficient In-memory Olfactory System

Hunhee Shin, Kangwook Choi, Joon Hwang, Jinwoo Park, Gyuweon Jung, Jong-Ho Lee
Seoul National University

This work presents complementary horizontal floating-gate (FG) FET gas sensors monolithically integrated with amplifiers and an AND-type nonvolatile flash memory array for a highly efficient in-memory olfactory system. By altering the charge-compensation path, the proposed coupled- and direct-type sensors generate opposite threshold voltage shifts to the same gas for the identical SnO_2 sensing layer. This complementarity drastically reduces mixed-gas prediction error to approximately one-fifth, eliminating the need for complex signed-weight computing. In NO_2/NH_3 mixed gases, real-time bit-line (BL) currents proportional to each gas concentration are extracted using strictly non-negative weights, demonstrating a breakthrough sensor-amplifier-memory co-design for a compact artificial olfactory system.

11:10 AM

13-5 Sensitivity-Configurable Active-Matrix Tactile Sensor Array Enabled by High-Performance Atomic Layer Deposited In_2O_3 Transistors

Bowen Zhu[1], Jiayi Mao[2], Shuhui Ren[2], Bangbang Qin[2], Weilong Feng[2], Fanfan Li[2], Yitong Chen[2], Tianzheng Zhou[2], Shuo Yang[2], Siyu Zhang[2], Guanlei Zhao[2], Saisai Wang[3], Qi Huang[3]
[1]Electronic and Information Engineering, Westlake University, [2]Westlake University, [3]Westlake Institute for Optoelectronics

In this work, we demonstrate an active-matrix tactile sensing platform based on high performance atomic layer deposited (ALD) ultrathin (5 nm thick) In_2O_3 thin-film transistors (TFTs) and piezoresistive pressure sensors. The fabricated TFTs achieve an average field effect mobility of 48.79 $cm^2 V^{-1} s^{-1}$, a peak transconductance (g_m) of 150 μS, and a low contact resistance of 2.55 kΩ·μm, enabling effective current modulation for tactile signal readout. By adjusting the TFT bias conditions, the sensitivity of the 1T1R tactile pixel is tuned from 175.48 to 986.78 kPa^{-1}, enabling configurable imaging contrast of pressure distribution. A 16×16 active-matrix tactile sensor array with an effective sensing area of 0.85 × 0.85 cm^2 is further demonstrated for pressure mapping. This work demonstrates a new role of ALD In_2O_3 TFTs in tactile sensing, extending their application beyond display and back-end-of-line (BEOL) logic.

11:35 AM

13-6 Energy-efficient Ferroelectric HZO and ITO TFT-based Dual-mode 2T1C Processing-in-sensor Array with Motion/Static Data Sensing and Processing

Shiqian Chen, Linghan Wang, Qihang Zhong, Gengyu Li, Yingyuan Weng, Runxiao Shi, Weifeng Sun, Hao Cai, Zhiyuan Fu, Wangran Wu
Southeast University

In this work, ferroelectric (FE) HZO and ITO TFT based 2T1C processing-in-sensor (PIS) hardware is proposed and experimentally demonstrated for the first time, featuring dual-mode of unified self-powered dynamic event detection and static data sensing functions in single PIS device. By utilizing the polarization modulated pyroelectric of FE-HZO and charge coupling effect of ITO transistor, the 2T1C PIS device can implement both of *in-situ* dynamic and static data sensing, and multi-bit multiplication of non-volatile weight and input dynamic/static sensing data in single 2T1C sensor under ultra-low power consumption (self-power for dynamic and 0.5 μW/pixel for static mode), alone with the advantage of high endurance, long retention and linearity. Moreover, the reconfigurable PIS array for versatile workloads is further designed and experimentally demonstrated under IR, enabling in-array MAC for varied task requirements. Based on the design, pattern recognition tasks achieves high accuracies showing its promising potential for edge intelligent sensing systems.

Session 14: MS | Memory, Device and System-Technology Co-Optimization

9:00 AM, Continental 4

Co-Chairs: Mohammad Hasanuzzaman, IBM Reserach and Albert Fayrushin, Micron

This session includes six papers focusing on memory, device and system technology co-optimization. The first paper by Sungwon Cho from SK Hynix describes propagation of device level trade-offs of various memory cells to the system level in HBM-based AI servers in terms of power, capacity and speed. The next paper by Meng-Lin Wu from National Taiwan University presents a method of closer integration of SRAM to the through silicon vias without having the stress from those vias degrading the memory cells. The third paper by Haobo Huang from Fudan University demonstrates connection between atomistic defects at the MoS_2/HfO_2 interface and the failure rate of a real SRAM cell. The fourth paper by Shiyang Li from Peking University demonstrates predicting heating from stacking memory chips in a full scale range from single atoms up to the whole package to redesign thermally optimized HBM tower. The next paper by Yue Yu from Peking University focuses on aging aware model that tracks oxide transistor BTI from single devices to neural network data retention. The last paper in this session by Mainakh Mukherjee from Purdue University demonstrates BEOL IGZO-based eDRAM arrays enabling large on-chip GPU scratchpads with minimal overhead, reducing memory bottlenecks that provide energy savings and lower latency for LLM workloads.

9:05 AM

14-1 System-Technology Co-Optimization Platform for Next-Generation High-Bandwidth Memory on Server-Scale LLM Serving System

Sungwon Cho[1,2], Po-Kai Hsu[1], Ashkan Moradifirouzabadi[1,4], Kiseok Lee[2], Kangsik Choi[3], Jungwon Seo[3], Suock Chung[3], Hyeonwoo Park[2], Eknath Sarkar[2], Mingu Kang[4], Jungmin Choi[1], Suman Datta[2], Shimeng Yu[2], Hoshik Kim[1]
[1]SK hynix, [2]Georgia Institute of Technology, [3]SK Hynix, [4]University of California, San Diego

Monolithic 3D DRAM technologies enable scaling beyond 2D physical limits and are capable of being seamlessly integrated into High-Bandwidth Memory (HBM). However, conventional Design-Technology Co-Optimization (DTCO) fails to intuitively project microscopic device trade-offs onto macro-scale server performance. This work introduces an end-to-end System-Technology Co-Optimization (STCO) platform for Large Language Model (LLM) serving workloads. We present a TCAD-to-system evaluation of Silicon 3D DRAM, Amorphous-Oxide-Semiconductor (AOS) 3D DRAM, and AOS 3D FeRAM. Utilizing a two-stage cascaded analytical framework calibrated against TCAD and SPICE, device parameters are seamlessly propagated onto 16-Hi HBM stack performance. Design space exploration propagated to server-scale workloads using an extended discrete-event LLM serving simulator, capacity-dense AOS FeRAM achieves a 2.3x throughput enhancement for large-batch operations targeting free-tier services. Conversely, faster AOS 3D DRAM achieves up to a 22% memory subsystem energy reduction in small-batch premium services.

9:30 AM

14-2 TSV-Stress-Aware SRAM DTCO with Circuit-Selective Keep-Out-Zone Utilization for High-Density Heterogeneous 3D Integration

Meng-Lin Wu[1], Zhenliang Pan[2], Yan-Cheng Su[1], Shuhang Lyu[3], Yu-Cheng Lu[1], Han-Lin Huang[1], Yi-Ming Tseng[1], Cheng-Hung Wu[1], Yu-Hsuan Lin[1], Tiwei Wei[2], Vita Pi-Ho Hu[1]
[1]National Taiwan University, [2]University of California, Los Angeles, [3]Purdue University

This work presents a circuit-selective SRAM DTCO framework for high-density heterogeneous 3D (H3D) integration. Case B (TSV perpendicular to channel) reduces five-stage inverter chain delay by up to 18.7% relative to the stress-free baseline. Conversely, Case A (TSV parallel to channel) improves the read/write times of 128×128 SRAM array by up to 18.2%3.87%, respectively. Leveraging these contrasting responses, the proposed DTCO places the SRAM array near TSVs while maintaining peripheral logic in stress-favorable regions. Under high TSV stress (σ_{11} = -640 MPa and σ_{22} = 453 MPa), conventional KOZ-boundary placement incurs substantial area overhead, whereas the proposed SRAM placement increases macro density by 1.11× while reducing read and write times. Experimental Raman measurement further shows reduced residual stress (σ_{11} = -95.7 MPa and σ_{22} = 65.1 MPa), limiting mobility variation within 5%, allowing reduced circuit-to-TSV spacing and 1.21× macro density with negligible variations in read/write times, RSNM, and logic delay.

9:55 AM

14-3 From Interface Defects to SRAM Failure: A Noise-Aware Atomistic-to-Circuit DTCO Path for MoS_2 FETs

Huang Haobo[1], Jia Yuntian[2], Xu Qinzhi[2], Li Zhiqiang[2], Zhao Yudi[3], Lu Ye[1]
[1]Fudan University, [2]Institute of Microelectronics of Chinese Academy of Sciences [3]Beijing Information Science and Technology University

Defects associated with the MoS_2/HfO_2 interface affect both channel transport and dielectric trapping, yet their impact on circuits reliability remains unclear. Thus, we develop a noise-aware atomistic-to-circuit DTCO path. At process level, DFT and KMC are combined to extract defect energetics and spatial distributions, revealing an inverse relation between sulfur vacancies in molybdenum disulfide and oxygen vacancies in hafnium dioxide. At the device level, to link these defect properties to electrical variability and noise, we develop a noise-aware NEGF framework that captures defect-induced transfer-characteristic variability and hysteresis. At the circuit level, we establish a variability-aware SRAM simulation flow using a VAE-based compact model that preserves the NEGF-generated device ensemble. The resulting SRAM simulations identify a defect-regime-dependent transition in the dominant failure mode, from oxygen-vacancy-induced C2C degradation at low sulfur-vacancy density to sulfur-vacancy-induced D2D static-margin loss at high density, providing a mechanism-aware DTCO path for MoS_2/HfO_2 interface, device, and circuit co-optimization.

10:45 AM

14-4 MASS-HBM: A Multiphysics Atom-to-System Simulation Tool for Thermal Reliability-Aware System-Technology Co-Optimization of HBM

Shiyang Li[1], Chenglin Ye[2], Siyuan Chen[2], Nan Tang[2], Yongjia Wang[2], Xiaoyan Liu[2], Peng Huang[2]
[1]School of Integrated Circuits, Peking University, [2]Peking University

We present MASS-HBM, a multiphysics, atom-to-system simulation tool for thermal-reliability-aware system-technology co-optimization (STCO) of HBM. MASS-HBM hierarchically integrates DFT-informed Monte Carlo phonon transport, state-dependent material and interface thermal properties, physics-based DRAM leakage, retention, and aging models, and adaptive-mesh thermal simulation. Using MASS-HBM, we systematically quantify the individual and coupled impacts of interface thermal resistance, material-property variation, and power-temperature feedback across 2.5D, planar-3D, and vertical-3D HBM architectures. We further evaluate a suite of STCO strategies, including package-level liquid cooling, geometry optimization, and thermal-aware workload mapping. The optimized designs increase the thermally allowable HBM capacity from 121 GB to 818 GB while simultaneously enhancing bandwidth scalability, thermal performance, mechanical-stress resilience, and retention reliability.

11:10 AM

14-5 BTI-Limited Endurance of OSFET-Based 2T0C eDRAM for Last-Level Cache: Device-to-System Co-Analysis and Experimental Demonstration

Yue Yu, Kaifeng Wang, Boxin Zheng, Yini Liu, Yuan Wang, Qianqian Huang, Ru Huang

Peking University

This work presents the first aging-aware system-technology co-optimization (STCO) framework for OSFET-based 2T0C eDRAM, bridging device aging and workload-dependent cache lifetime. A cycle-dependent bias-temperature instability (BTI) model, calibrated by stress–recovery measurements, captures the competing recoverable and permanent mechanisms under random bias, temperature and duty conditions. The calibrated V_{TH} shift then propagates through cycle-dependent variables, exposing BTI-limited cell endurance and aging-limited array scalability in closed form. Finally, a Ramulator-based mapping converts AI-workload access patterns into transistor stress, supporting workload-resolved lifetime assessment and guiding heterogeneous-cell co-optimization, with system-level validation confirming the LLC potential of OSFET eDRAM (165×/966× energy-efficiency recovery).

11:35 AM

14-6 Integrating BEOL IGZO eDRAM as On-die Scratchpad for GPUs via System Technology Co-optimization

Mainakh Mukherjee[1], Revanth Koduru[1,2], Arvind Sharma[2], Dwaipayan Biswas[2], Kaushik Roy[1]

[1]Purdue University, [2]imec

Back-end-of-line (BEOL) embedded DRAM (eDRAM) alleviates off-chip memory bottlenecks while restricting front-end-of-line (FEOL) overhead to peripherals. BEOL-compatible indium-gallium-zinc-oxide (IGZO) transistors provide ultra-low leakage, enabling capacitor-less, high-retention gain-cells (4.3s at 85°C). We integrate these IGZO arrays as GPU scratchpads bridging L2 cache and high-bandwidth memory (HBM). By placing them in intermediate BEOL layers (M6–M10 of 7nm node) directly above area-matched FEOL peripherals, M11 and beyond remain preserved for global routing. Under an iso-die constraint, these peripherals displace baseline compute cores (SMs) and L2 cache, dictating a rigorous memory-compute trade-off. We present a system-technology co-optimization (STCO) framework co-designing this reallocation with realizable eDRAM capacity and bandwidth. Utilizing 2T0C and 3T0C macros in an A100-like architecture, we achieve effective densities of 5.30 and 4.05MB/mm^2, yielding scratchpads up to 800 and 600MB. Across LLM workloads, this integration delivers up to 35% energy and 15% latency savings, with refresh overhead below 0.001% of system energy.

Session 15: NC | Probablistics and In-Memory Computing for Robotics and Optimization

9:00 AM, Continental 5

Co-Chairs: Tuo-Hung Hou, National Yang Ming Chiao Tung University and Dmitri Strukov, UCSB

Probabilistic and in-memory computing architectures leverage novel device dynamics to bypass traditional memory bottlenecks, delivering energy-efficient solutions for edge robotics, closed-loop control, and complex optimization. The first paper from Peking University introduces a heterogeneous memristive system combining an 8-Mb RRAM CIM processor with a VO2 stimulation chip for biohybrid robots. The second paper from Zhejiang University demonstrates a multi-timescale pipeline using self-selective memristors for edge robotics. The third paper from the University of Hong Kong presents an all-PWM RRAM CIM architecture with in-loop state memory for closed-loop evolution. The fourth paper from the University of Southern California demonstrates a fully analog system integrating sample-and-hold, MAC, and activation for end-to-end sensing and computing. The fifth paper from Southeast University presents continuous optimization via Markov-chain-based Gaussian p-bits using VCMA-MTJs. The sixth paper from the University of Maryland introduces time-domain p-computing using pulsed sMTJ dynamics and temporal logic. The final paper from Tsinghua University proposes a fast GNN-guided SAT solver using dual-domain RRAM CIM and technology co-optimization.

9:05 AM

15-1 A Heterogeneous Memristive System Integrating an 8-Mb RRAM Perception-to-Policy CIM Processor and a VO_2 Neural Stimulation Chip for Closed-Loop Biohybrid Robots

Longhao Yan[1], Qihang Ding[1], Jiakang Qiu[2], Bo Yang[1], Jinghao Yang[1], Pek Jun Tiw[1], Yuxiang Yang[1], Teng Zhang[1], Daijing Shi[1], Zelun Pan[1], Zhe Zhan[1], Bowen Wang[1], Kechao Tang[1], Yuchao Yang[1]
[1]Peking University, [2]Beijing Normal University

Coupling artificial intelligence with biological adaptability enables autonomous biohybrid robots that combine electronic perception and decision-making with energy-efficient biological actuation. This requires low-power onboard intelligence, biologically compatible neural-signal transduction, and closed-loop regulation of variable responses. We present a heterogeneous memristive system integrating an 8-Mb 16-core RRAM perception-to-policy compute-in-memory (CIM) processor and a VO_2 neural stimulation chip. The RRAM processor performs spatiotemporal thermal perception and generates dual-channel stimulation policies, while the variation-calibrated VO_2 chip converts them into charge-balanced biomimetic bipolar spikes (BBSs) for biological locomotion control. In vivo experiments on Madagascar hissing cockroaches demonstrate closed-loop thermal-target navigation, increasing success rate from 18% to 94% and reducing average task time from 108 s to 54 s. The platform further supports Bluetooth-enabled multi-agent thermal-target search. Compared with an MCU-based implementation, the system reduces task energy by 82.1%, demonstrating a compact and energy-efficient hardware path toward biohybrid embodied intelligence.

9:30 AM

15-2 First Demonstration of a Multi-Timescale Pipeline Using Self-Selective Memristors for Edge Robotics

Xuemeng Fan[1,2,3], Guobin Zhang[1,2,3], Zijian Wang[1,3], Pengtao Li[1,2,3], Zhejia Zhang[1,3], Ding Zou[3], Yiqi Zhou[4], Daying Sun[4], Xiaolei Zhu[1], Bin Yu[1], Qing Wan[2], Dawei Gao[1,3], Yishu Zhang[1,2,3]
[1]Zhejiang University, [2]Yongjiang Laboratory, [3]Zhejiang ICsprout Semiconductor Company Limited, [4]Nanjing University of Science and Technology

Edge robotics faces severe latency and energy constraints due to the von-Neumann bottleneck. This work presents a brain-inspired processing system that overcomes this limitation using self-selective memristor arrays. By integrating three SSM types, including short-term volatile, long-term volatile, and non-volatile, our hardware embeds temporal filtering, working memory, and synaptic storage directly within the compute fabric. The system orchestrates seven cortical-like regions across five parallel arrays to execute a real-time six-phase gesture-to-action pipeline. This architecture eliminates external memory accesses, achieving 53× energy efficiency and 28× latency reduction over an edge GPU, establishing a scalable pathway for complex cognition on autonomous agents.

9:55 AM

15-3 All-PWM RRAM Compute-in-Memory with In-Loop State Memory for Energy-Efficient Closed-Loop State Evolution

Jichang Yang[1], Xinyuan Zhang[1], Songqi Wang[1], Hegan Chen[1], Zimo Li[1], Shuang Liang[1], Hang-Ming Zhang[1], Ning Lin[1], Yi Li[1], Zhongrui Wang[2], Han Wang[1]
[1]THE UNIVERSITY OF HONG KONG, [2]Southern University of Science and Technology

Modern AI is shifting from static inference toward dynamic state evolution in emerging network architectures and generative applications. Existing RRAM CIM hardware remains optimized for isolated MVM, requiring repeated A/D and D/A conversions and state movement across solver steps. We present PRIDE, a pulse-width-modulated (PWM) RRAM in-memory differential equation solver with PWM-encoded inter-module dataflow and recurrent feedback, and capacitive in-loop state memory. Dual-mode charge integration supports incremental and normal MVM, while pulse-width logic implements nonlinear activation. Fixed-voltage, single-cycle charge-domain MAC suppresses RRAM nonlinearity and averages read noise without in-loop ADC/DAC operations. The experimentally demonstrated prototype implements a state-space model (SSM) and flow matching (FM) for pedestrian trajectory prediction, achieving near-software average displacement errors (ADEs) of 0.41 m and 0.29 m, respectively, with 41-fold and 51-fold energy reductions relative to a GPU. PRIDE thereby extends RRAM CIM beyond isolated MVMs to closed-loop state evolution at the edge.

10:20 AM

15-4 A Fully Analog System Integrating Sample-and-Hold, MAC, and Activation for Efficient End-to-End Sensing, Processing, and Computing

Zixu Wang[1], Yichun Xu[1], Maysara Hamada[1], Mostafa Toubar[1], Wenhao Song[1], Max Zhang[2], Miao Hu[2], Kangjun Bai[3], Sabyasachi Ganguli[3], Michael R. Page[3], Qing Wu[3], Mike Shuo-wei Chen[1,2], J. Joshua Yang[1,2]
[1]University of Southern California, [2]TetraMem Inc., [3]Air Force Research Laboratory

While analog multiply-accumulate (MAC) on 1T1R memristor arrays is well established, reported systems remain bounded by data converters. Thus, the sensing, signal-processing, and computing stages of an analog chain remain disjoint, unlike in biological systems. Here we present a fully analog system that unifies them, operating end-to-end from a live analog input to ReLU-activated analog outputs. A 64-channel sample-and hold (S&H) front-end samples at 320 MS/s, driving rows of a 128×64 1T1R array directly. An array of transimpedance amplifiers with embedded ReLU (ReLU-TIA) reads 32 columns in parallel, producing voltage-domain activations compatible with next-layer row inputs, thereby enabling analog inter-layer communication. The compute path is free of DACs, ADCs, and digital activation logic, and sustains continuous end-to-end analog data flow for real-time signal processing. We characterize MAC accuracy under DC and streamed inputs, and demonstrate hardware RF classification at 76.73% accuracy, a 3.08% absolute drop from the 79.81% software baseline.

11:10 AM

15-5 From Discrete to Continuous Optimization with Markov-Chain based Gaussian Probabilistic Bits using VCMA-MTJs

Huanghui Wang[1], Chao Fang[2], Yaoru Hou[3,4], Yihan He[2], Xiangyi Meng[1], Yantong Di[1], Shirui Zhao[3], Zhihua Xiao[3,4], Xiao Gong[2], Bo Liu[1], Gengchiau Liang[5], Qiming Shao[3,4], Hao Cai[1]
[1]Southeast University, [2]National University of Singapore, [3]Hong Kong University of Science and Technology (HKUST), [4]AI Chip Center for Emerging Smart Systems, [5]National Yang Ming Chiao Tung University

Real-world combinatorial optimization problems increasingly involve both binary decisions and continuous variables, challenging probabilistic solvers such as Ising machines with binary operation. This work introduces a Markov-chain (MC) based Gaussian probabilistic bit (G-Bit) for mixed discrete–continuous optimization problem. By decomposing a target Gaussian distribution into sequential conditional probabilities, a single voltage-controlled magnetic anisotropy magnetic tunnel junction (VCMA-MTJ) probabilistic bit (*p*-Bit) is time-multiplexed to generate an N-bit Gaussian state. G-Bit generation is validated on a VCMA-chip prototype and applied to mixed-variable and continuous optimization tasks. Compared with the binary QUBO baseline, the proposed approach improves solution accuracy while reducing computational complexity. A projected 28-nm CMOS integrated with VCMA-MTJ ASIC achieves a core area of 4876 μm^2 and a 6.25 MSPS sampling rate at 100 MHz, representing a 208× improvement over prior sMTJ-based G-Bit work. These results establish device-native G-Bit computing as a scalable approach for mixed-variable and continuous optimization.

11:35 AM

15-6 Time-domain p-computing: Pulsed sMTJ Dynamics and Temporal Logic for Synchronous

Temitayo Adeyeye[1], Matthew Daniels[2], Sidra Gibeault[1], William Borders[2], Mark Stiles[2], Jabez McClelland[2], Matt Carey[3], Chunhui Dai[3], Patrick Braganca[3], Daniel Lathrop[1], Advait Madhavan[2]
[1]University of Maryland, [2]National Institute of Standards and Technology, [3]Western Digital

Probabilistic computing (p-computing) frame works solve energy-based models by sampling from Boltzmann distributions over accumulated model energies. Conventional voltage-domain sampling of stochastic magnetic tunnel junctions (sMTJs) faces conflicting system-level tradeoffs. Digital implementations require many stochastic transitions per state update, demanding complex materials and fast de vices. Conversely, “instantaneous-interactions” in analog asynchronous designs constrain sMTJs to operate slower than resistive crossbars, inflating latency and energy. We propose a digital-friendly time-domain architecture leveraging sMTJ’s single-well-escape physics for faithful exponential sampling from a single switching event alongside delay-based temporal logic for energy accumulation. The time domain enables low-activity-factor operation, achieving low latencies without complex device engineering. We experimentally demonstrate a synchronous temporal p-compute unit cell performing annealing over a 32× timescale range. 7 nm simulations demonstrate advantages over conventional approaches of 5× in power and 56× in area, establishing a scalable pathway for integrated p-computing.

12:00 PM

15-7 A Fast Graph Neural Network-Guided SAT Solver for Combinatorial Optimization via Dual-Domain RRAM CIM and System–Technology Co-Optimization

Junyang Zhang[1], Peng Yao[1], Siyao Yang[1], Songtao Wei[1], Qiumeng Wei[1], Keer Liu[1], Binbin Cheng[1], Xiangrui Li[1], Zhixing Jiang[1], Yanbo Su[1], Xiangchao Ma[2], Kun Wang[2], Haixia Yu[3], Chao Yue[3], Xingjun Wu[1], Liyang Pan[1], Dong Wu[1], Zhigang Zhang[1], Jianshi Tang[1], Bin Gao[1], He Qian[1], John Paul Strachan[4], Huaqiang Wu[1]
[1]Tsinghua University, [2]Beijing InnoMem Technologies Co., Ltd., [3]Tongxin Microelectronics Co., Ltd., [4]Forschungszentrum Jülich

Boolean satisfiability (SAT) is a canonical formulation for combinatorial optimization problems (COPs) in scheduling, routing, resource allocation, and graph optimization applications. Compared with feed-forward AI inference, SAT repeatedly evaluates irregular variable–clause dependencies, incurring costly iterations and data movement. This work proposes the first GNN-guided hardware SAT solver for fast and efficient combinatorial optimization by introducing dual-domain RRAM compute-in-memory (CIM) and system–technology co-optimization (STCO). Algorithmically, a two-stage method combining GNN initialization and GNN-guided ProbSAT search is adopted to accelerate convergence. At the hardware level, RRAM characteristics and array circuits are optimized in dual-domain CIM macros for fast GNN and efficient ProbSAT solver, respectively. Based on this dual-domain CIM system, the hardware solver achieves 100% solvability with a state-of-the-art 5.2-μs time-to-solution (TTS) and 11.5-nJ energy-to-solution (ETS) on 50-218 3-SAT benchmarks, supporting demonstrations of Max-Cut solving and automated circuit-layout generation.

Session 16: EDT | Novel Computing Paradigms

9:00 AM, Continental 6

Co-Chairs: Saptarshi Das, Penn State and Jens Trommer, Namlab

This session explores emerging approaches to computing and data processing in silicon-based and related technologies, spanning secure computing, stochastic and probabilistic computing, quantum technologies, thermal management, superconducting electronics, and the sustainability of future computing platforms. The first paper presents the monolithic integration of emerging reconfigurable transistors with FDSOI technology for secure edge applications. The second paper demonstrates a Gaussian random-number generator that exploits band-to-band tunneling in a single FDSOI transistor. The third paper introduces a multi-tier diamond scaffolding approach for thermal management in densely integrated 3D circuits. The fourth paper demonstrates dense computational logic using nonlinear polar materials, enabling compact and potentially energy-efficient logic functionality beyond conventional CMOS. The fifth paper assesses the critical factors governing qubit-control electronics for quantum computing and provides practical guidelines for future device design. The sixth paper presents an ultrafast resonant-tunneling-diode-based p-bit for reliable probabilistic computing at cryogenic temperatures. The seventh paper demonstrates NbTiN superconducting resonators with three metal layers fabricated using a scalable 300 mm process for superconducting electronic circuits. Finally, the eighth paper provides a comparative assessment of emerging FET technologies, quantifying trade-offs among power, performance, area, and total carbon footprint for future VLSI logic.

9:05 AM

16-1 Monolithic CMOS Integration of Reconfigurable FET Hybrid Circuits for Secure Edge Applications

Violetta Sessi[1], Giulio Galderisi[2], Niladri Bhattacharjee[1], Brojogopal Sapui[2], Yuxuan He[2], Juan Pedro Martinez[2], Jose Maria Gonzalez Medina[3], Ananya Karmakar[4], Annekathrin Zeun[1], Kerstin Poenisch[1], Konstantin Li[1], Binit Syamal[1], Dirk Utess[1], Fernando Koch[1], Michael Zier[1], Peter Baars[1], Sophie Schallert[3], Viktor Havel[2], Stefan Slesazeck[2], Oskar Baumgartner[3], Alexander Kloes[4], Thomas Mikolajick[2,5], Peter Javorka[1], Jens Trommer[2]
[1]GLOBALFOUNDRIES, [2]NaMLab gGmbH, [3]Global TCAD Solutions GmbH, [4]THM University of Applied Sciences, [5]TU Dresden

This work demonstrates a novel technology solution for low-power edge applications. Thereto, three-independent-gate reconfigurable FETs have been integrated monolithically together with CMOS transistors for the first time, without negatively affecting the FDSOI baseline technology. Both transistor types work seamlessly together to create hybrid logic cells with polymorphic functionality. A TCAD design-technology co-optimization study was carried out to analyze the impact of channel length scaling and extracting parasitics. A related compact model, for

the first time including intrinsic capacitances, was developed to enable standard-cell characterization. Finally, the technology's potential is illustrated in a PRESENT substitution-box design, providing a cipher block with increased resistance to a variety of side-channel attack scenarios.

9:30 AM

16-2 Breaking the Entropy Wall: A Single-Transistor FDSOI Gaussian Random Number Generator Exploiting Band-to-Band Tunneling

Pengyu Ren[1], Xingtian Wang[1], Likai Pei[1], Jiahao Zheng[1], Boyang Cheng[1], Jianbo Liu[1], Jiahui Duan[1], Sizhe Ma[1], Yixin Qin[1], Halid Mulaosmanovic[2], Stefan Duenkel[2], Jérôme Mazurier[2], Sven Beyer[2], Thomas Kämpfe[3], Yiyu Shi[1], Xiaobo Sharon Hu[1], Ningyuan Cao[1], Kai Ni[1]

[1]University of Notre Dame, [2]GLOBALFOUNDRIES, [3]Fraunhofer Institute for Photonic Microsystems

Artificial intelligence is increasing demand for high-throughput entropy sources for uncertainty-aware computing and this demand is especially important for edge systems operating in uncertain environments. However, conventional Gaussian random number generators (GRNGs) have over 10^4× lower throughput density than modern computing hardware, creating an "entropy wall."

This work demonstrates a single-transistor GRNG on an FDSOI platform using band-to-band tunneling (BTBT). Stochastic hole injection modulates the body potential and produces a Gaussian drain-current distribution. BTBT-generated entropy is over 10× larger than read noise. Device and bias optimization achieves a 0.01 μm^2 cell area, 4 fJ/sample energy, sub-20 ns latency, endurance above 10^{10} cycles, less than 20 mV threshold-voltage shift after 10^4 s at VOV = -2V and operation at 80 °C. A temporal-differential circuit further suppresses drift and supports accurate probabilistic inference. These results show that BTBT-based FDSOI can provide compact, energy-efficient, and scalable entropy generation for probabilistic AI hardware.

9:55 AM

16-3 BEOL-Compatible 2-Tier Diamond 3D Thermal Scaffolding for Non-Aligned Cooling of Dense 3D ICs

Junrui Lyu[1], Eda Deniz Demirel[2], Kaiyan Shan[2], Zhengliang Bian[2], Navid Anjum Aadit[2], Anna Kasperovich[2], Dennis Rich[2], Mohamadali Malakoutian[2], Subhasish Mitra[2], Srabanti Chowdhury[2]

[1]Electrical Engineering, Stanford University, [2]Stanford University

In dense 3D ICs, heat extraction through conventional copper thermal vias depends on precise inter-tier alignment, constraining manufacturability and design. 3D thermal scaffolding provides a non-aligned cooling architecture by combining a thermally conductive dielectric for lateral heat spreading with vertical thermal vias for heat extraction. Here, we experimentally demonstrated the first 2-tier, BEOL-compatible, void-free polycrystalline-diamond (PCD) thermal scaffolding architecture enabling non-aligned 3D thermal management. Using selective nucleation and ion-miling, we stack multiple PCD layers, achieving conformal, near-isotropic PCD dielectric layers and 2-μm-diameter PCD thermal vias with low surface roughness (~18 nm). We quantify the relationship between via non-alignment length (ΔL) and hotspot temperature-rise penalty (ΔT), limiting ΔT to 5 K at ΔL = 60.8 μm and 10 kW/cm^2. A calibrated model projects 12-tier AI accelerator and RISC-V designs, reducing worst-case ΔT by up to 12.3× and increasing tolerable ΔL by up to 4.2× versus copper.

10:45 AM

16-4 A Dense Computational Logic Technology with Nonlinear Polar Materials

Sasikanth Manipatruni[1], Neal D. Reynolds[1], Rafael Rios[1], Peter Meisenheimer[1], Ikenna Odinaka[1], James Clarkson[1], Abel Fernandez[1], Akshata Naidu[1], Tyler Whyland[1], Noriyuki Sato[1], Meera Menezes[1], Gaurav Thareja[1], Vishantak Srikrishna[1], Darshak Doshi[1], Tanay Gosavi[1], Antun Domic[1], Paul Hartke[2], Ralph Wittig[2], Rajeev Dokania[1], Debo Olaosebikan[1], Ramamoorthy Ramesh[1], Dmitri E. Nikonov[1], Amrita Mathuriya[1]

[1]Kepler Computing, [2]Advanced Micro Devices, Inc. (AMD)

We outline a logic densification technology that utilizes a quasi-paraelectric (QPE) in the BEOL coupled with standard FEOL CMOS transistor and lithography technology for a VLSI-compatible majority gate logic. Experimentally, we demonstrate the feasibility of these Nonlinear Polar Majority (NLPM) gates in QPE with unconditional endurance (>10^15 cycles at 4V accelerated voltage testing) and <50ps switching time at <1V. We also develop the logic synthesis tools to show how the high expressivity of MAJ-n gates (n=3,5,7) reduces the circuit size for an equivalent CMOS function and provides >4x density improvements (>2 generations of traditional scaling) to enable logic density beyond 1000MXtors/mm2. For 4-bit arithmetic as a benchmark in AI/GPU, we find a

8.6x power-performance-area (PPA) improvement for 4-bit adders and 3.2x PPA improvement for 4-bit multipliers for the sharpest QPE measured with Vx=129mV. Due to the CMOS driver in MAJ-n logic, improving transistor roadmaps will continue to benefit MAJ-n technology.

11:10 AM

16-5 Investigating Impacts of Control-Electronics Non-idealities on Silicon Spin Qubits Quantum Gate Performance Using Quantum–Classical Co-Simulation

Rubaya Absar[1,2], Bingjian Du[3], Zach D. Merino[3], Chenao Yuan[3], Bohdan Khromets[3], Shiyu Su[3], Jonathan Baugh[3], Lan Wei[3]

[1]Electrical and Computer Engineering, University of Waterloo, [2]Institute for Quantum Computing, University of Waterloo, [3]University of Waterloo

Fault-tolerant quantum computing requires classical control electronics capable of generating precise qubit-control signals. However, conventional circuit-design metrics, including accuracy, noise, linearity, and timing performance, provide limited insights into their impacts on quantum-gate performance, making it difficult to derive practical design specifications and potentially leading to under- or over-designed control circuits. This work establishes quantitative design guidelines for cryogenic CMOS control electronics using a SPICE-compatible quantum--classical co-simulation framework that directly links circuit-generated control waveforms to quantum-state evolution and gate fidelity in scalable silicon spin-qubit quantum processors. Representative single- and two-qubit gate operations are analyzed to quantify the impact of static voltage errors, stochastic noise, DAC non-idealities, and timing imperfections. The results reveal that static voltage errors together with low-frequency noise sources, including random telegraph noise and $1/f$ noise, dominate fidelity degradation, whereas DAC quantization, current-source mismatch, jitter, and settling effects contribute negligibly over practical operating ranges.

11:35 AM

16-6 Achieving Reliable Cryo-Compatible Probabilistic Computing with Ultrafast Resonant-Tunneling-Diode P-Bit

Chenhao Xia[1], Yuze Lu[1], Hongjie Zeng[1], Weican Chen[1], Chengming Zhang[1], Pingyang Huang[1], Hongbo Wang[1], Ding Wang[1], Wei Tan[2], Xiaoyan Liu[1], Juan Su[2], Runsheng Wang[1], Fei Liu[1]

[1]Peking University, [2]China Academy of Engineering Physics

Probabilistic computing based on cryo-compatible resonant-tunneling-diode (RTD) p-bit is demonstrated for optimization and low-latency cryogenic quantum-error-correction workload for the first time. A Poisson-NEGF-Monte Carlo model predicts nearly 10^2 noise enhancement with GHz-scale entropy bandwidth, enabling high-speed, low-correlation p-bit generation. Experimentally, AlAs/$In_{0.8}Ga_{0.2}As$/AlAs RTDs preserve NDR from 8K to 300K, exhibit broadband super-Poissonian noise, and remain stable after 10^{12} pulse cycles, confirming a temperature-robust and reliable entropy source. A 25-unit RTD parallel p-computer further demonstrates seven-level probability calibration, <5% probability drift after 10^{12} pulse cycles, NIST-qualified bitstreams at 8K and 300K, and circuit-limited 10 MHz operation. Using this 25-unit p-bit circuit, we experimentally demonstrated a sparsified 48 p-bit decoder that stably solves 8 qubits surface-code quantum-error-correction (QEC) instances within 10μs for 10 hardware runs, and moreover, extended to a fault-tolerant quantum-computing application scenario, yielding high *T*-state throughput and $<10^{-5}$ propagated VQE electron-density error.

12:00 PM

16-7 NbTiN-based 3 metal layer resonators using a 300 mm process for superconducting electronics

Kamal Brahim, Gilles Delie, Blake Hodges, Benjamin Huet, Oleksandr Hryhorenko, Julian Gil Pinzon, Steven Brebels, Peter Nicholas Koufalis, Ankit Pokhrel, Adham Elshaer, Daniel Perez Lozano, Sujan Kumar Sarkar, Luke Zhang, Steven Deckers, Seifallah Ibrahim, Sabine O'Neal, Richard Rouse

imec

Superconducting electronics are a promising solution for energy-efficient high-performance computing (HPC), quantum computing readout/control, and cryogenic sensing, requiring low-loss passive components and scalable interconnects. This work presents the fabrication and characterization of material resonators using a planarized three-metal layer (3ML) NbTiN process that enables compact multilayer integration with 60 nm minimum feature sizes and 200 nm line pitch. Cryogenic measurements between 2.6 K and 4 K demonstrate robust superconducting

performance, with effective loss tangents on the order of tan δ ~ 1e-4. Losses are primarily attributed to the bulk and interface contributions of the (SiO2 and SiN) dielectric fill within the stack. These results highlight the viability of NbTiN multilayer platforms for dense superconducting RF and digital circuits targeting HPC systems.

12:25 PM

16-8 Quantifying Trade-Offs in Power, Performance, Area, and Total Carbon Footprint (PPAtC) of Next Directions in FET Technologies for VLSI Digital Logic Circuits

David Kong[1], Danielle Grey-Stewart[1], Mariam Elgamal[1], Zhenyang Chen[2], Zhengyin Piao[2], Jalil Morris[1], Yuan Yao[2], Gage Hills[1]

[1]Harvard University, [2]Yale University

Computing's carbon footprint is a major consideration when designing future VLSI circuits, accounting for embodied carbon ($C_{embodied}$) due to fabrication, and operational carbon ($C_{operational}$) due to use. We quantify trade-offs in power, performance, area, and total carbon footprint of advancing from FinFETs to: (1) Nanosheet (NSH) FETs; (2) Sequential Complementary FETs (sCF); and (3) Monolithic Complementary FETs (mCF). We develop detailed models for $C_{embodied}$, accounting for energy, material, and gas use from >500 fabrication process steps. We then leverage VLSI design flows to create processor core designs with each FET technology, to quantify power, performance, area, and $C_{operational}$. Despite higher $C_{embodied}$ per wafer, $C_{embodied}$ per die improves for advanced FET technologies, due to smaller circuit area. Cembodied per die is 2.13× better for NSH, 2.08× better for sCF, and 2.44× better for mCF. Thus, advanced FET technologies can improve carbon efficiency, despite their more complex fabrication processes.

Session 17: FS | Semiconductors for Quantum Computing

9:00 AM, Continental 7-9

Co-Chairs: Benoit Bertrand, CEA-Leti and Changhwan Shin, Korea University

This Focus Session addresses the manufacturing path to large-scale quantum processors: 300 mm CMOS qubit fabrication and yield, cryogenic CMOS control, and superconducting packaging. It matters because scaling is now limited by materials, interconnects, and footprint mismatch rather than qubit physics, making industrial integration the decisive bottleneck.

9:05 AM

17-1 Scalable Quantum Computing in CMOS: Materials and Integration Perspectives

Benjamin Lilienthal-Uhlig[1], Maximilian Lederer[1], Marcus Wislicenus[1], Felix Mende[1], Sergej Liberda[1], Oliver Ostien[1], Conrad Guhl[1], Katharina Lilienthal[1], Yannick Raffel[1], Maik Simon[1], Varvara Brackmann[1], Johannes Hertel[1], Steffen Bickel[2], Pascal Muster[3,4], Wolfram Langheinrich[3], Arnau Sala[4], Lars R Schreiber[4], Hendrik Bluhm[4]

[1]Fraunhofer Institute for Photonic Microsystems, [2]Fraunhofer Institute for Reliability and Microintegration, [3]Infineon Technologies, [4]RWTH Aachen University

Over the past decade, quantum computing has transitioned from predominantly academic research to an emerging industrial field. This shift has been driven by steady improvements in qubit fidelity and coherence, rapid scaling in qubit count but also significant public and private investment. In this context, the semiconductor fabrication ecosystem, already central to the global technology landscape and further elevated by the recent AI boom, has emerged as a key enabler for scalable quantum computing. In this talk, we examine how large-scale quantum computing systems can be realized using microelectronics fabrication approaches and identify the critical challenges that must be addressed for industrial implementation. We argue that scalable quantum processors will require tightly integrated, CMOS-compatible solutions spanning materials, interconnect technologies, and device architectures. As illustrative examples, we discuss thermal management strategies based on CMOS-compatible Bragg-reflector multilayers and advanced wiring approaches enabled by monolithic interposer concepts. At the device level, we highlight recent progress in silicon spin shuttling architectures and emphasize the complex interplay between materials engineering, process integration, and device design. These results outline key pathways and remaining bottlenecks toward the realization of manufacturable, large-scale quantum computing systems.

9:30 AM

17-2 28 FD-SOI Based Qubit/Logic 300mm Platform for Quantum Processor Unit Chip Compatible with Volume Manufacturing

Franck Arnaud[1], bruna Cardoso[2], Etienne Novak[2], Carlos Suarez-Segovia[3], Giselle Elbaz[2], Nicolas Daval[2], Maud Vinet[2], Tristan Meunier[2], Miltiadis Alepidis[4], Thomas Bedecarrats[4], Jose Lugo-Alvarez[4], Mikael Casse[4], Aurelie Bajolet[4], Flavio Enrico Bergamaschi[4], Victor Sabia Pereira Carpes[5], Philippe Galy[3], Florence Sonnerat[3], Magali Gregoire[3], Angelo Moroni[3], Patrick Torresani[2]

[1]R&D, STMicroelectronics, [2]QUOBLY, [3]STMicroelectronics, [4]CEA-Leti, [5]Université Grenoble Alpes

We present for the first time a full CMOS platform for Quantum Processor Unit (QPU) based on 28nm FDSOI commercial technology in 300mm fab, so called QSOI®. Optimized process modules such as triple-gate-oxide, dual contacted-poly-pitch and un-landing contact, enabling for the first time, a control of the Spin Qubit by the 'adjacent' CMOS integrated circuit. Leveraging superior advantage of the back-bias, signal contrast of the Coulomb peaks has been significantly improved. Rabi oscillations have been observed for a single electron spin thanks to a magnetic field generated with co-integrated RF antenna. Our platform provides the entire devices suite, including all the passives components. Thanks to back-bias, the $V_T^{4K} \sim V_T^{300K}$ for MOS transistors, improving by +50% the ring-oscillators performance at low temperature. QPU circuit has been designed with our cryo-design-kit, incorporating qubit control, RF antenna and qubit readout based on charge-capacitance technique. This fully integrated technology can easily migrate on superconducting wires for future quantum chip performance enhancement

9:55 AM

17-3 Cryo-optimized CMOS for ultra-low temperature applications

Janne Lehtinen, Nikolai Yurttagül, Cameron Lauvitel, Md Najmussadat, Iago Rossetto, Janne Ikonen, Jan Toivonen, Esa Tuovinen, Henrik Romppainen, Arto Rantala, Mika Prunnila, Markku Kainlauri

SemiQon

Cryogenic CMOS has emerged as a key enabler for scaling solid-state quantum computers from hundreds to hundreds of thousands of qubits. It offers a proven industrial pathway for large-scale integration and remains the only technology to successfully demonstrate such scalability. CMOS technology can deliver the stability and uniformity required for cryogenic electronics. However, most existing cryogenic CMOS approaches rely on conventional room-temperature designs operated at low temperatures, which limits the benefits that could be achieved through manufacturing and circuit design optimized for low temperatures. In this article, we present examples illustrating some of these benefits and outline a path toward ultra-low drive-voltage circuits—an essential step for scaling cryogenic control systems beyond small-scale prototypes.

10:45 AM

17-4 Type II InAs/GaSb heterostructures a promising material platform for quantum computing

Wei Pan[1], Aaron Muhowski[1], Joseph J. Cuozzo[2], Enrico Rossi[3]

[1]Sandia National Laboratories, [2]University of Texas at El Paso, [3]William & Mary

There have been intense research efforts on identifying Majorana zero modes (MZMs) for topological quantum computing (TQC). Most of the current approaches to TQC, however, require applying a magnetic field, which pose challenges to scalability and operation of TQC. Consequently, realizing MZMs without the need for magnetic fields is highly desirable. One promising alternative involves Majorana Kramers pairs—pairs of MZMs that can be used to form Majorana Kramers qubits (MKQs). Notably, MKQs benefit from preserved time-reversal symmetry, a larger topological gap and, thus, enhanced protection against quasiparticle poisoning, making them a compelling alternative to conventional MZM-based qubits. Here, we will present our results on helical edge state transport in an InAs/GaSb quantum spin Hall insulator with a superconducting constriction. We demonstrate that this device structure can enable the realization of Majorana Kramers pairs without the application of magnetic fields.

11:10 AM

17-5 Programmable Electro-Optic Quantum Photonic Circuits for Hybrid Quantum Simulation

Hyounghan Kwon[1], Uijoon Park[2], Jin-il Lee[2], Sang Uk Jeong[3], Sunghyun Moon[2], Minho Choi[2], Hyeong-Soon Jang[2], Min-Kyo Seo[4], Jinhyung Lee[2], Dae-Hwan Ahn[2], Sang-Wook Han[2], Hyang-Tag Lim[2], Yong-Su Kim[2], Jaewoo Park[3], Hojoong Jung[2]

[1]Center for Quantum Technology, Korea Institute of Science and Technology, [2]Korea Institute of Science and Technology, [3]Chungbuk National University, [4]Korea Advanced Institute of Science and Technology (KAIST)
Scalable high-speed quantum photonic hardware is essential for practical photonic quantum systems, yet conventional integrated photonics rely on thermo-optic phase shifters constrained by slow speed and high power dissipation. Here, we demonstrate programmable electro-optic quantum photonic circuits that address the dual challenges of source integration and module packaging for hybrid quantum simulation. The platform is experimentally validated through two complementary demonstrations First, we achieve the first electro-optic implementation of an integrated variational quantum eigensolver combining on-chip periodically poled nonlinear photon sources with reconfigurable interferometric meshes to estimate molecular ground-state energy within chemical accuracy. Then, we show a fiber-pigtailed processor module enabling robust hybrid quantum simulations. These results establish thin-film electro-optic devices as a viable hardware platform for next-generation quantum processors.

11:35 AM

17-6 Qubit-Device Yield & Validation for a Silicon Exchange-Only Quantum Processing Unit

Clayton A C Jackson, Bryan J Thomas, Tyler A Cain, Max D Choi, Dan R Ward, James M Chappel, Sieu D Ha, Sam J Whiteley, Trevor M Fowler, J P dodson, June Suh, Christi M Peterson, Mike P Jura, Aaron M Jones, Andrew S Pan, Thaddeus D Ladd, Matt G Borselli
HRL Laboratories, LLC
Commercially viable quantum processors require qubit technologies that combine high performance, manufacturability, and ability to scale. Silicon exchange-only (EO) qubits present a compelling path due to their baseband electrical control and compatibility with advanced semiconductor fabrication. To demonstrate feasibility of this architecture, EO qubits need to have sufficiently high yield and be operable via scalable control architectures. Here, we present qubitdevice wire yield, uniformity analysis and quantum yield for a 54-dot, 18-qubit Si/SiGe EO qubit array designed for integration within a custom designed and fabricated cryogenic-CMOS controlled quantum processing unit (QPU). We present wafer-level per-wire yield ("wire-functional"), electron loading voltage distributions, and two-dimensional maps of qubit singlet return probabilities ("quantum-functional yield") within an 18-qubit array. These results establish that the qubit device provides a foundation for scalable, silicon-based fault tolerant quantum computing.

Session 18: RSD | Reliability of Ferroelectric Materials and Devices

9:00 AM, Imperial A
Co-Chairs: Kai Ni, U. of Notre Dame and Reika Ichihara, Kioxia
This session contains 8 papers investigating different reliability aspects of ferroelectric materials and devices. This first sub-session explores novel processing routes, phase engineering, and structural optimization to advance scaled ferroelectric device integration. Paper 1 uncovers the phase-evolution mechanisms driving wake-up in 3 nm HZO and demonstrates a TiN re-deposition strategy to achieve immediate, wake-up-free low-voltage operation. Paper 2 presents a 3D vertical ferroelectric tunneling junction array using sub-2 nm HZO combined with sidewall plasma-oxygen treatment to achieve endurance beyond 10^12 cycles. Paper 3 introduces a low-temperature (≤200 C) field-directed crystallization process that directly forms c-axis-oriented orthorhombic HZO, yielding a Quasi-rectangular P-V loop and record 2Pr of 75 μC/cm2. Paper 4 investigates DC and AC time-dependent dielectric breakdown in ultra-thin HZO, uncovering a functional bilayer structure controlled by an interfacial TiOxNy layer that dictates high-field reliability. The second sub-session focuses on understanding and mitigating endurance degradation, defect dynamics, and polarization fatigue in HfO2-based ferroelectric capacitors. Paper 5 establishes a predictive fatigue model using a single I-V measurement by demonstrating that total polarization remains conserved while switchable domains convert to non-switchable ones. Paper 6 identifies charged oxygen-vacancy dynamics as the primary driver of duty-cycle degradation and offers strain- and bias-tuning mitigation strategies. Paper 7 reveals that a first-cycle oxygen vacancy surge accounts for half of all early leakage in 5 nm HZO, presenting a dual WOx oxygen reservoir layer to achieve 10^10-cycle endurance at 145 C. Paper 8 introduces an oxygen-gradient engineering approach using graded ZrO2 interlayers to stabilize the polar phase, delivering 1.6×10^13-cycle endurance in capacitors and high-speed operation in 2D MoS2 FeFETs.

9:05 AM

18-1 Unraveling the Wakeup Mechanism in Ultra-Thin (~3 nm) HfO_2-based Ferroelectric Films: Synergistic *Pbca*-to-*Pca2₁* Antiferroelectric Loop Depinching and Tetragonal-Phase-Evolution-Driven Pr Enhancement

Xiaopeng Li[1], Chaoming Wu[1], Xiaoyu Dou[2], Dong Zhang[1], Yang Feng[1], Xiaolin Wang[1], Yaoyu He[2], Pengpeng Sang[2], Jixuan Wu[2], Ning Ji[1], Duy Hieu Trinh[1], Yi Tong[1], Xinming Li[1], Zhilun Zhang[1], Gengchiau Liang[3], Jiezhi Chen[2], Xiao Gong[1]

[1]National University of Singapore, [2]Shandong University, [3]National Yang Ming Chiao Tung University

For the first time, we elucidate the wakeup mechanism in ultra-thin (~3 nm) HfO_2-based ferroelectric (FE) films, addressing a fundamental bottleneck in low-voltage, high-density FE technologies. We reveal that the wakeup effect originates from two defect-driven, synergistic phase-evolution processes: (1) *Pbca*-to-*Pca2₁* transition that depinches the antiferroelectric (AFE) loop, and (2) t-to-*Pbca*-to-*Pca2₁* transition that enhances remanent polarization (Pr). Guided by this understanding, we propose and experimentally demonstrate a TiN re-deposition strategy that directly activates AFE loop depinching without electrical wakeup cycling. The resulting BEOL-compatible 3 nm-thick HZO capacitor exhibits wakeup-free operation, a 2Pr of ~27 μC/cm^2 at 1 V, complete polarization switching under low-voltage and high-frequency operation (1 V, 1 MHz), and a record-high breakdown voltage of ~3.3 V, enabled by the retained t-phase fraction. This work establishes a direct mechanism-to-device pathway toward reliable, low-voltage ultra-thin HZO capacitors for high-density 3D FE memory integration.

9:30 AM

18-2 Highly Reliable Ferroelectric HZO with Record 2P_r of 75μC/cm^2 and Quasi-rectangular P-V through *in-situ* Field-directed *c*-axis-oriented o-phase Crystallization at Low-Temperature (≤200℃)

Lingxuan Zou[1], Shengjie Cao[1], Jiayu Li[2], Mingxi Jin[1], Jiayan Zhu[1], Yueyang Liu[2], Zhiyuan Fu[3], Qianqian Huang[1], Ru Huang[1]

[1]Peking University, [2]Institute of Semiconductors, Chinese Academy of Sciences, [3]Southeast University

For the first time, this work demonstrates a highly reliable ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ (HZO) capacitor with quasi-rectangular P-V, a record-high 2Pr of 75μC/cm^2 and abrupt switching via low-temperature (≤200℃) field-directed crystallization (LTFDC), free of high-temperature annealing. Enabled by in-situ direct crystallization of amorphous HZO into the c-axis-oriented o-phase, the device shows decreased leakage and 10-year retention at 125℃ and a quasi-rectangular P-V retained after 10^{10} cycles. Combining calculations, characterizations, first-principles and phase-field simulations, we reveal that the field selectively lowers the polar o-phase nucleation barrier with out-of-plane c-axis preference, yielding well-textured pure-o grains that lower the switching barrier and suppress oxygen-vacancy pinning, while the ultra-low temperature reduces bulk oxygen vacancies. Based on these, LTFDC delivers 2Pr of 48μC/cm^2 at 1V in scaled 5nm HZO for 3D 1T1C FeRAM and disturb-free operation for nC FeRAM, offering a BEOL-compatible route toward next-generation ferroelectric memories.

9:55 AM

18-3 3D CMOS-Compatible Vertical FTJ Array with Record-Thin 1.x-nm Ferroelectric HZO and Record Endurance (>10^{12}-cycles)

Renhao Xue, Ruizhan Yan, Quanyi Tu, Wenjuan Zhou, Mansun Chan, Xiwen Liu

Hong Kong University of Science and Technology (HKUST)

This work demonstrates a CMOS-compatible three-dimensional vertical ferroelectric tunnel junction (3D V-FTJ) array that combines 3D integration with record-thin 1.x-nm HZO scaling. An all-ALD process and trench-sidewall plasma-oxygen treatment enable conformal ultrathin barrier formation while suppressing etch-induced defects and oxygen-vacancy-related leakage. A kb-scale 3 × 16 × 32 array is fabricated within a ≤400 ℃ thermal budget, achieving record-level reliability with write endurance beyond 10^{12} cycles and a breakdown field above 12.5 MV/cm. The devices further exhibit ~20-ns switching, ON-state current density exceeding 3 A/cm^2 at ~0.25 V, stable retention, and uniform operation across three vertical layers. Array-level conductance mapping and multilevel storage additionally verify scalability and reliable analog functionality. By jointly advancing lateral and vertical scaling without sacrificing reliability, the proposed platform bridges ultrathin FTJ operation and CMOS-compatible 3D integration, offering a practical pathway toward highly dense, low-voltage, and reliable BEOL nonvolatile memory.

10:20 AM

18-4 Insights into the Reliability of HZO Ferroelectric Capacitors Through Comprehensive DC/AC TDDB and Leakage Studies

Philémon Kouaté Wafo, Simon Martin, Olivier Billoint, Sam Fiette, Abygaël Viey, Carine Jahan, Fabien Grimaud, Julien Patouillard, Laurent Grenouillet
CEA-Leti

We report a combined DC/AC Time-Dependent Dielectric Breakdown (TDDB) and leakage study on $Hf_{0.5}Zr_{0.5}O_2$ (HZO) ferroelectric capacitors (FeCaps). DC TDDB on 5, 6 and 7nm films shows unimodal Weibull statistics enabling 10-year maximum-field extrapolation and confirming that reducing HZO film thickness improves the breakdown field (F_{BD}) (4.48MV/cm i.e., 2.24V for 5nm HZO 100µm diameter FeCaps). In contrast, AC TDDB reveals bimodal Weibull distributions and strong frequency dependence, evidencing an effective bilayer oxide stack. Temperature-dependent leakage measurements and modeling support the bilayer assumption, where HZO multi-phonon Trap Assisted Tunneling (MTAT) occurs under a reduced effective electric field imposed by an interfacial TiO_xN_y layer, likely associated with a Ti-scavenging-driven oxygen-vacancy (Vo) profile in HZO. These results reveal an effective bilayer nature of HZO FeCaps, providing broader insight into HZO-based ferroelectric devices. Finally, bipolar AC TDDB reveals a transition from HZO- to TiO_xN_y-limited degradation under Ferroelectric Random-Access Memory (FeRAM)-relevant bipolar stress conditions.

11:10 AM

18-5 New Insights of Generalized Fatigue Effects in HfO_2 and $Hf_{0.5}Zr_{0.5}O_2$ Ferroelectric Capacitors: Polarization Conservation, Frequency Independence, Universal Degradation and Predictive Method

Keying Sun[1], Hao Li[2], Ning Ji[3], Zongwei Shang[2], Jiajun Qiu[2], Hongcheng Jiang[4], Zuoyuan Dong[2], Dong Zhang[3], Chaoming Wu[3], Ming Li[2], Lang Zeng[1], Huixiong Deng[4], Xiao Gong[3], Runsheng Wang[2]
[1]Beihang University, [2]Peking University, [3]National University of Singapore, [4]Institute of Semiconductors, Chinese Academy of Sciences

In this work, the generalized fatigue effects across HfO_2 and $Hf_{0.5}Zr_{0.5}O_2$ ferroelectric capacitors are statistically investigated. The long-standing ambiguity about fatigue effect is elucidated by revealing the vital role of PUND amplitude. The domain dynamics extracted by the proposed novel decoupling method shows that the total polarization remains conserved during fatigue cycles, while switchable domains gradually transform into non-switchable domains. Such domain transformation exhibits the independence on fatigue cycling frequency within the range of 10KHz to 1MHz. It is further demonstrated that both HfO_2 and $Hf_{0.5}Zr_{0.5}O_2$ ferroelectric capacitors present universal fatigue degradation curve when plotted against to the switching ratio rather than the switching voltage. Based on these new findings, a fatigue degradation predictive method is established. With only once ***I-V*** curve measurement, the method can achieve accurate degradation prediction for any applied voltage. This work provides deep understandings for fatigue mechanism and practical method for rapid early-stage endurance estimation.

11:35 AM

18-6 New Mechanistic Insights and Experimental Evidences Enable Effective Mitigation of the Duty-cycle Induced Degradation in $Hf_xZr_{1-x}O_2$-Based Ferroelectric Capacitors

Yu Li[1], Jiaolong Liu[1], Guan Feng[1], Yuqi Gao[2], Hao Jiang[1], Xiaodong Wang[1], Linhui Hu[2], Feng Wang[2], Shucheng Zhang[1], Xianzhe Chen[1], Xumeng Zhang[1], Yushi Hu[3], Yong Kang[3], Qian Tao[3], Yingfen Wei[1], Qi Liu[1], Ming Liu[1]
[1]Fudan University, [2]Shanghai GTA Semiconductor Co., Ltd, [3]Wuxi Smart Memories Technologies, Co., Ltd.

This work provides insights into duty-cycle or access-rate-dependent degradation in $Hf_xZr_{1-x}O_2$ (HZO)-based ferroelectric capacitors: 1) we show that multi-voltage capacitance-frequency measurements can quantify evolution of charged oxygen-vacancy (V_o^{2+})-related border traps during duty-cycling; 2) despite nearly thickness-independent switchable-polarization degradations across 4-10 nm HZO films, coercive-field (E_c)-shift dominates in thicker films, whereas domain pinning dominates in thinner films; 3) experiments and simulations provide evidence for two-step defect process where V_o^{2+}s are generated by hot-atom damage and subsequently redistribute under built-in field ($E_{built\text{-}in}$) during delay time (T_d), thereby accelerating degradation; 4) guided by mechanistic insights, we develop a comprehensive physical model that attributes the distinct degradation behaviors across HZO thicknesses to thickness-dependent V_o^{2+} dynamics and spatial distributions; optimizations targeting both V_o^{2+} generation and migration are proposed and validated: well-tuned HZO strain simultaneously stabilizes ferroelectric orthorhombic

phase and increases average Hf/Zr-O bond energy, while an appropriate bias applied during T_d suppresses E_{dep}-driven V_o^{2+} migration.

12:00 PM

18-7 The 1st Cycle Effect: A Hidden Reliability Bottleneck in Scaled $Hf_{0.5}Zr_{0.5}O_2$ and Its Mitigation

Salma Soliman[1], Andrea Padovani[2], Chanyoung Kim[1], Pochun Hsieh[3], Antoine Arnould[4], Jihoon Choi[5], Nashrah Afroze[1], Yu-Hsin Kuo[1], Priyankka Ravikumar[1], Mengkun Tian[1], Malak Desouky[1], Sanghyun Kang[1], Michelle Ukiwe[1], Tyson Back[6], Chengyang Zhang[1], Suman Datta[1], Jun Hee Lee[5], Susan Trolier-McKinstry[3], Venkatraman Gopalan[3], Lauren Garten[1], Nicholas Barrett[4], Asif Khan[1]
[1]Georgia Institute of Technology, [2]University of Modena and Reggio Emilia, [3]Pennsylvania State University, [4]Université Paris-Saclay, [5]Ulsan National Institute of Science and Technology (UNIST), [6]Air Force Research Laboratory

$Hf_{0.5}Zr_{0.5}O_2$ (HZO) is the leading scalable ferroelectric for non-volatile memory, but its reliability at the 5 nm node under the elevated temperatures of 3D-stacked/BEOL integration remains poorly understood. While $VO\cdot\cdot$ (oxygen vacancy) accumulation during cycling is known to drive degradation via conductive paths, the outsized role of the first cycle has never been identified. We show that (1) the first cycle raises $VO\cdot\cdot$ concentration by 40-50% in 5 nm HZO, accounting for nearly half of 10^6-cycle leakage and setting the defect profile governing subsequent degradation; (2) dual ALD WO_x oxygen reservoir layers (ORLs) at both interfaces donate oxygen during this cycle, eliminating the surge and holding the profile near-pristine; (3) this first-cycle passivation yields 1000× endurance enhancement at 85–145°C, enabling 10^{10}-cycle operation at 145°C without breakdown. These results establish Cycle-1 $VO\cdot\cdot$ generation as a key, previously overlooked reliability bottleneck in scaled HZO, with implications for high-temperature 5 nm memory.

12:25 PM

18-8 A 10^{13}-Cycle Fatigue-Free HZO Ferroelectric Enabled by Oxygen-Gradient Engineering for Reliable Low-Voltage FeFETs

Chunlai Luo, Fan Wu, Shuxin Chen, Zhaoren Xie, Shuo Chen, Yu Liu, Xuelong Liu, Kaichen Zhu, Junhao Chu, Wenwu Li
Fudan University

HfO_2-based ferroelectrics suffer from pronounced polarization fatigue and reliability degradation, which severely limit the practical application of ferroelectric field-effect transistors (FeFETs). Here, we propose an oxygen-gradient engineering strategy by introducing ZrO_2 interlayers that gradually thin from the electrodes toward the film center into $Hf_{0.5}Zr_{0.5}O_2$ (HZO) films. This structure regulates the vertical oxygen-vacancy distribution, stabilizes the polar O-phase, and mitigates cycling-induced defect redistribution. The gradient-ZrO_2-engineered HZO (GZ-HZO) capacitors deliver a high remnant polarization of $2P_r > 40$ μC/cm^2 and an exceptional fatigue endurance **of 1.6**×10^{13} cycles. By integrating the GZ-HZO films into MoS_2-channel metal-ferroelectric-metal-insulator-semiconductor (MFMIS) FeFETs, the resulting devices exhibit a low operating voltage of 2.0 V, a high on/off ratio over 10^7, a large memory window of 1.8 V, ultrafast 20 ns program/erase speed, and robust 6×10^{12}-cycle switching endurance. This oxygen-gradient design provides an effective route to overcoming fatigue in HZO-based ferroelectrics, enabling high-speed, low-voltage, and highly reliable FeFETs.

Session 19: PMA | Advances in RF, mmWave, and Sub-THz III-V and Wide-Bandgap Transistor Technologies for 6G and Beyond

9:00 AM, Imperial B

Co-Chairs: Jose Jimenez, Qorvo, Inc. and Shamima Afroz, Northrop Grumman

This session showcases the latest breakthroughs in high-frequency transistor technologies that are enabling the next generation of 6G communications, radar, and sensing systems. The papers span a broad range of innovations in GaN HEMTs, including 28-V devices on 200 mm silicon for infrastructure applications, low-voltage Y-gate technologies for future 6G handsets, advanced AlN/GaN and AlScN/GaN heterostructures, and new approaches to achieving watt-class power densities and high efficiency at D-band frequencies. Together, they demonstrate significant progress in material engineering, device scaling, and RF performance. The session also highlights record-setting achievements in microwave and millimeter-wave power devices, including GaN HEMTs demonstrating up to

62 W/mm and nearly 60 W/mm output power density, as well as 120-W X-band power amplifiers enabled by novel fluorine ion termination technology. Concluding the session, an 830-GHz InAs quantum-well HEMT pushes the boundaries of transistor speed and provides insight into the ultimate room-temperature frequency limits of electronic devices. Collectively, these papers offer a compelling view of the technologies shaping RF and wireless systems for 6G and beyond.

9:05 AM

19-1 28V GaN HEMTs on 200mm Si Substrate for 6G Infrastructure Applications with 5.9W/mm Output Power and 46.5% PAE at 13GHz

Sachin Yadav, Muhammad Asad, Babak Kazemi Esfeh, Amratansh Gupta, Aarti Rathi, Arturo Sibaja-Hernandez, Rana ElKashlan, Sourish Banerjee, Erik Bury, Wei-Min Wu, Bjorn Vermeersch, Edoardo Brezza, Alexander Possberg, Ambra Neri, Joris Van Driessche, Barry O'Sullivan, Hao Yu, Uthayasankaran Peralagu, Alireza Alian, Bertrand Parvais, Nadine Collaert
imec

We report 120 nm gate length GaN HEMTs on 200 mm Si substrates operating at 28 V in the 6G FR3 band. Combining technology elements such as enhanced gate and source connected field plate designs, low resistivity ohmic contacts, robust ex-situ passivation, and 3-metal level Cu BEOL, the performance potential of GaN-on-Si for 6G infrastructure power amplifiers (PAs) is explored. Impact of the transistor design on the RF and reliability performance is studied and 28 V operation at 300 µm gate width is demonstrated with a continuous-wave (CW) Pout: 5.9 W/mm (32.45 dBm) with corresponding PAE: 46.5% and AM-PM: 2.8° at 13 GHz, highlighting the potential of GaN-on-Si technology for 6G wireless infrastructure applications.

9:30 AM

19-2 0.1-μm Y-gate Low-voltage GaN HEMT and MMIC PA with PC1.5-required Power Rating and Record-high PAE of 57% for U6G Handset Applications

Haochen Zhang[1], Yue Sun[2], Hongtu Qian[1], Yangyi Zhao[2], Qibing Fan[1], Zhaozi Chen[1], Rongyu Tang[1], Zili Wu[1], Xiao Han[1], Yongsheng Zhang[1], Xinchuan Zhang[1], Yi Pei[1], Jianan Liu[2], Kevin Chen[3], Haiding Sun[1], Naiqian Zhang[1]
[1]Dynax Semiconductor Inc., [2]Xiaomi Communications Company Ltd., [3]Hong Kong University of Science and Technology (HKUST)

The upper-6-GHz-band has emerged as an important spectrum extension for 5G-Advanced and potential bridge towards 6G, requiring PA to deliver sufficient P_{out}, high PAE, and wideband at elevated frequency. Herein, based on deep-sub-μm low-voltage (LV) GaN-HEMT technologies, we present a unified device-to-MMIC solution tailored for U6G-band, delivering record-breaking power-performance at both transistor & circuit levels. A 0.1-μm-air-spacer-Y-gate and high-quality regrowth-ohmic-contacts are adopted to minimize device parasitics and enable enhanced power-performance at low-drain-bias and high frequency. A maximum P_{out} of 2.8-W/mm and 82.2%-PAE are achieved at 10-V 7-GHz. A two-stage LV-GaN MMIC-PA is further fabricated, delivering a typical linear-gain of 27-dB, 39-dBm saturated-P_{out} and 57%-PAE across n104 band. The linearity of the MMIC-PA is evaluated using 100-MHz DFT-s-OFDM-QPSK, yielding 33-dBm-P_{out} with well-controlled linearity. This work marks the compound-semiconductor PA to achieve record efficiency while satisfying Power-Class-1.5 requirements for U6G-handset applications, setting a leading performance benchmark for GaN-handset-PA solutions in future mobile communication.

9:55 AM

19-3 AlN/GaN/AlN XHEMTs on 100-mm Bulk AlN Substrates by MOCVD

Hanqiao Zhang[1], Yuxuan Deng[1], Jeongwoo Kim[1], Aias Asteris[1], Kazuki Nomoto[1], Akshey Dhar[1], Clara Fleisig[1], Shake Karapetyan[1], Yu-Hsin Chen[1], Jimy Encomendero[1], Kasey Hogan[2], James Grandusky[2], David A Muller[1], Debdeep Jena[1], Hari Nair[1], Huili Grace Xing[1]
[1]Cornell University, [2]Crystal IS Inc

Driven by the ultrawide bandgap of 6.2 eV, the availability of large-size high-quality single-crystal AlN substrates has enabled the development of numerous ultraviolet (UV) photonic devices. In this work, we demonstrate AlN/GaN/AlN pseudomorphic HEMTs (XHEMTs) grown by MOCVD on 100-mm bulk AlN substrates. While remaining compatible with high-volume manufacturing, we construct heterostructures with managedstrain taking advantage of the high thermal conductivity of bulk AlN. We show uniform 2DEG densities

and low channel resistance in ~15 nm and 10 nm channels. The fabricated XHEMTs achieve an output power (P_{out}) of 9.62 W/mm at X-band. These findings provide a path toward unifying the benefits of both epitaxy design and manufacturing scalability, positioning XHEMT on bulk AlN as a promising candidate for next-generation RF electronics.

10:20 AM

19-4 D-Band MOCVD-Grown AlScN/GaN RF HEMTs with >1.1 W/mm and >22% PAE at 120 GHz

Utku Agtas[1], Hakan Cankat Gur[1], Onder Soydal[1], Francesco Marcantoni[1], Yuan Zong[1], Teresa Duarte[2], Byeongchan So[2], Lutz Kirste[2], Patrik Stranák[2], Mario Prescher[2], Stefano Leone[2], Rüdiger Quay[2], Elison Matioli[1]
[1]École Polytechnique Fédérale de Lausanne (EPFL), [2]Fraunhofer Institute for Applied Solid State Physics

We report, for the first time, large-signal performance of MOCVD-grown AlScN/GaN RF HEMTs at D-band frequencies. The passivated devices deliver a high current density of 3.4 A/mm and a peak transconductance of 750 mS/mm. The unpassivated device achieves f_T/f_{MAX} of 195/300 GHz, the highest f_T reported to date among AlScN/GaN HEMTs, and 175/290 GHz after passivation. Load-pull measurements at 120 GHz yield a PAE of 22% with an output power density of 1.1 W/mm, the highest PAE reported for any Ga-polar GaN technology at D-band, and 1.46 W/mm at the poweroptimized bias. At 160 GHz, the devices achieve 18.9% PAE and up to 1.02 W/mm. Notably, the devices had not reached compression at the maximum available input power, indicating that the intrinsic performance exceeds these values. These results establish MOCVD-grown AlScN/GaN HEMTs as a scalable platform for next-generation high-power mm-wave electronics.

11:10 AM

19-6 First Demonstration of 62 W/mm X-band GaN HEMT and 120W 7.7-8 GHz PAs utilizing Fluorine Ion Termination technology

Bin Hou[1], Qingyuan Chang[1,2], Jiafeng Wu[1,3], Jingshu Guo[2], Tianqi Pan[1], Youjun Zhu[1], Zelin Lai[1], Shuyang Liu[1], Qingyuan Hou[1], Weiyu Ren[1], Jingshen Chen[1], Jiale Du[1], Qing Zhu[1], Meng Zhang[1], Mei Wu[1], Hao Lu[1], Mao Jia[1], Yang Lu[1], Minhan Mi[1], Weiwei Wang[1], Yi Wang[1,3], Wenrao Fang[2,4], Hao Shao[2], Ling Yang[1], Xiaohua Ma[1], Yue Hao[1]
[1]Xidian University, [2]Qianyuan Laboratory, [3]The 13th Research Institute of China Electronics Technology Group Corporation, [4]Northwest Institute of Nuclear Technology

Ultra-high output power density of 62 W/mm and peak PAE of 60.5% at 8 GHz and 135 V for GaN HEMTs were firstly reported using fluorine ion termination (FIT) technology based on AlN buffer structure. Molecular dynamics (MD) calculations reveal that the implanted fluorine ions are strongly confined by Si-F bonding within the upper a-SiN_x region, enabling stable negative fixed-charge formation and flattened electric-field profile. Off-state stress measurement at V_{gs} = -6 V and V_{ds} = 135 V shows that the device exhibits only a slight threshold voltage shift (ΔV_{th}) of -0.15 V and limited drain current variation ($\Delta I_{d,max}$) of -30 mA/mm. Finally, a PA with a single 3-mm gate-width FIT-GaN HEMT demonstrates broadband operation from 7.7 to 8 GHz, delivering a maximum P_{out} of 120.3 W (40.1 W/mm) at 100 V. These results highlight the potential of FIT technology for high-voltage and high-power GaN RF applications.

11:35 AM

19-7 GaN RF HEMTs with Record P_{out} of 59.8 W/mm at 10 GHz

Kun Zhang[1], Hong Zhou[1], Yixing Wang[1], Min Zhou[1], Xinchuan Zhang[2], Wenjun Liu[1], Qifeng Lyu[2], Yachao Zhang[1], Chenlu Wang[1], Chunfu Zhang[1], Zhihong Liu[1], Yi Pei[2], Han Wang[3], Yuhao Zhang[3], Yue Hao[1], Jincheng Zhang[1]
[1]Xidian University, [2]Dynax Semiconductor Inc., [3]The University of Hong Kong

GaN-on-SiC RF HEMTs featuring an AlN buffer have broken a decade-old record by enabling maximum output power density (P_{out}) exceeding 40 W/mm. This work fills this gap by identifying that current collapse at high drain bias (V_{DS}) dominates this P_{out} upscaling limitation and introduces distinct device pathways to overcome it. We demonstrate that a thicker AlN buffer, combined with in-situ/ALD/PECVD SiN stacked passivation layers, effectively mitigates buffer and surface traps. By increasing the AlN buffer thickness from 200 nm to 350 nm, the current collapse ratio (CCR) is reduced from >15% at V_{DSQ} = 40 V to 9.6% at V_{DSQ} = 80 V. Additionally, MOCVD-regrown Ohmic contacts are implemented to achieve a contact resistance (R_C) of 0.12 Ω·mm. By integrating these epitaxial and process optimizations, the fabricated GaN/AlN-on-SiC RF HEMTs achieve a record P_{out} of 59.8 W/mm at 10 GHz with a peak PAE of 50.7%.

12:00 PM

19-8 f_T = 830-GHz InAs Quantum-Well HEMTs Enabled by High-Mobility Carrier Transport and Physics-Based Delay Analysis Toward the Ultimate Room-Temperature Frequency Limit

Seung-Woo Son[1], In-Geun Lee[1], Sang-Pyeong Son[1], Ji-In Byeon[1], Se-Hun Kim[1], Min-Gyu Song[1], Sang-Kuk Kim[2], Jacob Yun[2], Ted Kim[2], Taro Sasaki[3], Takuya Hoshi[3], Hiroki Sugiyama[3], Fumito Nakajima[3], Tae-Woo Kim[4], Kyounghoon Yang[5], Dae-Hyun Kim[1]

[1]Kyungpook National University, [2]QSI, [3]NTT, Inc., [4]Texas Tech University, [5]Korea Advanced Institute of Science and Technology (KAIST)

We present a 35-nm-L_g InAs HEMT exhibiting a record-high cutoff frequency (f_T) of 830 GHz , representing the highest f_T reported among all transistor technologies. The fabricated device simultaneously possesses excellent DC performance with R_{ON} = 228 Ω·μm, g_{m_max} = 2.76 mS/μm, S = 65 mV/dec, and $I_{ON}/I_{OFF} > 10^5$. These outstanding DC and RF performances originate from three key technology and physics enablers: (1) an InAs sub-channel QW with an effective electron mobility (μ_{eff}) of 17,000 $cm^2/V{\cdot}s$, simultaneously reducing the R_s and enhancing carrier transport; (2) a physics-based carrier transport framework incorporating *quasi-ballistic* transport with accurate R_s extraction, enabling quantitative analysis of carrier transport in aggressively scaled InAs HEMTs; and (3) a physics-based delay analysis revealing the physical origin of the record-high f_T. Consequently, the ultimate room-temperature frequency limit of InAs HEMTs is projected to be approximately 930 GHz, establishing a fundamental design guideline for future HEMT technologies.

Session 20: MT | Nonvolatile Memories

2:00 PM, Grand Ballroom A

Co-Chairs: Mattia Boniardi , STMicroelectronics and Kai-Shin Li, National Tsing Hua Univ.

This session is devoted to seven groundbreaking advances in Nonvolatile Memory technologies, spanning embedded NVM, RRAM, MRAM, and selector-only memory (SOM). The first contribution will be focused on a 320-kb anti-fuse eNVM in 3nm FinFET that employs an STI-less layout to naturally thwart hardware attacks with 100% yield. Next, Infineon will go through an automotive-grade 28nm embedded RRAM as conventional flash replacement with 1000-hour data retention at 175oC, followed by a 3D vertical Gain-Cell RRAM that exploits FinFET transconductance amplification to enable record-low 1 V 4-bit (QLC) operation with 24 ns read latency. Then TSMC will show a scaled 0.0187um^2 universal MRAM delivering sub-10ns writes to bridge the gap between storage and last-level cache in 5nm nodes, followed by the first 16-Kb field-free Type-Cone SOT-MRAM achieving deterministic 1 ns switching at an 80 nm scale. For extreme environments, Avalanche Technology will showcase the industry's highest-density monolithic 2Gb STT-MRAM, providing complete heavy-ion radiation immunity for in-orbit AI. Finally, a binary SOM will be shown, utilizing a sub-2 nm 2DEG electrode to tightly confine Joule heating, delivering a >1.4V window and >5 x10^10 cycling endurance.

2:05 PM

20-1 A 3nm FinFET Secure PUF-Obfuscated 320-kb Anti-Fuse eNVM:STI-less Bitcell Design with In-Cell Redundancy

HsieHsien Ho, Tsao-Hsin Yang, Dennis Ho, Chi-Chang Lin, Claire Su, Wilson Liu, David Hoang, Zhe Yi Lin, Po-Hao Huang, Benedict Hu, Allen Lin, Chun-Lin Chien, Chih-Min Wang, Chun-Fu Lin

eMemory Technology Inc.

This paper presents a fully logic compatible PUF-obfuscated 320-kb anti-fuse (AF) eNVM in 3nm FinFET. The proposed STI-less bitcell utilizes a continuous active region, eliminating LOD penalties without Poly-On-Diffusion-Edge (PODE) overhead while providing intrinsic mitigation against Passive Voltage Contrast (PVC) attacks. To enable 4V programming on 0.75V core AF devices, an auxiliary following transistor forms a voltage divider, effectively suppressing GIDL leakage and preventing punch-through. In-cell redundancy is implemented to mitigate scaling-induced defects for stable manufacturing. Silicon measurements demonstrate 100% OTP yield with zero PUF bit error across corners and qualification lots, enabling a robust, secure macro for advanced SoCs.

2:30 PM

20-2 Embedded RRAM for Automotive Microcontrollers

Robert Strenz, Thomas Kern, Robert Wiesner, Karl Hofmann, Edvin Paparisto, Sebastian Kiesel, Thomas Roehr, Thomas Liebermann, Christoph Roll, Stephan Kassenetter, Roland Renardy, Nora Lawal, Thomas Rabenalt, Martin Stiftinger, Stefano Aresu, Lucas Reganaz, Rainer Duschl, Akshaya Narayanan, Rudolf Ullmann, Nellina Mautone, Felix Tengler, Jonas Kiemle, Florian Konder, Philipp Zimmermann, Thomas Borst, Gunnar Diestel, John Power, Corneliu Colesniuc, Sanjeeb Bharali, Po Yuan Shih, Ingolf Rau, Jacek Szatkowski, Gianfranco Capetti, Martin Keck, Matteo Coppetta, Alessandro Grossi, Pierre Scaramuzza, Federico Finotto, Daniele Giordano
Infineon Technologies

Starting with 28nm CMOS technology node, embedded flash technologies are getting more and more complex and costly to be integrated into CMOS baseline technology. This difficulty was driving significant effort in recent years of developing alternative embedded NVM concepts like MRAM, PCM and RRAM. This paper gives an overview demonstrating that besides 22/28nm MRAM and PCM also transition metal oxide based RRAM has proven its capability to not only cover consumer and industrial grade applications, but also to reach most stringent automotive requirements. RRAM technology characteristics are presented enabling a wide variety of Infineon automotive microcontrollers (MCUs) in 28nm generation. The data presented are based on Infineon proprietary RRAM NVM design IP used in automotive MCU products. Mission profiles and requirements are fully covering those of embedded flash technologies. RRAM has been chosen due to its low process complexity, corresponding process robustness and cost competitiveness, immunity to magnetic disturb, as well as its compatibility to CMOS type of fab processes. This paper covers characterization data of both RRAM program (PMEM) and data (DMEM) type of memories, as well as some further technology aspects.

2:55 PM

20-3 Gain Cell RRAM: An Ultimate Solution of 1V 3D Vertical Resistive-Gate RAM with QLC Operation for FinFET Generation and Beyond

Y. T. Wang[1], S. P. Ho[1], C. F. Zhan[1], C. H. Sung[1], Y. T. Yang[1], P. H. Shih[1], J. C. Guo[1], Z. H. Lin[2], Osbert Cheng[2], Steve Chung[1]
[1]National Yang Ming Chiao Tung University, [2]United Microelectronics Corporation (UMC)

For the first time, a new concept of **Gain cell RRAM** was introduced. It was built with a 3D AND architecture and 2TnR Resistive-Gate RAM (RG-RAM), enabling a record low voltage (1V) and stackable vertical MIM layers on a 16nm FinFET platform. The unit cell is a 2T gain cell RRAM where 1T serves as a control transistor for programming while the other 1T serves as READ output. MIM RRAM is forming free with low SET/RESET operating voltages (-1V and +1V, respectively) and excellent HRS/LRS ratio (>10^5). Also, excellent endurance over 10^8 cycles and excellent retention performance under 125°C baking with 10-year lifetime can be achieved. Finally, a V-RRAM memory core is integrated with peripheral circuit, forming a fully functional memory of 1kb RG-RAM Macro. Ultra-fast speed of READ operation at 1V within 24ns can be achieved. This architecture has the potential to be extendable to Angstrom generation, e.g., nanosheet.

3:45 PM

20-4 Universal MRAM: Enabling Sub- 10 ns RAM-like Applications and Bridge the FLASH-to-RAM Memory Performance Gap in Advanced CMOS

Allen Wang, Ding-Shuo Wang, Ming-Han Tsai, Han-Ting Lin, Chih-Chen Peng, Yin-Yu Hsueh, Nien-Yu Dai, Jian-Cheng Huang, Yu-Sheng Chen, Butsurin Jinnai, Chien-Min Yang, Roger Wang, Kuo-Chyuan Tseng, Yuan-Jen Lee, Ku-Feng Lin, Yi-Chun Ethan Shih, Kuo-Feng Huang, Tien-Wei Chiang, Eric Chien, Yi-Ching Ong, Chia-Yu Wang, Chia-Fu Lee, Jenn Jou Wu, Yih Wang, Yu-Der Chih, Harry Chuang, Chue-San Yoo
Taiwan Semiconductor Manufacturing Company (TSMC)

This work presents a fast-write embedded MRAM with a 0.0187 μm^2 bitcell in a 16nm CMOS node, delivering a high retention “FLASH-plus” and a high-endurance “RAM-like” solution. We also report the first sub-10ns write in a 128 Mb MRAM testchip with write current optimized for the 5nm nodes. The technology aims at increasing on-chip last level cache (LLC) density, thereby reducing cache misses, minimizing off-chip memory access (both bandwidth and power), and enabling energy-efficient standby.

4:10 PM

20-5 16-Kb Field-Free Type-Cone SOT-MRAM Enabled by Domain-Wall Chirality Engineering

Peiyue Yu[1], Fengyi Yu[1], Meiyin Yang[1], Dashan Shang[1], Jingsong Huang[1], Yanru Li[1], Shuo Xu[1], Shuaiyu Gong[1], Lisha Liu[1], Jianfeng Gao[1], Weibing Liu[1], Shasha Wang[2], Lei Zhao[2], Enlong Liu[2], Wenlong Yang[2], Yang Gao[2], Shikun He[2], Jun Luo[1]

[1]Institute of Microelectronics of Chinese Academy of Sciences, [2]Zhejiang Hikstor Technology Co., Ltd.

We demonstrate a CMOS-compatible, field-free Type-Cone SOT-MRAM. The Ta/W multilayer channel is consistent with a stable easy-cone anisotropy in CoFeB, enabling deterministic switching through domain wall chirality control. Fabricated on a 300-mm platform, SOT-MTJ devices exhibit 113% tunnel magnetoresistance, 1 ns switching, high thermal stability (Δ = 73), and 10^{12} endurance cycles with BEOL compatibility. Integration into a 16-Kb array demonstrates field-free write feasibility alongside tight resistance distributions and robust read margins. This Type-Cone SOT-MRAM successfully resolves long-standing density versus field-free trade-offs, providing an aggressively scaled, industry-compatible pathway for next-generation non-volatile memory.

4:35 PM

20-6 Monolithic Multi-Interface 2Gb MRAM for In-Orbit AI Application

Zihui Wang, Guanzhong Wu, Zhiqiang Wei, Longqian Hu, Mengchun Wu, Uday Chandrasekhar, Thinh Tran, Dean Nobunaga, Mourad Baraji, Jun Li, Fred Jaffin, Ebi Abedifard, Yiming Huai

Avalanche Technology

This paper presents a monolithic multi-interface 2Gb Spin-Transfer Torque MRAM (STT-MRAM) for in-orbit artificial intelligence (AI) applications. This device is the highest-density monolithic MRAM chip currently on the market and enables stacked products up to 16Gb. The multi-interface design ensures broad compatibility, facilitating the deployment of MRAM across diverse environments to supersede traditional alternatives. Design and technology co-optimizations (DTCO) improve switching uniformity, retention, write-error behavior, TMR/Resistance distribution, and read margin. Heavy-ion testing showed no observed SEL, SEFI, or SEU events across the tested LET range at both room and elevated temperature.

5:00 PM

20-7 Switching Volume Confinement in Selector-Only Memory (SOM) via a Sub-2 nm Two-Dimensional Electron Gas Electrode for Large Memory Window (> 1.4 V) and Excellent Write Endurance (> 5 x 10^{10})

Jangseop Lee[1], Dongmin Kim[1], Yoori Seo[1], Min Kim[1], Yu Bin Park[2], Tae Hoon Lee[2], Wu Zhe[3], Kwangmin Park[3], Seunghun Lee[3], Hyunsang Hwang[1]

[1]Pohang University of Science and Technology (POSTECH), [2]Kyungpook National University, [3]Samsung Electronics

In this study, we demonstrate an extremely scaled selector-only memory (SOM) based on a two-dimensional electron gas (2DEG) as a resistive electrode. Formed at the TiO_2/Al_2O_3 interface, the atomically thin 2DEG confines the conduction path to below 2 nm, while the oxide layers localize the Joule heating within this region. COMSOL and DFT analyses reveal that this localized heating promotes the carrier-induced bonding state modulation of the chalcogenide material, achieving a large memory window (MW > 1.4 V), suppressed Vth variability, and an ultra-low write current (Iwrite ~ 10 μA). This highly localized switching region, combined with suppressed overshoot current, enables excellent write endurance (> 5x10^{10}). The vertically stacked structure confirms the strong potential of the proposed SOM for high-density 3D integration.

Session 21: ALT | BEOL, Thermal and Packaging

2:00 PM, Grand Ballroom B

Co-Chairs: Tetsu Ohtou, TEL and Shay Reboh, IBM

The session covers eight papers addressing key advances in BEOL, BSPDN, and thermal engineering for advanced logic technologies. imec presents the feasibility of 18 nm-pitch Ru interconnects fabricated using High-NA single patterning, while TSMC demonstrates highly oriented PtCoO2 as a novel back-end-of-line conductor. CEA-Leti further advances low-temperature integration with a 400°C analog devices enabled by nanosecond laser annealing for 3D sequential CMOS. On the BSPDN field, Samsung develops a novel BSPDN design scheme while Peking University shows detailed thermal characterization of advanced logic technologies incorporating BSPDN. Complementing these works, SUSS MicroTec presents a paper on low-distortion bonding for advanced backside

interconnects. On the thermal engineering area, KAIST showcases an InGaAs HEMT-embedded active SiC interposer platform, and National Taiwan University presents a new superlattice-engineered approach with pyroelectric energy harvesting and electrocaloric cooling for thermal management.

2:05 PM

21-1 DTCO-enabled scaling via High-NA single patterning with electrical validation in 18nm pitch Ru DME

Subham Dastidar[1], Yun Zhou[1], Shubhankar Das[1], Federica Lucciano[1], Kenichi Miyaguchi[1], Balakumar Baskaran[1], Yasser Sherazi[1], Hsinlan Chang[1], Subash Sharma[1], Halil Kükner[1], Lynn Verschueren[1], Bart De Wachter[1], Stefan Decoster[1], Yannick Hermans[1], Ryoung-Han Kim[1], Geert Hellings[1], Victor M. Blanco Carballo[1], Yugyeong Lee[2], Ajeong Kim[2], Howon Jung[2], Taehoi Park[2], Sangho Yun[2], Changmin Park[2], Seunghun Lee[2], Zhengtao Chen[3], Elisabeth Camerotto[3], Shruti Jambaldinni[3], Anuja De Silva[3], Tsann-Bim Chiou[4], Chun-Kuang Chen[4], Martin O'Toole[4], Etienne De Poortere[4]
[1]imec, [2]Samsung Electronics, [3]Lam Research, [4]ASML

We demonstrate a single-exposure 0.55 NA EUV dark-field flow using MOR and dry development, eliminating pitch walk and simplifying integration process. Electrical test structures show defect-free yield at 18 nm pitch over 1.2 mm line lengths. High-NA-enabled DTCO with curvilinear layouts improves BEOL routing and pattern fidelity, while aggressive scaling with small minimum island and tight tip-to-tip (T2T) spacing supports continued standard cell scaling, establishing a viable path for sub-20 nm interconnects. Early DTCO studies show curvilinear M0 layouts enhances routing efficiency, increasing available M1 resources for intercell routing by 11% while reducing 9% M3-M12 and 13% V2-V12 vias.

2:30 PM

21-2 Overcoming the Scaling Limit: Highly Oriented $PtCoO_2$ as a Novel Back-End-of-Line Conductor

Ming Han Lee[1], Yu Chen Chan[1], Shu Wei Lee[1], Shin Yi Yang[1], Donglai Zhong[1], Chuang Han Hsu[1], Meng Pei Lu[1], Wei Sheng Yun[1], Shu Wei Chang[1], Mark van Dal[1], Greg Pitner[1], Iuliana Radu[1], En Cheng Chang[2], Chih I Wu[2], Hsin Ping Chen[1], Ru Gan Liu[1], Min Cao[1]
[1]Taiwan Semiconductor Manufacturing Company (TSMC), [2]National Taiwan University

This work demonstrates the potential of $PtCoO_2$ as a transformative Back-End-of-Line (BEOL) interconnect solution to overcome the scaling challenges of conventional metallization. We report a width-scaling benchmark for narrow $PtCoO_2$ lines down to 10 nm, demonstrating that they outperform traditional Cu/barrier systems. This impressive scaling behavior stems from a combination of low in-plane resistivity and suppressed surface scattering. Furthermore, leveraging the intrinsic transport properties of $PtCoO_2$ through optimized contact engineering yields a contact resistivity comparable to typical metal-metal interfaces. Additionally, an exceptional breakdown current density is demonstrated, verifying its outstanding electrical robustness. These results firmly establish $PtCoO_2$ as a viable candidate to power next-generation interconnects.

2:55 PM

21-3 Bringing 400°C Analog Devices to the Next Level: Mastering nano second Laser Annealing for 3D Sequential CMOS SOI

Daphnee Bosch[1], William Vandendaele[1], Anne-Sophie Royet[1], Patrick Reisinger[2], Sebastien Kerdiles[1], Mathieu Opprecht[1], Antoine Albouy[1], Francois-Xavier darras[1], veronique benevent[1], Lucie Laraignou[1], Flavien Berthaud[1], obaid Adami[1], Elhadji-Alhousseyni DIALLO[1], Leo Matokwong[1], Mickael Ribotta[1], Adeline grenier[1], Clemence Hellion[1], remi gassilloud[1], Xavier Garros[1], Christoforos Theodorou[3], Perrine Batude[1]
[1]CEA-Leti, [2]TU Wien, [3]Université Grenoble Alpes

This work presents key breakthroughs in nanosecond laser annealing (NLA) for 400°C CMOS SOI devices, tailored for 3D sequential integration (3DSI). Going well beyond previous studies, our results stand out due to the following aspects: i) Record nMOS low trap densities (N_t= 5×10^{15} $eV^{-1}\cdot cm^{-3}$) for narrow transistors, even surpassing state-of-the-art high-temperature devices; ii) First demonstration of 20 years hot-carrier injection lifetime for both n & pMOS. The NLA tool is pivotal in gate stack healing and junction optimization, operating in both melt and solid-phase epitaxial regrowth (SPER) regimes. This paper provides strong insights into the origin of these improvement, including the temperature dependency of gate oxide healing and the fine decomposition of activation mechanisms in different junction regions, demonstrating the full functionality of the optimization toolbox (access resistance and

leakage). These advancements highlight the readiness of 400°C Silicon devices for large-scale industrial adoption in 3D sequential integrated analog and mixed-signal applications.

3:20 PM

21-4 Enabling Manufacturable BSPDN through a Novel Scheme Design: Element Technology Development and Validation

Juggil Yang, Jinkyu Kim, Junsun Hwang, ByoungGi Kim, Heebum Hong, Hyoungjun Kim, Hidenobu Fukutome, Jeewoong Kim, Hyeonseok Jeong, Myungjin Chung, Soohang Chae, Goun Kang, Younghan Kim, Euncheol Kim, Yoolim Ahn, Taeil Eom, Hyeongwoo Jo, Beomjin Kim, Seoyeon Park, Yoonbeom Park, Jeongwoong Shin, Ra You, Yongwoo Lee, Myeonggyoon Chae, Shinkyu Choi, Daeyoung Chu, Sunghwan Kim, Man Chang, Yunsuk Nam, Young Gun Ko, Eunguk Chung, Hyung Suk Lee, Yonghee Park, Kyoung-Woo Lee, ChangMin Park, Kuntack Lee, Hoyoung Kim, Hyoseok Choi, Jaehoon Jeong, Sunjung Kim, Keun Hwi Cho, Ilryong Kim, Jaihyuk Song
Samsung Electronics

By developing and validating key element technologies, this work provides a novel direct backside contact (DBC) scheme on a 2 nm-node GAA technology platform that successfully enables manufacturable BSPDN architecture. We introduce novel element technologies, including optimized bonding and Si thinning processes, which effectively suppress bonding-related defects and reduce backside contact overlay degradation by 57%. The logic density is reduced to 0.92× that of a conventional FSPDN while achieving either an 8 % speed increase or a 15 % power reduction. A key contribution of this work is the development and validation of a thin-bulk compatible lateral Gate-Bounded Diode, which successfully enables essential IP functions such as Band-Gap Reference (BGR) circuits and Electrostatic Discharge (ESD) protection devices within the BSPDN environment. This work demonstrates the novel DBC-based BSPDN platform for next-generation AI and HPC products.

4:10 PM

21-5 First Three-Dimensional Thermal Characterizations and Optimization-Oriented Reconstruction for Advanced Logic Technology with BSPDN Beyond 2 nm Node

Junliang Fan, Zirou Chen, Haoyue Huang, Zifeng Huang, Ming Li, Runsheng Wang, Zhe Cheng
Peking University

In this work, we for the first time presented an experiment-based framework spanning three-dimensional thermal characterizations, thermal reconstruction and AlN-centered thermal optimization of advanced backside power delivery networks (BSPDN) chips. By measuring thermal conductivities (TCs) of the chip's critical structures via time-domain thermoreflectance (TDTR), we characterized and discussed key thermal optimization features, including dense, thick and high-TC backside Cu wires (up to 400 W/m-K) and vias, thermally favorable backside dielectrics and carrier Si bonding layer. As a natural extension of the existing thermal optimization, we investigated low-temperature AlN films as a dielectric replacement and experimentally observed phonon ballistic transport in films below 100 nm, informing future bonding-layer and backside-dielectric design. On this basis, we reconstructed a thermal model and assessed system-level thermal strategies. Taken together, the results established a practical metrology and thermal database for BSPDN characterizations, laying the groundwork for thermal-aware design of future BSPDN chips.

4:35 PM

21-6 Low Post Assembly Distortion Bonding Technology Enabling Advanced Backside Nano Interconnects

Philippe Muller[1], Arne Bahr[1], Utkarsh Jain[2], Oguzhan Orkut Okudur[2], Philipp Schmidt[1], Koen D'have[2], Steven Brems[2], Koen Kennes[2], Damien Leech[2], Thomas Schmidt[1], Zsolt Tokei[2]
[1]SUSS MicroTec Solutions Gmbh, [2]imec

This paper presents pneumatic-deflected wafer to wafer bonding for a novel low distortion fusion and hybrid bonding applications required in CFET and backside Power Delivery Network (PDN). Impact of wafer thickness variation and activation plasma on post bond distortions are observed. Initial control loop and bond recipe optimization examples demonstrated potential for <5nm postbond lithography distortions with low-order corrections-perexposure (CPE), with <100nm max bonding overlay vectors.

5:00 PM

21-7 InGaAs HEMT-embedded Active SiC Interposer platform for Low-Loss RF Routing, Switching, and Thermal Managements

Nahyun Rheem[1], Yoon-Je Suh[1], Chaeyeon Han[1], Seokang Choi[1], Jaeyong Jeong[2], Sanghyeon Kim[1]
[1]Korea Advanced Institute of Science and Technology (KAIST), [2]Ulsan National Institute of Science and Technology (UNIST)

This work presents a semi-insulating SiC (SI-SiC) active RF interposer integrating InGaAs HEMT active cells with passive RF routing. SI-SiC provides glass-like low-loss transmission and high thermal conductivity, enabling compact passive/active heterogeneous integration. CPW lines on SI-SiC show a transmission loss of ~0.5 dB/mm up to 50 GHz, ~9× lower than Si/2μm SiO_2. The integrated InGaAs HEMT exhibits g_m = 1.47 S/mm, R_{on} = 0.4 Ω·mm, and f_T/f_{MAX} = 444.5/753.6 GHz. TRM measurements reveal a 10× lower device-level thermal resistance than Si/2μm SiO_2, and calibrated chip-level simulations confirm lateral heat spreading, reducing peak temperature from 430 to 415 K and thermal resistance by 11.8 %. The HEMT maintains stable operation from 303 to 353 K and achieves $R_{on}{\cdot}C_{off}$ = 295 fs. These results establish SI-SiC as a low-loss, thermally manageable active RF interposer for high-bandwidth heterogeneous integration in future sub-THz systems and advanced heterogeneous packaging applications.

5:25 PM

21-8 Promising Superlattice-Engineered $Hf_{1-x}Zr_xO_2$ with Record-High Pyroelectric Energy Harvesting and Electrocaloric Cooling for On-Chip Thermal Management

Kun-Yuan You, Cheng-Hong Liu, Mochamad Januar, Chia-Chun Hung, An-Chi Luo, Yi-Shin Hsu, Min-Hung Lee
National Taiwan University

$Hf_{1-x}Zr_xO_2$ (HZO) thin films offer a CMOS-compatible route for low-voltage on-chip solid-state cooling and pyroelectric energy harvesting (PEH). This study introduces the first ferroelectric–antiferroelectric (FE-AFE) transitional HfO_2/ZrO_2 superlattice (SL) capacitor as a monolithic hafnia-based platform that integrates electrocaloric (EC) cooling and pyroelectric energy harvesting. Over a chip-relevant temperature span (ΔT_{HL}) of 100 K from 300 to 400 K, the SL FE-AFE capacitor achieved a temperature-normalized harvestable energy density (HED/ΔT_{HL}) of 0.1124 J/(cm^3·K), corresponding to an experimentally measured HED of 11.24 J/cm^3. Under unipolar operation, the same device delivered an EC responsivity of 0.02043 K·mm/kV and a refrigerant capacity (RC) of 2.96 J/cm^3. These results demonstrate that the SL FE-AFE capacitor is a promising candidate for concurrent on-chip cooling and energy harvesting in 3D ICs.

Session 22: SMB | MEMS Resonators and Transducers

2:00 PM, Continental 1-3

Co-Chairs: Humberto Campanella, Silicon Austria Labs GmbH and Jeronimo Segovia-Fernandez, Texas Instruments

This session presents 6 papers that describe the latest advances in micro electro-mechanical system (MEMS) resonators and transducers, and their technologies. The first paper by S. Madhvapathy et al. presents an optomechanical sensor based on Si micro-ring ultrasound resonators that reports 4x sensitivity improvement through Si backside etching. The second paper by Cai et al. presents a tunable acoustic resonator for filter applications enabled by a transferred piezo-ferroelectric perovskite oxide NaNbO3 layer and demonstrating a record coupling coefficient of 34-36% and a tuning range of 15%. The third paper by Fan and Li presents a CMOS-MEMS Lorentz-force magnetometer that employs current re-circulation and piezoresistive detection with low current consumption, full-scale range and high sensitivity within a compact footprint. The fourth paper by Reilly et al. reports a piezoelectric pressure sensing technology tailored to harsh environments, demonstrating functional stability up to 800 ºC in vacuum, and monolithic integration with JFET-R transistor technology. The fifth paper by S. Nayak et al. investigates the impact of ferroelectric switching on ScAlN-based BAW performance, linking the prolonged cycling with a degradation of the crystalline lattice. The sixth paper by J. Liu et al. combines PZT and ScAlN in a single, monolithically stacked PMUT to achieve 1.6x better pulse-echo sensitivity by employing PZT for actuation and ScAlN for detection.

2:05 PM

22-1 Backside-Etch Enabled 4× Sensitivity Enhancement of Optical Ultrasound Sensors in a 45-nm Electronic-Photonic CMOS Process

Sarika Madhvapathy[1], Panagiotis Zarkos[2,3], Tianqi Lu[4], Danielius Kramnik[4], Manuj Singh[5], Vladimir Stojanovic[4,6], Ali Niknejad[4], Jun-Chau Chien[4]

[1]Electrical Engineering and Computer Sciences, University of California, Berkeley, [2]Lawrence Berkeley National Laboratory, [3]Technical University of Crete, [4]University of California, Berkeley, [5]QuEra Computing Inc., [6]Ayar Labs

This work demonstrates sensitivity enhancement of silicon microring resonator-based ultrasound sensors for endoscopic imaging applications through silicon substrate removal and partial etch of the buried oxide (BOX) layer of a CMOS-SOI (silicon-on-insulator) die. Substrate removal is performed using xenon difluoride (XeF2) vapor etch and the BOX is thinned with reactive ion etch (RIE). We achieve an average intrinsic sensitivity (Sint) of 160.11 fm/kPa at a BOX thickness of < 0.1 μm, corresponding to an average improvement of 2.23× and maximum improvement of 4× compared to the baseline 2 μm BOX thickness.

2:30 PM

22-2 Ultra-High K^2_{eff} and Tunable $NaNbO_3$ Acoustic Devices on Silicon

Weifan Cai[1], Zhi Shiuh Lim[1], Tiancheng Luo[1], Qibin Zeng[1], Shengwei Zeng[1], Samantha Faye Duran Solco[1], Hui Kim Hui[1], Celine Sim[2], Chee Kiang Ivan Tan[1], Huajun Liu[1]

[1]Agency for Science, Technology and Research (A*STAR), [2]Nanyang Technological University

In this work, we present a transferred perovskite oxide $NaNbO_3$ (NNO) voltage-reconfigurable FBAR and filter on silicon. Through polymer-assisted transfer, the NNO layer is integrated as the active piezoelectric film on silicon while preserving its crystalline quality and field-switchable electromechanical functionality. The NNO FBAR exhibits electrically activated acoustic resonance, a ultra-high effective electromechanical coupling coefficient K^2_{eff} of 34–36%, a frequency tuning range of 15.2%, and a tuning efficiency of 2.5%/V. A compact 1.5-stage NNO filter further demonstrates an 18.4% fractional bandwidth and a 33% frequency-tuning range within a 0.2×0.2 mm^2 compact footprint. These results establish transferred NNO as a promising oxide acoustic platform for compact, tunable RF front-end filtering applications.

2:55 PM

22-3 An 83 nT/√Hz 0.032-mm^2 CMOS-MEMS Lorentz Force Magnetometer with ±49 mT Full-Scale Range

Ming-Huang Li, Wei-Bin Fan

National Tsing Hua University

This work presents a CMOS-MEMS Lorentz force magnetometer that employs a current-recirculation scheme and piezoresistive detection to enhance sensitivity, implemented in a TSMC 0.35-μm 2P4M CMOS process. The magnetometer is a torsional resonator operating at 168 kHz with a quality factor (*Q*) of 600 at ambient pressure. It consists of a central plate with an embedded stacked spiral coil, in which the recirculating current multiplies the Lorentz-force-induced torque, and is torsionally suspended by four stepped beams that enhance the piezoresistive sensing efficiency via the stress-concentration effect. With a 200-μA coil driving current and a 2.4-V bias voltage for the piezoresistive bridge, the proposed magnetometer achieves a sensitivity of 0.25 V_{rms}/T, a resolution of 83 nT/√Hz, a bias instability of 34.5 nT at an averaging time of 10.6 seconds, and a full-scale range of ±49 mT within 1% nonlinearity, all within a compact device footprint of only 0.032 mm^2.

3:45 PM

22-4 Technology Development Toward 800 °C Integrated Harsh Environment Transducers

Mark Sheplak[1], Alexander J. Reilly[2], Philip G. Neudeck[3]

[1]Interdisciplinary Microsystems Group, University of Florida, [2]University of Florida, [3]NASA Glenn Research Center

This invited paper describes the technology development of extreme-temperature (T), high-bandwidth aluminum nitride (AlN)-based piezoelectric dynamic pressure sensors towards advancing the state-of-the-art in harsh-environment integrated microsystems technology. Air-atmosphere testing of discrete AlN pressure sensors demonstrated bandwidths > 1 MHz and minimum detectable signals of less than 1 Pa/√Hz (at ≥ 1 kHz) approaching 900 °C. High-temperature testing of AlN films on tantalum silicide ($TaSi_2$) electrodes was performed to demonstrate viability for monolithic integration within NASA Glenn's 500 °C durable silicon carbide (SiC) Junction Field Effect Transistor – Resistor (JFET-R) integrated circuit (IC) technology. Functional stability of the film stack was verified up

to 800 °C in vacuum. With further development to fully demonstrate extreme-T durable monolithic integration, the combined AlN + SiC JFET-R platform would enable a variety of integrated sensors, resonators, and ferroelectric memory operational in numerous relevant harsh environment applications.

4:10 PM

22-5 Electrical Polarity Switching Induced Lattice Strain and Reliability of Ferroelectric ScAlN Thin Films and GHz BAW Resonators

Sanjay Nayak, Zhi-Qiang Lee, Ying Zhang, Shashidhara Acharya, Minghua Li, Poh Chong Lim, Jihang Liu, Chen Liu, Xinghua Wang, Kui Yao, Huamao Lin, Yao Zhu, Yee-Chia Yeo

Agency for Science, Technology and Research (A*STAR)

Ferroelectric switching in multilayer ScAlN bulk acoustic wave resonators (BAW) enables polarization-controlled frequency tuning, but repeated cycling can induce performance drift. In this work, we report the first use of 2D-X-ray diffraction to investigate the evolution of lattice strain in switched ferroelectric $Sc_{0.3}Al_{0.7}N$ and correlate the structural changes to GHz BAW resonator performance. Complementary test structures, including clamped capacitors, released membranes, and functional BAW resonators, are employed to decouple the contributions from lattice-strain, macroscopic deformation and leakage to the acoustic response and ferroelectric fatigue. The results reveal that moderate polarization switching initially enhances the coupling coefficient by ~1% and **Q-**factor by 300% of BAW resonator at 3.5 GHz, whereas prolonged cycling results in increased leakage and deterioration of resonator performance. This work provides insight into why BAW resonator performance evolves under repeated ferroelectric switching and establishes a framework for assessing the reliability of ScAlN-based reconfigurable ferroelectric BAW resonators and filters.

4:35 PM

22-6 An 8-inch PMUT Platform with Vertically Integrated PZT for TX and $Sc_{0.3}Al_{0.7}N$ for RX Achieving Record High Transceiver Sensitivity at 158 kHz

Jihang Liu, Xinghua Wang, Prakasha Chigahalli Ramegowda, David Choong Sze Wai, Trivedi Shyam, Goh Duan Jian, Shashidhara Acharya, Kui Yao, Huamao Lin, Qingxin Zhang, Yul Koh, Yao Zhu, Yee-Chia Yeo

Agency for Science, Technology and Research (A*STAR)

Single-material PMUTs face a fundamental transceiver sensitivity ceiling: PZT's high transverse piezoelectric coefficient ($e_{31,f}$) maximizes transmission, while its large permittivity (ε_r~1100) severely limits voltage reception. Conversely, $Sc_{0.3}Al_{0.7}N$'s low ε_r (~18) enables high-voltage sensitivity but its moderate $e_{31,f}$ limits output. A hybrid PZT + ScAlN PMUT that assigns PZT to TX and ScAlN to RX on the same membrane can circumvent this constraint, but prior realization has not seen the desired performance boost due to piezoelectric texture degradation when integrating both materials on the same wafer. We present the first 8-inch PMUT platform with vertically integrated PZT and $Sc_{0.3}Al_{0.7}N$ on silicon, achieving supreme film quality. By driving TX with PZT and receiving with $Sc_{0.3}Al_{0.7}N$, a record-high PMUT TRX figure-of-merit of 626×10^{-5} at 158 kHz in air, a 1.6× improvement of the best single-material PMUT, confirming the hybrid PZT + ScAlN approach as a scalable strategy for next-generation high performance ultrasonic transceivers.

Session 23: MS | Atomic Simulation: Devices, Circuits, and Process

2:00 PM, Continental 4

Co-Chairs: Changwook Jeong, UNIST and Gengchiau Liang, NYCU

This session highlights how atomistic and multiscale simulation is evolving from a tool for fundamental materials analysis into a predictive framework spanning semiconductor processes, devices, reliability, and circuits. The papers bridge first-principles calculations, molecular dynamics, machine-learning interatomic models, and quantum transport with process, compact-model, and circuit-level simulation. Applications include the scaling limits of ultra-thin Si nanosheets, process-aware quantum transport in strained SiGe nanosheet FETs, atom-to-device modeling of logic and memory technologies, and defect-mediated dynamics in ferroelectric capacitors. The session further extends atomistic modeling beyond the device level: one work connects first-principles material properties through compact modeling to array-level optimization of selector-only memory, while another translates atomistic etch reaction kinetics into compact, aspect-ratio-dependent process laws for advanced IC fabrication. Together, the six papers demonstrate a common direction toward cross-scale, physics-based simulation in which

atomic-scale information is systematically propagated to experimentally relevant device, circuit, and process metrics.

2:05 PM

23-1 Atom-to-Circuit Modeling Framework for Low-Power Selector-Only Memory: From First Principles to Array-Level Optimization

Ha-Jun Sung, Minwoo Choi, Jaechul Park, Jun Hyuk Park, Youngmin Lee, Zhe Wu, Youngjae Kang, Jinseong Heo, Kiyeon Yang
Samsung Electronics

We present an atom-to-circuit multiscale framework for the design of low-power SOM. The framework spans DFT, MD, TCAD device simulation, and array-level circuit analysis. Partial incorporation of Te into the Ge-As-Se baseline lowers the threshold voltage while keeping the leakage increase gradual, unlike the Sb substitution. The DFT-derived band gap and trap density feed a physics-based compact model whose parameters are extracted by Bayesian optimization. Array-level simulation under a half-bias scheme then quantifies write and read energy and sensing margin. The results show that the optimal material is not fixed but depends on array size, read scheme, and application.

2:30 PM

23-2 Atom-to-Device Scale-bridging Simulation for Angstrom Node

Uihui Kwon, Min Yong Jeong, Kyoung Yeon Kim, Bokyeom Kim, Kantawong Vuttivorakulchai, Woon Ih Choi, Yong-Hee Cho, Mohammad Ali Pourghaderi, Yoon-Suk Kim, Hyo-Shin Ahn, Márton Vörös, Guido Gandus, Byounghak Lee, Hong-Hyun Park, Seonghoon Jin, Woosung Choi, Jaehoon Jeong, Jaihyuk Song
Samsung Electronics

To enhance the performance and yield of ultra-scaled CMOS devices entering the Angstrom-node era, it is crucial to establish an atom-to-device (A2D) scale-bridging simulation framework that seamlessly integrates machine learning interatomic potentials - molecular dynamics (MLIP-MD), density functional theory (DFT), and non-equilibrium Green's function (NEGF) quantum transport. By applying this workflow to several representative logic and memory devices, we demonstrate how A2D simulations can reveal the underlying physics governing state-of-the-art devices and overturn long-standing technical dogmas in this field.

2:55 PM

23-3 Scaling Limit of Ultra-Thin Si Nanosheets: Insights from DFT-NEGF Transport and Experiments

Aryan Afzalian, Sylvain Baudot, Dmitry Batuk, Cristian Rebolledo Espinoza, Tom Schram, Farid Sebaai, Naoto Horiguchi, Paul Heremans
imec

Using the DFT-NEGF formalism and experiments, we explore the fundamental scaling limits of ultra-thin Si Nanosheets. Simulations predict Si could be scaled down to t_{Si} ~ 1.5 to 2 nm and L = 7 nm. NEGF predictions are supported by experimental electron mobility measurements.

3:45 PM

23-4 Quantum Transport Simulation of p-type SiGe Nanosheet Transistors: A Rigorous Process-to-Device Approach Including Strain and Alloy Effects

Gyeongwon Roh[1], Min-Seo Jang[2], Phil-Hun Ahn[1], Sung-Min Hong[2], Mincheol Shin[1]
[1]Korea Advanced Institute of Science and Technology (KAIST), [2]Gwangju Institute of Science and Technology

A rigorous process-to-quantum-transport simulation framework is developed for p-type SiGe nanosheet FETs by addressing two key limitations of conventional modeling: empirical treatment of alloy disorder and uniform strain. An in-house process simulator extracts the full position-dependent strain tensor accumulated throughout process. Random-alloy disorder is captured using a first-principles-based absolute potential model, enabling the alloy contribution to be evaluated without calibration against empirical measurements. NEGF simulations then quantify the combined effect of strain, alloy-disorder, and hole-phonon scattering

4:10 PM

23-5 First ML-Driven Atomistic Compact Etch Model: From Molecular Dynamics to Predictive Etch-Rate Laws for Advanced IC Manufacturing

Yuxuan Zhai, Yunxiao Zhang, Zemeng Feng, Minrong Zhang, Rui Ge, Ziyi Hu, Hongli Song, Qikai Zhao, Hua Shao, Dashan Shang, Zhiqiang Li, Guang Yang, Xiaolei Wang, Junfeng Li, Rui Chen, Junjie Li, Ling Li
Institute of Microelectronics of Chinese Academy of Sciences

As IC manufacturing advances to sub-3-nm nodes and complex 3D architectures, precise process control requires a unified treatment of geometry-dependent species transport and surface reactions. Conventional models often treat these effects separately and have limited applicability. This work presents the first ML-driven atomistic compact etch model, a general cross-scale methodology that bridges reaction kinetics and feature-scale profile evolution to derive compact aspect-ratio (AR)-dependent etch-rate laws. Validated for ion-assisted vertical Si etching and radical-driven lateral SiGe release in GAAFET fabrication, the framework captures experimental profile evolution, while the extracted laws describe AR-dependent etch-rate variations with R^2 up to 0.994. Compared with direct atomistic MD simulations, the model achieves up to 26.7× speedup while retaining atomistic reaction information, enabling efficient process exploration for advanced IC manufacturing.

4:35 PM

23-6 A Defect-centric Framework for Universal Description of Ferroelectric Capacitor Dynamics: from Wake-up to Access-dependent Fatigue

Xiaoyu Dou[1], Xiaopeng Li[2], Yilin Hou[3], Yaoyu He[3], Yang Feng[2], Pengpeng Sang[3], Xuepeng Zhan[3], Jingrui Guo[3], Jixuan Wu[3], Jiezhi Chen[3]
[1]School of Information Science and Engineering, Shandong University, [2]National University of Singapore, [3]Shandong University

In this work, a defect-centric multi-scale modeling framework is developed to enable a universal description of the dynamic behaviors of HfO_2-based ferroelectric capacitors. The framework integrates *ab initio* calculations with a fully-coupled physical model that accounts for phase-field and oxygen vacancy (V_O) dynamics, including defect generation, drift-diffusion, and charge state conversion. Leveraging the proposed unified defect-centric ferroelectric dynamics (UDFD) framework, for the first time, all dynamic behaviors of HZO capacitors are captured self-consistently within a unified simulation platform, including wake-up, fatigue, imprint, and access-rate dependence. The proposed UDFD model provides a comprehensive physical understanding of the defect-related dynamic evolution of ferroelectric film and offers theoretical guidance for process optimizations.

Session 24: NC | LLM Acceleration Using In-Memory Compute

2:00 PM, Continental 5

Co-Chairs: Ravi Kumar, Infineon and Jun Deguchi, Kioxia

The seven papers in this session explore advanced compute-in-memory and processing-in-memory architectures designed to overcome memory and communication bottlenecks in edge LLM and Mixture-of-Experts inference. The first paper from Peking University demonstrates a 288-Mb PCM compute-in-memory processor that enables dynamic KV-cache processing for edge LLMs by combining hot/warm/cold tri-tier placement, asymmetric K/V quantization, and low-stress multilevel programming to improve endurance, attention fidelity, and write efficiency. The invited paper from SK hynix discusses use of Accelerator-in Memory (AiMX) cards for LLM inference acceleration. This is achieved by using a multi-rank parallel architecture for PIM based attention disaggregation that increases effective memory bandwidth. The third paper from Chinese Academy of Sciences demonstrates a hybrid analog–digital compute-in-memory architecture for efficient LLM inference by co-fabricating 3D high-endurance SLC FeFET arrays for dynamic KV-cache processing and high-density MLC FeFET arrays for static FFN-weight computation on the same wafer. The fourth presentation by Peking University demonstrates a 232-layer 3D-NAND in-memory-computing architecture with a source-line shutter array that enables dynamic expert selection and routing-weighted accumulation for accelerating LLM inference with MoE. The paper from Tsinghua University presents a monolithically integrated 3D hardware architecture that combines RRAM-CIM, ITZO-based routing, and CNT-CMOS logic to accelerate computation and dynamic token communication in MoE Transformers. The paper from Peking University demonstrates a new expert-mapping and weight-encoding architecture on FeNAND to boost MoE inference in LLMs. This architecture-circuit-device co-design achieves 90x higher throughput and more than 4x

efficiency. The last presentation from Macronix presents a 192-layer 3D-NAND-based digital in-memory-computing accelerator that reuses existing FBD/FBC circuits and combines query-adaptive threshold prediction with IVF and CPU/GPU processing to accelerate large-scale approximate nearest-neighbor search while reducing data movement and DRAM usage. Together, these papers highlight hardware–algorithm co-design as a key path toward faster, more efficient LLM systems.

2:05 PM

24-1 A 288-Mb PCM Compute-in-Memory Processor with Drift-Resilient Tri-Tier KV-Cache Computing and Low-Stress Multilevel Programming for Edge LLM Inference

Longhao Yan[1], Xiangyuan Xue[1], Yihang Zhu[1], Xi Li[2], Chenchen Xie[2], Zhiyuan Li[1], Minghui Yin[1], Bowen Wang[1], Zelun Pan[1], Zeyao Tang[1], Yaoyu Tao[1], Zhitang Song[2], Yuchao Yang[1]
[1]Peking University, [2]Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences

As context grows, edge large language model (LLM) inference is bottlenecked by KV-cache storage and repeated access. Phase-change memory (PCM) compute-in-memory (CIM) provides dense nonvolatile storage and in-situ attention computing, but frequent writes consume endurance, conductance drift degrades attention fidelity, and abrupt RESET transition causes over-reset and corrective SET operations, increasing energy and stress. We present a 40-nm 288-Mb PCM CIM processor with architecture–programming co-design. Drift-resilient tri-tier KV-cache computing maps hot, warm, and cold data to an SRAM processing engine, digital PCM CIM, and analog PCM CIM. Its wear-aware asymmetric quantization and hotness-guided scheduling extend device lifetime by 4.67× and reduce attention-output error by 66.9%. Low-stress multilevel programming adaptively scales RESET pulses with the current-to-target conductance gap, reducing programming pulses by 54% and write energy by 68.5%. A four-board 1.125-Gb prototype achieves 235–297× and 14–18× reductions in energy and latency per token versus an A100 GPU on Qwen2.5-7B and LLaMA2-7B.

2:30 PM

24-2 Addressing the AI Memory Bottleneck with PIM/PNM-Based Attention Disaggregation (Invited)

Haerang Choi, Gyeongcheol Shin, Changhyun Kim, Hyunha Joo, Byeongju An, Jangyong Lee, Taeha Kim, Inhwan Lee, Woojae Shin, Jongsoon Won, Guhyun Kim, Jeongbin Kim, Dayeon Yun, Jaehan Park, Yosub Song, Byeongsu Yang, Yousub Jung, Hyeongdeok Lee, Wonjun Lee, Seungyeong Park, Seonghun Kim, Yonghoon Park, Ilkon Kim, Euicheol Lim
SK Hynix

Processing multiple large language model (LLM) requests concurrently improves model-weight reuse, but Decode attention remains bandwidth-bound because each request accesses its own KV cache. We propose a multi-rank parallel architecture for PIM-based attention disaggregation that increases effective memory bandwidth, providing decode attention with throughput gains otherwise unavailable from batching. We implement and evaluate it on a dual-rank AiMX prototype, where broadcasting data-bus-free PIM compute commands enables concurrent execution across ranks. Single- and dual-rank execution deliver average effective memory bandwidths of 8.5× and 12.4× the device-interface baseline, respectively; dual-rank execution reaches 13.3× when emulating refresh-free disaggregated memory mode under continuous KV-cache access. These results demonstrate multi-rank scaling of both capacity and throughput in attention-oriented PIM/PNM systems.

2:55 PM

24-3 Hybrid Analog–Digital LLM-CiM Enabled by 3D Co-Fabricated Vertical Dual-Gate SLC and GAA MLC FeFET Arrays with Distinct Architectures

Chen Gu, Haochen Yu, Zhaori Cong, Weiwei Li, Qianqian Wei, Zhenkun Chen, Yuan Xue, Junzhe Shen, Minggao Zhan, Yiyang Yuan, Cailian Ma, Wenfeng Jiang, Chuanke Chen, Xinlv Duan, Xin Liu, Zuowei Zhang, Congyan Lu, Shengjie Zhao, Kaiping Zhang, Jiebin Niu, Yu Liu, Pengfei Jiang, Jinshan Yue, Guanhua Yang, Zhicheng Wu, Nianduan Lu, Di Geng, Qing Luo, Ming Liu
Institute of Microelectronics of Chinese Academy of Sciences

We demonstrate a hybrid analog–digital (A&D) large language model (LLM) compute-in-memory (CiM) design enabled by same-wafer co-fabrication of 4-layer 3D dual-gate SLC and gate-all-around MLC FeFET arrays with distinct device and array architectures. The SLC devices achieve 100-ns PGM/ERS operation and $>10^{11}$ endurance. The MLC devices provide 3-bit/cell storage with projected 10-year retention. At the array level, the SLC architecture

enables current-domain in-array matrix–vector multiplication for runtime-updated KV-cache/attention data. The MLC architecture supports multi-bit weight readout followed by digital shift-and-add for static FFN weights. The measured arrays demonstrate sub-7.5-ns read latency, low-disturb operation, linear SLC MVM, and high-linearity MLC weight fitting. At the architecture level, the proposed coarse-grained inter-layer operator pipeline and fine-grained head-streaming pipeline, together with KV-cache-capacity-aware buffer reuse, coordinate the MLC-based DCiM and SLC-based ACiM paths. The design achieves 28.13 TOPS/mm^2, 273.44 TOPS/W, 9.65 μJ/token, and 339 token/s, validating a hybrid A&D-CiM architecture for LLM inference.

3:20 PM

24-4 Compute-in-3D NAND with Source-Line Shutter and Xtacking Technology for Efficient Mixture-of-Experts Large Language Model Acceleration

Tianyang Luo[1], Zongwei Wang[1], Dongxue Zhao[1], Guanqi Wang[1], Kuijun Zeng[1], Ling Liang[1], Lin Bao[1], Lei Jin[2], Zhiliang Xia[2], Zongliang Huo[2], Nanxiang Chen[2], Yimao Cai[1], Ru Huang[1]
[1]Peking University, [2]Yangtze Memory Technologies Co., Ltd. (YMTC)

In this work, we propose and implement Compute-in-3D NAND architecture to accelerate Large Language Model (LLM) inference with Mixture-of-Experts (MoE). The weight-encoded NAND cells and bit-line (BL) current accumulation perform parallel vector-matrix multiplication (VMM). A source-line shutter array (SSA) is introduced to partition common source line of 3D NAND blocks into fine-grained segments, enabling token-dependent expert selection and weighted accumulation. The SSA enables two operating modes: 1) Binary Shutter (B-shutter) for parallel execution of dynamically selected experts; and 2) Multilevel Shutter (M-shutter) with 16 calibrated levels for **weighted accumulation of selected experts' down-projection outputs.** Furthermore, leveraging Xtacking technology, a cost-effective three-tier architecture is proposed by integrating the NAND computing array with dedicated compute/readout circuits and shutter-control CMOS. System-level evaluation demonstrates a peak throughput of 47.9 TOPS@INT8 with only 3.9-percentage-point accuracy loss compared with an FP16 software baseline, highlighting the scalability and bandwidth efficiency of 3D-NAND IMC for next-generation MoE inference.

4:10 PM

24-5 Enabling All-to-All Connectivity for MoE-based Transformers Through Monolithic 3D Integration of ITZO-based NoC with CNT CMOS and RRAM CIM

YIBEI ZHANG[1], Jianshi Tang[1], Lei Gao[2], Ningfei Gao[2], Haitao Xu[2,3], Siyu Xu[1], Kexin Wang[1], Weijin Wang[1], Han Zhao[1], Yanbo Su[1], Zhidong Tang[1], Yijia Fan[1], Zhixing Jiang[1], Jiaqi Yang[4], Bonan Yan[4], Peng Yao[1], Qingtian Zhang[1], Dong Wu[1], Bin Gao[1], He Qian[1], Huaqiang Wu[1]
[1]Tsinghua University, [2]Beijing Institute of Carbon-based Integrated Circuits, [3]Institute of Carbon-based Thin Film Electronics, Peking University, China, [4]Peking University

MoE Transformers can reduce the number of activated parameters per token through the expert selection mechanism, but meanwhile introduce large-volume dynamic token routing between the gating logic and distributed expert blocks. This bidirectional all-to-all traffic dominates the latency and wiring cost when implemented with conventional 2D interconnects. To address this bottleneck, here we present a monolithic 3D network-on-chip (M3D-NoC) chip designed for accelerating MoE-oriented Transformers. On the BEOL of Si CMOS, the M3D-NoC chip vertically integrates a RRAM-based CIM layer for expert computations, an ITZO-FET-based NoC layer for token routing, and a CNT CMOS layer for gating network and weighted sum. Token movement is handled by the dedicated ITZO-FET-based NoC layer with router arrays and inter-layer vias, rather than through long 2D global wires. System-level benchmark reveals a 4.41× speed-up compared to GPU while reducing the area by 81.73%.

4:35 PM

24-6 Novel Mask-Routed Charge-Domain Δ-In-Storage-Computing on FeNAND Array with Sudoku-Like Expert-Mapping and Coprime-Factor Weight-Encoding for High-Throughput Mixture-of-Experts LLM Inference

Bingrui Song[1], Shaodi Xu[1], Weikai Xu[1], Yuxin Lin[1], Runze Han[1], Shengjie Cao[1], Jin Luo[1,2], Qianqian Huang[1,3], Ru Huang[1,3]
[1]Peking University, [2]Beihang University, [3]Beijing Advanced Innovation Center for Integrated Circuits

A novel mask-routed charge-domain delta-in-storage-computing (MR-ΔISC) is proposed and experimentally demonstrated on ferroelectric-NAND (FeNAND) array through architecture-circuit-device co-design to boost

throughput for mixture-of-experts (MoE) inference in edge LLMs. Architecturally, proposed Sudoku-like mapping interleaves partitions of expert matrix across a word-line (WL) layer in 3D FeNAND structure, enhancing spatial parallelism along both source-line (SL)-input and bit-line (BL)-output dimensions under sparse MoE activation. Two hardware innovations support this mapping: (1) Mask WL layers employ series-connected multilevel-V_{th} cells to realize string-level combinatorial gating for *in-situ* expert-ID selection and mapped-partition activation. (2) Weight WL layers implement charge-domain ΔISC using coprime-factor weight-encoding and a compact ferroelectric Δ-DAC for sparse bipolar Δ-input. Prime factorization increases zero-valued weight cells, while string-current-clamped temporal V_{th} access enables accurate BL-charge accumulation. Using a FeNAND array with a record-high memory-window-to-subthreshold-swing ratio for temporal V_{th} access optimization, the demonstrated MR-ΔISC system achieves 89.6×/4.6× improved throughput and energy efficiency for edge MoE LLMs.

5:00 PM

24-7 3D-NAND Digital In-Memory Computing (DIMC) Accelerator with Predicted Threshold and Inverted File Index Algorithm

Po-Hao Tseng[1], Chien-Liang Yu[1], Chung-Han Wu[1], Cheng-Hsuan Lin[2], Feng-Ming Lee[1], Yu-Hsuan Lin[1], An-Yeu Wu[2], Hsiang-Yun Cheng[3], Ming-Hsiu Lee[1], Kuang-Yeu Hsieh[1], Keh-Chung Wang[1], Kuen-Long Chang[1], Chih-Yuan Lu[1]
[1]Macronix International Co., Ltd., [2]National Taiwan University, [3]Academia Sinica

We proposed a high performance approximate nearest neighbor search (ANNS) system using 3D-NAND based digital in-memory computing (DIMC) accelerator with predicted fail bit detection (FBD) threshold and inverted file index (IVF) algorithm for AI applications in large database. This IVF-ANNS system achieves device-algorithm co-optimization by integrating the advantages of CPU/GPU (fast computing) and the DIMC accelerator (high memory density, accurate parallel distance computing), resulting in low DRAM usage, high recall rates and low system latency. Typical fail bit detection (FBD) or fail bit counting (FBC) circuits in the 3D NAND chip are reused to execute the distance computing function, where a novel pre-trained dynamic threshold prediction model is used to optimize the FBD threshold setup for each user query. The proposed system is experimentally demonstrated on SIFT-1B, achieving over 98% recall rate, fast digital computing with FBD, robust digital computing reliability, 20X latency reduction, and reduces DRAM usage by 22.5X.

Session 25: EDT | Oxide-Channel Devices

2:00 PM, Continental 6

Co-Chairs: Ching-Tzu Chen, IBM Research and Michelly de Souza, Centro Universitario FEI

This session highlights the latest advancements in oxide-channel devices, spanning module development, device characterization, and device integration. The session starts with a report of record-high mobility in In_2O_3 channel FETs via kinetically enhanced crystallization. The second paper teaches how to mitigate the mobility, threshold-voltage and stability tradeoff in amorphous InO_x devices via N-Ge co-doping. The third paper demonstrates contact scaling down to 10 nm, maintaining low contact resistance via Ru source/drain engineering in channel-all-around IGZO devices. The following paper introduces a high-thermal-conductivity BeO layer under In_2O_3 for heat dissipation, which significantly improves hot-carrier degradation stability. The next paper explores how channel passivation with thermal treatment reduces back-channel defects, enhancing thermal stability and device performance in p-type SnO devices. Also highlighted in this session is a device-level DoS mapping method that reveals how band-tail states affect device performance, using InO_x and IGZO as examples. The section concludes with two monolithic integration papers. One features the 1T-1V (varactor) capacitive gain-cell on 40 nm foundry Si-CMOS for digital CIM. The other showcases a fully functional monolithic 3D integrated IP library, featuring IGZO/CNT CFET-based SRAM-CIM/Gain Cell-PIM hybrid macro.

2:05 PM

25-1 Record-High Mobility (236.4 $cm^2 \cdot V^{-1} \cdot s^{-1}$) in BEOL-Compatible Crystalline In_2O_3 FETs Featuring Kinetically Enhanced Crystallization

Huangbai Liu[1], Yan Le[2], Leming Jiao[2], Wei Shi[2], Weibing Hao[2], Jiawei Xie[2], Rui Shao[2], Yuxuan Wang[2], Xinming Li[2], Dong Zhang[2], Kai Ni[3], Gengchiau Liang[4], Chen Sun[2], Xiao Gong[2]

[1]ECE, National University of Singapore, [2]National University of Singapore, [3]University of Notre Dame, [4]National Yang Ming Chiao Tung University

In this work, we propose a novel low-temperature KEC process at only 300 ℃ to realize high-performance crystalline In_2O_3 FETs (C-In_2O_3 FETs). Compared with amorphous In_2O_3 (A-In_2O_3) counterparts, the KEC-enabled C-In_2O_3 FETs exhibit a 118.4% enhancement in effective mobility (μ_{eff}), achieving a record-high 236.4 $cm^2 \cdot V^{-1} \cdot s^{-1}$ among BEOL-compatible oxide-semiconductor devices. We further experimentally reveal, for the first time, that the mobility of C-In_2O_3 exhibits weak dependence on in-plane channel orientations. Unlike A-In_2O_3 FETs that suffer from substantial mobility degradation at 125 ℃, the C-In_2O_3 FETs maintain highly stable mobility over a wide temperature range from 25 to 125 ℃. Moreover, aggressively scaled devices with a L_{CH} of 30 nm deliver an ultrahigh I_{ON} of 7.7 mA/μm at V_{OV} = 4 V and V_{DS} = 2 V. Bias-temperature-instability measurements further reveal that KEC significantly suppresses hydrogen-related instability in C-In_2O_3 FETs, yielding a ΔV_{TH} of only ~40 mV after 1 ks of stress.

2:30 PM

25-2 Breaking the Mobility–Threshold-Voltage–Stability Tradeoff in Amorphous Indium Oxide through Anion–Cation Co-Doping

Xingtian Wang[1], Md Abdullah Al Mamun[2], Jiahui Duan[1], Aman Chandrakar[1], Yiyan Yu[1], Yubo Wang[1], Guanyu Zhou[1], YuShan Lee[1], Pengyu Ren[1], Siyuan Li[1], Sizhe Ma[1], Yixin Qin[1], Xiao Gong[3], Kyeongjae Cho[2], Christopher Hinkle[1], Kai Ni[1]

[1]University of Notre Dame, [2]University of Texas at Dallas, [3]National University of Singapore

AOS TFTs face a tradeoff among mobility, V_{TH}, and electrical stability. Conventional cation doping controls carrier concentration and V_{TH}, but disrupts the In-5s network and reduces mobility. Here, we demonstrate an anion-first, cation-assisted strategy in In_2O_3. N incorporation lowers carrier concentration while largely preserving the In-5s network. DFT shows that N suppresses V_O, and acts as an acceptor with minimal disruption to the In-5s states. Because N alone does not fully suppress bias-stress instability, a small amount of Ge is added to stabilize the oxide network. This approach enables amorphous Ge- and N-co-doped In_2O_3 TFTs with V_{TH} =0.3 V, mobility of 60 $cm^2/V \cdot s$, a V_{TH} shift of only 50 mV under 2 MV/cm bias stress for 1000 s, and a drive current of 140 $\mu A/\mu m$ at V_{OV}=V_D=1 V for L_G=100 nm. These results show that anion-first, cation-assisted co-doping can separately control carrier density, transport, and stability in high-performance AOS TFTs.

2:55 PM

25-3 Enabling Contact Scaling Down to 10 nm with Ultra-low Contact Resistivity via Ruthenium Source/Drain Engineering in Channel-All-Around IGZO FETs

Qingding Tong, Tong Yu, Congyan Lu, Yu Liu, Jiebin Niu, Kaiping Zhang, Shengjie Zhao, Chuanke Chen, Weiwei Li, Nianduan Lu, Di Geng, Ling Li, Ming Liu

Institute of Microelectronics of Chinese Academy of Sciences

For the first time, we demonstrate an aggressive contact length (L_C) scaling down to 10 nm in vertical Channel-All-Around (CAA) IGZO FETs while maintaining an ultra-low specific contact resistivity (ρ_C) of $6.32\times10^{-8}\ \Omega \cdot cm^2$ through Ruthenium Source/Drain (Ru S/D) engineering. The scaled CAA FET with L_C of 10 nm and CD of 65 nm exhibits a low contact resistance (R_C) of 632 Ω·μm, together with a high I_{ON} of 122 μA/μm(@V_{OV}=2V, V_{DS}=1V) and a steep SS of 74 mV/dec. First-principles calculations further reveal the microscopic origin of the superior Ru/IGZO contact. This work establishes Ru as an effective contact technology for aggressive contact scaling in oxide-semiconductor transistors, providing a practical pathway toward future high-density 3D DRAM arrays.

3:20 PM

25-4 First Demonstration of High-Performance Top-Gate In_2O_3 RF Transistors with Alleviating Self-Heating via BEOL-Compatible ALD Polycrystalline BeO

Peiyan Hong[1], Yixing Wang[2], Jiayao Hao[1], Kexin Zhang[1], Kun Zhang[2], Kai Wang[1], Puqing Jiang[1], Hong Zhou[2], Xiangshui Miao[1], Xuefei Li[1]

[1]Huazhong University of Science and Technology, [2]Xidian University

In this work, we first demonstrate high-performance top-gate (TG) In_2O_3 radio-frequency (RF) transistors integrated with high-thermal-conductivity beryllium oxide (BeO) substrates via atomic-layer deposition (ALD). The thermal conductivity (κ_{th}) of the ALD ultrathin BeO thin films is 10.79 W/m·K, which is 7.7 times greater than that of conventional amorphous SiO_2. Owing to this superior thermal management, In_2O_3 power transistors on BeO

substrates exhibit higher breakdown voltages (V_{BD}) and breakdown powers (P_{BD}) along with significantly improved hot-carrier degradation (HCD) stability. Moreover, our short-channel TG In_2O_3 FETs exhibit excellent electrical performance, including a high on-state current (I_{on}) of 3.8 mA/μm, an ultrahigh on/off ratio exceeding 10^{12}, and a small DIBL of 20 mV/V. At a gate length (L_g) of 140 nm, the RF transistors have a record-high cutoff frequency (f_T) × L_g of 6.5 GHz·μm. These comprehensive results highlight ALD BeO as a promising high-thermal-conductivity material for efficiently dissipating heat in BEOL monolithic 3D integration.

4:10 PM

25-5 Distinct Role of Passivation Engineering in p-Type Oxide Transistors: Enabling Thermal Stability and Performance Improvement

Chi-Hsin Huang[1], Yalun Tang[2], Yilong Song[2], Shengman Li[3], Yu-Sen Jiang[3], Koustav Jana[3], Xinxin Wang[3], Hugo J.-Y. Chen[3], Shuhan Liu[3], Stacey F. Bent[3], Kenji Nomura[2], H.-S. Philip Wong[3]
[1]Electrical Engineering, Stanford University, [2]University of California, San Diego, [3]Stanford University

We demonstrate that passivation plays a critical yet fundamentally distinct role in *p*-type SnO back-gated field-effect transistors compared with *n*-type oxide transistors. Beyond terminating back-channel defects, passivation preserves the SnO phase during high-temperature processing, widening the process window and enabling back-end-of-line (BEOL)-compatible integration. After passivation processing incorporating thermal treatment, SnO transistors with a channel length of 80 nm achieve an enhanced on-current of **8.4** μA/μm at a drain-to-source voltage of –1 V and a gate-to-source voltage of –4 V, and a record-low subthreshold swing of 197 mV/dec.

4:35 PM

25-6 Direct In-Device Band-Tail DOS Mapping of OS TFTs Using Electric-Field-Penetration Technique

Kisoo Nam, Chang Niu, Aravindh Shankar, Jian-Yu Lin, Sanjeev Khare, Sumi Lee, Pramey Upadhyaya, Peide D. Ye
Purdue University

In this work, we demonstrate a device-level density-of-state (DOS) mapping method in fabricated oxide semiconductor (OS) thin-film-transistors (TFTs) using an electric-field-penetration technique (eFPT) with a single gate-bias sweep. By comparing the eFPT-extracted DOS with transfer and transconductance characteristics, we clarify how band-tail states affect transistor operation. Frequency-dependent measurements show a dispersive low-energy band-tail state and frequency-independent higher-energy band-tail state. A systematic comparison of InO_x and IGZO channel TFTs shows that the extracted DOS consistently follows the transfer characteristics shift across different channel lengths (L_{ch}). Furthermore, larger L_{ch} devices exhibit improved subthreshold swing (SS), which cannot be explained solely by shallow/interface trap variation. By correlating this trend with the extracted band-tail widths, we show that the higher-energy band-tail width provides an additional contribution to SS degradation. This work provides an experimental framework for linking band-tail DOS in OSs to transistor operation, offering insight into optimization of device design and scaling.

5:00 PM

25-7 Logic-Compatible Capacitive Gain Cell Monolithically Integrated on Si CMOS for Bit-Cost-Scalable 3D Digital Compute-in-Memory

Junmo Lee[1], Kiseok Lee[1], Chengyang Zhang[1], Tae-Hyeon Kim[1], Mengkun Tian[1], Leo Jen-Chieh Liu[2], Meng-Fan Chang[2], Suman Datta[1], Shimeng Yu[1]
[1]Georgia Institute of Technology, [2]Taiwan Semiconductor Manufacturing Company (TSMC)

We experimentally demonstrate a BEOL-compatible 1T–1V capacitive gain-cell (Cap-GC) monolithically integrated on foundry 40nm CMOS for 3D Cap-GC digital compute-in-memory (DCIM). The Cap-GC combines an ALD W-doped In_2O_3 (IWO) field-effect varactor (FE<u>V</u>) and an IWO FE<u>T</u>, where the storage-node (SN) voltage modulates gain-cell "capacitance". Unlike conventional 2T-GCs, capacitive coupling is leveraged to suppress SN disturbance and avoid the large read-bitline (RBL)-capacitance-induced delay. Compared with ferroelectric nonvolatile capacitors (nvCAPs), Cap-GCs achieve a ~20× capacitive on/off ratio at a lower, logic-compatible write voltage (≤1V). Further, multiply-and-sense (MAS), the key step in Cap-GC read, is experimentally demonstrated from the device to the array level. Building on the experimental M3D integration, we propose a bit-cost-scalable (BiCS) 3D Cap-GC DCIM architecture, comprehensively evaluated through SEMulator3D® process emulation, 3D parasitic extraction, and experimentally-calibrated simulations. The projected architecture achieves <1ns read latency, <0.1V SN disturbance, and a 33.1× higher density FoM than conventional 10T SRAM-DCIM at 3nm.

5:25 PM

25-8 A 340-kHz M3D Integrated IP Library Featuring IGZO/CNT CFET-based SRAM-CIM/Gain Cell-PIM Hybrid Macro for Operator-Aware LLM Inference

Yanbo Su[1], Jianshi Tang[1], Yijia Fan[1], Jiaqi Yang[2], Kexin Wang[1], Zhidong Tang[1], Zhixing Jiang[1], Junyang Zhang[1], Yibei Zhang[1], Bowen Shen[1], Nianzi Sui[3], Yide Li[1], Bingsheng Qiu[3], Jianwen Zhao[3], Song Qiu[3], Qingwen Li[3], Bonan Yan[2], Peng Yao[1], Bin Gao[1], He Qian[1], Huaqiang Wu[1]

[1]Tsinghua University, [2]Peking University, [3]University of Science and Technology of China (USTC)

In this work, we experimentally demonstrate a fully functional monolithic 3D (M3D) integrated IP library, featuring a hybrid macro with SRAM computing-in-memory (CIM) and gain cell (GC) processing-in-memory (PIM) based on back-end-of-line (BEOL)-compatible IGZO/CNT CFET technology. Comprehensive statistical evaluations across 4-inch wafers confirm exceptional manufacturing uniformity with an overall high yield exceeding 92%. In particular, the fabricated BEOL circuits achieve the state-of-the-art (SOTA) operating frequency of 340 kHz at a 2μm technology node. Diverse complex computing modules are fully verified. This M3D hybrid architecture successfully implements operator-aware hardware mapping, delivering over 3× system-level inference speedup. Our work establishes a high-performance versatile layout-to-fab platform for advanced AI hardware.

Session 26: ODI | Emerging Photonic Devices

2:00 PM, Continental 7-9

Co-Chairs: Michael Cich, Google and Chao Shen, Fudan University

This session covers emerging materials and device structures for photonics applications. To start, three papers discuss image sensors that can be configured to operate in multiple wavelength bands. The first paper by Luo et al. from the University of Science and Technology of China (USTC) present a 5-μm-pixel filter-free UV-visible micro-spectrometer array that leverages vertically stacked GaN-based asymmetric dual PN junctions for ultra-compact multispectral imaging. The second paper by Hahn et al. from IMEC uses PbS colloidal quantum dots in a back to back diode configuration to operate in either visible or shortwave infrared. The third paper by Gao et al. from USTC reports a double-sided integrated optoelectronic co-encoding architecture utilizing a GaN/AlGaN device platform toward compact, high-density hyperspectral imagers. The fourth paper by Li et al. from Xiamen University discusses the epi-growth and fabrication of high-speed, low power consumption InGaN-based quantum dot micro-LEDs, providing a new transmitter technology for optical interconnects. Then, Zhang et al. from Hong Kong University of Science and Technology (HKUST) create a common GaN platform to address emission, reception, and nonvolatile synaptic storage for free-space photonically interconnected neuromorphic computing. Next, Guo et al. from Nanjing University cover the use of SiC for UV detection, taking advantage of the SiC miscut growth angle and an electric field profile to get spatial information. Finally, Liu et al. from Shenzhen Pinghu Laboratory demonstrate a photoconductive switch operating at 11kV using Mg-doped Ga2O3.

2:05 PM

26-1 A 5-μm-Pixel Filter-Free Micro-Spectrometer Array via Asymmetric Dual PN Junctions for Multispectral Imaging from Deep Ultraviolet to Visible Band

Yuanmin Luo, Zhixiang Gao, Huabin Yu, Wengang Gu, Wei Chen, Dongyang Luo, Haochen Zhang, Zixun Wang, Zhengming Yao, Yuchen Du, Linjun Wang, Yang Kang, Yong Yan, Haiding Sun

University of Science and Technology of China (USTC)

Miniaturized on-chip spectrometers with broadband detection range while having high spectral resolution remains a significant challenge. Traditional approaches by using external discrete optical components, such as narrow-band filters or dispersive metasurfaces, constrain device downscaling and require complex nano-lithography process. Herein, we present the first-of-its-kind 5-μm-pixel, filter-free micro-spectrometer array based on a stacked asymmetric dual PN junction (ADJ) photodiode (PD). By incorporating a composition-graded InGaN layer as the broadband light absorptive region, the ADJ-PD architecture eliminates the necessity for external bulky optical components. Fabricated via CMOS-compatible processes, this two-terminal diode realizes an electrically tunable, continuous absorption profile spanning from 250 nm (Deep UV) to 510 nm (Visible). The device achieves a 5-μm footprint while having a high peak resolution over ~2 nm. Leveraging the superior micro-spectrometer and its array,

we further demonstrate simultaneous spatio-spectral plastic classification and multi-spectral terrestrial remote sensing, providing a viable pathway for next-generation ultra-compact spectral imaging system.

2:30 PM

26-2 Reconfigurable Thin-Film Imager with 3T + 2PD Active Pixel Structure

Sangin Hahn[1,2], Donghyeok Jeong[1,3], Vladimir Pejovic[1], Minhyun Jin[1], Myonglae Chu[1], Seunghyup Yoo[4], Jung-Hoon Chun[3], Sang Yeon Lee[1], Gökhan Kara[1], Tristan Weydts[1], Pawel E. Malinowski[1], Arman Uz Zaman[1], Luis Moreno Hagelsieb[1], Yoora Seo[1,5], Jiwon Lee[1,6], David Cheyns[1], Itai Lieberman[1], Joo Hyoung Kim[1]
[1]imec, [2]School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), [3]Sungkyunkwan University, [4]Korea Advanced Institute of Science and Technology (KAIST), [5]Hanyang University, [6]Pohang University of Science and Technology

We demonstrate a reconfigurable image sensor by combining a thin-film photodetector with a conventional CMOS active pixel, forming a 3T + 2PD structure. A two-terminal PbS colloidal quantum dot (cQD) thin-film photodetector with a symmetric back-to-back photodiode configuration is monolithically integrated on a CMOS 3T active-pixel backplane, enabling bias-controlled modulation of photocurrent magnitude and polarity from the top device stack. As a result, the fabricated 512 × 768 image sensor with a 5 μm pixel pitch exhibits pixel-level output modulation under different operation bias conditions. Using the reconfigurable pixel output, we further demonstrate low-level image preprocessing based on image sets acquired under varying bias conditions. In addition, by selectively addressing each sub-diode, we implement a bias-switchable dual-band imager by combining spectrally distinct active layers in the device stack. This work establishes a practical design framework for a bias-programmable imager and provides a straightforward route toward advanced vision systems.

2:55 PM

26-3 The First Optical-Electrical Co-Encoded Photodiode Array Integrated with On-Chip Bragg-pair Filter for Miniaturized Hyperspectral Ultraviolet Imaging Sensor

Zhixiang Gao, Yuanmin Luo, Dongyang Luo, Zhengming Yao, Wei Chen, Haochen Zhang, Yuchen Du, Muhammad Hunain Memon, Huabin Yu, Haiding Sun
University of Science and Technology of China (USTC)

Miniaturized hyperspectral ultraviolet (UV) imagers are highly constrained by limited number and high correlation of spectral response channels. Here, we report the first optical-electrical co-encoded photodiode array integrated with on-chip Bragg-pair filter for miniaturized hyperspectral UV imager. The imager consists of a bias-tunable 8×8 photodiode array while integrating a 64-channel binary Bragg-pair filter on the backside of photodiode array. It possesses tunable and effective photoresponse that multiplicatively couples optical and electrical encoding process**es** at the same time. Compared with conventional optical-and electrical-only encoding schemes, the proposed co-encoding approach significantly reduces the inter-channel mean correlation by an order of magnitude from ~0.7 to ~0.06. A deep neural network is further implemented, with a demonstration of extremely high dual-peak spectral resolution of 1.9 nm from 245 nm to 365 nm. Finally, an integrated readout circuit and single-pixel imaging system demonstrate its superior wavelength-resolved UV imaging ability.

3:45 PM

26-4 First Demonstration of Thermally Robust Wafer-Scale AlGaN/InGaN QD Micro-LEDs with 1.3-GHz Bandwidth at 100 A/cm⊃2] for Optical Interconnects

Guanqi Li[1], Jianhui Liu[2], Jiansong Chen[2], Yanan Xu[2], Jinjian Yan[2], Chunyu Liu[2], Jinchai Li[2], Zhaoxia Bi[2], Hongjian Li[2,2], Panpan Li[2], Kai Huang[2], Rong Zhang[2]
[1]Department of Physics, Xiamen University, Xiamen University, [2]Xiamen University

For the first time, we demonstrate 4-inch wafer-scale AlGaN/InGaN quantum dot (QD) micro-light-emitting diodes (micro-LEDs) for low-current-density GHz modulation targeting short-reach optical interconnects. The AlGaN-capped InGaN QD active region integrates In-rich localized recombination centers with AlGaN-cap-enhanced carrier confinement, significantly reducing the effective carrier lifetime. This design enables: 1) 1.23-GHz -3 dB bandwidth at low *J* (100 A/cm^2), with 3.2× and 1.7× improvements over InGaN QW and QDs references; 2) robust thermal stability with >1 GHz bandwidth at 500 K (only 4.1% bandwidth loss); 3) 4-inch wafer-scale manufacturability statistics show reproducible GHz-class response, with 98% high-bandwidth yield (>1

GHz). Below-FEC NRZ-OOK transmission up to 2.2 Gb/s further verifies data-link capability. These results establish AlGaN/InGaN QDs as a scalable active-region platform for low-power micro-LED optical interconnects.

4:10 PM

26-5 Optoelectronic Neurons and Synapses on a GaN Platform for Free-Space Photonically Interconnected Neuromorphic Computing

Yuanke Zhang, Yichen Liu, Zheng Wu, Peixuan Bai, Tao Chen, Kevin Jing Chen
Hong Kong University of Science and Technology (HKUST)

Neuromorphic hardware with colocated processing and memory units addresses the latency and energy bottlenecks in conventional computing architecture. However, the interconnections among electronic neurons require dedicated electrical routing and peripheral circuits for signal readout/conversion/driving. Here we demonstrate optoelectronic neurons and synapses on the p-GaN/AlGaN/GaN platform for free-space photonically interconnected neuromorphic computing. The optoelectronic neuron is built on p-GaN/AlGaN/GaN p-i-n diodes gated by Al2O3/AlN/p-GaN charge-trapping stacks. Electrical modulation of interfacial charge trapping produces tunable optical emission and endows the device with short-term memory for temporal information processing. The optoelectronic synapse integrates GaN-based photodetectors with charge-trapping non-volatile memories, coupling optoelectronic conversion with non-volatile storage. The OES delivers multi-level states, long retention time of 6×108s, and endurance exceeding 1011cycles. Employing the OEN as a physical reservoir, we perform a chaotic time-series prediction with NRMSE of 0.26. These results demonstrate GaN-based optoelectronic devices as a promising solution for scalable, free-space photonically interconnected neuromorphic systems.

4:35 PM

26-6 A Symmetry-Broken Semiconductor Drift Chamber with 0.1 e^- Noise Charge Enabling Pixel-Free In-Sensor Spatial Computing and 25-ns Shot Imaging

Yue Guo, Yu Chen, Yao Li, Mengtao Han, Qixiang Guo, Weiqi Gao, Dong Zhou, Feng Zhou, Fangfang Ren, Jiandong Ye, Dunjun Chen, Rong Zhang, Youdou Zheng, Changjin Wan, Weizong Xu, Hai Lu
Nanjing University

This study proposes a semiconductor drift chamber that utilizes the structural anisotropy of an off-axis 4H-SiC epitaxial layer and combines it with a droplet-style concentric potential modulation scheme to establish a charge collection architecture that completely breaks symmetry. By vertically integrating the photon absorption region with the modulation electrodes and employing a dedicated system-in-package process, we achieved a record-low noise charge of 0.1 e^- and ultrafast charge output within 25 ns. Crucially, within an effective area of 3 mm^2, the device demonstrates a physically decodable mapping relationship between two-dimensional spatial positions and output waveform shapes, achieving >93% accuracy in 64-bit spatial reconstruction that includes grayscale intensity information, thereby providing a widely applicable and scalable implementation path for pixel-less single-chip camera platforms.

5:00 PM

26-7 Mg-Engineered β-Ga_2O_3 Photoconductive Switches with >11 kV Breakdown, ~10^{12} Off/On Ratio and Sub-ns Response under Sub-Bandgap Excitation

Chenghao Liu[1], Xiang Li[1], Duanxi Wu[2], Mengqiao Hu[2], Jiaxiang Chen[1], Teng Jiao[1], Maojin Yang[1], Yuxi Wan[1], Longfei Xiao[2], Daohua Zhang[1]
[1]Shenzhen Pinghu Laboratory, [2]Shandong University

We demonstrate a Mg-engineered β-Ga_2O_3 photoconductive semiconductor switches (PCSS) with simultaneously high breakdown voltage (>11 kV), ultrahigh off/on-state resistance ratio (~10^{12}), and sub-ns response under 355 nm sub-bandgap excitation. The device delivers a ~0.2/0.6 ns rise/fall time with a ~0.55 ns pulse width, and maintains strong GHz-range spectral content with ~8 dB attenuation at 1GHz. Sub-bandgap photoresponse and fast recovery dynamics, mediated by Mg-related defect engineering, highlight Mg:β-Ga_2O_3 PCSS as a promising solid-state switching platform for high-power microwave and ultrawideband applications.

Session 27: RSD | Component-to-System Level Reliability and Thermal Management

2:00 PM, Imperial A

Co-Chairs: Tianhao Zhang, Synopsys and Kenji Okada, Rapidus

In this session, reliability is addressed in spanning device degradation physics, passive-component reliability, advanced packaging, and system-level resilience. First, a cluster-based framework is presented to bridge single-transistor BTI degradation mechanisms to high-bandwidth-memory (HBM) reliability in multi-stacked memory systems. The second paper introduces a low-frequency-noise-based methodology for reliability modeling in aggressively scaled gate oxides and multi-link interconnect structures. Next, dielectric polarization phenomena are connected to AC-TDDB behavior through a physics-based reliability model for high-density 3D MIM capacitors. The fourth paper presents high-density on-chip MIM capacitors based on scaled DRAM cell structures to improve power integrity in 3D-stacked AI systems. Building on this theme, the fifth paper then introduces system-level thermal management techniques for enhancing capacitor reliability in heterogeneous 3D integrated power-delivery architectures. The sixth paper investigates 4D electron-beam tomography for failure-critical defect diagnosis and in-situ healing of operating power devices. Finally, the session concludes with a two-phase cooling solution for kW-scale computing chips, demonstrating highly efficient thermal management with remarkably low coolant flow requirements for next-generation high-performance computing systems.

2:05 PM

27-1 From Single-Transistor BTI to HBM Reliability: A Cluster-Based Device-to-System Reliability Framework for Multi-Stacked Memory Systems

Seonhaeng lee, Gang-jun kim, Donghee son, NAMHYUN lee, Giho yang, Myeongkyu eo, Jaeyeop ahn, Jaesang lee, Sunghwan kim, Inhoe kim, Minjun song, YouMi sim, JungEun seok, Gunhee bae, KyungWoo lee, Kyungwon park, Chae-Hyuk yun, Sungchul park, Kichul chun

Samsung Electronics

We demonstrate that ppm-level frequency degradation of multi-stacked HBM products can be quantitatively predicted directly from single-transistor BTI characteristics. Massive statistical analysis reveals intrinsic clustering in NBTI-induced ΔVth distributions caused by progressive consumption of weak Si–H bonds, whereas PBTI exhibits nearly uniform degradation due to rapid trap saturation. The identified clustering behavior propagates from transistor-level degradation to inverter-chain delay and ultimately to module-level frequency degradation. Conventional single-distribution extrapolation significantly overestimates tail degradation and falsely predicts product specification failure. By incorporating the proposed NBTI cluster model, measured HTOL results of both one-chip and 12-high HBM products are reproduced with approximately 95% accuracy without additional fitting parameters. This work establishes a practical device-to-system reliability framework and demonstrates that intrinsic NBTI clustering governs ppm-level degradation in advanced multi-stacked memory systems.

2:30 PM

27-2 Low-Frequency Noise-Based Reliability Modeling in Gate Oxide Area Scaling and Multi-Link Interconnects

Pei-Jean Liao, Kuo-Yu Hsiang, Hsin-Chun Chang, Yu-Kai Chang, Chen-Han Chou, Jen-Hao Lee, Ryan Lu, Yi-Huang Wu

Taiwan Semiconductor Manufacturing Company (TSMC)

We present a low-frequency noise (LFN)-based reliability framework addressing characterization challenges in scaled technologies. In FEOL, LFN detects the initial SBD masked by parasitic leakage in large-area devices, while its progressive trajectories reveal that HBD-triggering defect dynamics remain area-independent. In BEOL multi-link interconnects, the framework captures early void nucleation via a >100x PSD jump at a minimal <1.1% resistance shift. The extracted 1.46 eV activation energy physically validates the sensitivity of LFN in probing the incubation time to overcome the critical nucleation barrier.

2:55 PM

27-3 From Dielectric Polarization to AC-TDDB Modeling in High-Density 3D MIM Capacitors

Abygaël VIEY[1], Alexis KRAKOVINSKY[1], Bruno REIG[1], Julien DELPRATO[1], Marios BARLAS[2], Alexis DIVAY[1], David ROY[2], Xavier FEDERSPIEL[2], Xavier GARROS[1]

[1]CEA-Leti, [2]STMicroelectronics

Last year, we reported that relaxation currents in high-density 3D Metal-Insulator-Metal (MIM) capacitors integrated in the Back-End-of-Line originate from dielectric polarization and can be accurately modeled in both the time and frequency domains under low electric-field conditions (< 4V). In this work, we extend this polarization framework to high-field reliability by introducing a novel physics-based AC-Time-Dependent Dielectric Breakdown (AC-TDDB) model that directly links dielectric polarization to time-to-breakdown (t_{BD}). Using only three fitting parameters, the model accurately reproduces the experimental dependence of t_{BD} on AC stress frequency, Duty Factor, and high-level AC voltage. It further explains the discrepancy between the effective AC and DC TDDB acceleration factors (γ), as well as the experimentally observed decrease of γ with increasing frequency. The model also provides physical insight by suggesting that defect generation is governed by electric-field-assisted dielectric mechanical deformation (electrostriction), thereby establishing a theoretical bridge between dielectric polarization and AC-TDDB degradation.

3:20 PM

27-4 High-Density On-Chip MIM Capacitors Utilizing Scaled DRAM Cell Structures for Power Integrity in 3D-Stacked AI Chips

Gang-Jun Kim, Giho Yang, Donghee Son, Jae-Won Kim, Eunjeong Kim, Youngjun Kim, Seonhaeng Lee, Nam-Hyun Lee, Kichul Chun

Samsung Electronics

To enhance power integrity in 3D-stacked AI chips and HBM, this paper presents high-density on-chip MIM capacitors utilizing aggressively scaled DRAM cell structures, achieving a capacitance gain of over 200x compared to conventional MOS capacitors. The trade-off between capacitance and leakage was overcome through dielectric material engineering and interlayer band-offset modulation. By suppressing thermal emission, the leakage characteristics were governed by the temperature-independent Trap-Assisted Tunneling mechanism under bottom electrode bias. The proposed structure exhibited an enhanced breakdown electric field (>7 MV/cm) and exceptional robustness, with no significant leakage increase during 1,000-hour HTOL stress at 1.4VDD. Furthermore, a correlation between initial leakage and TDDB was established to derive a 10-year lifetime specification, which was successfully integrated into a mass-production system for continuous quality monitoring. This work offers a scalable solution for ensuring the reliability of next-generation high-performance AI systems.

4:10 PM

27-5 System-Level Thermal Management and Capacitor Reliability Enhancement in Heterogeneous 3D Integrated Power Delivery System

Minhyeok Lee[1], Chan Jik Lee[1], Bong Ho Kim[1], Donggeon Lee[1], Jongun Beom[1], Sunghwan Lee[1], Chae Yeon Han[1], Heetae Kim[1], Youngkeun Park[1], Byung Jin Cho[1], Jaeyong Jeong[2], Sanghyeon Kim[1]

[1]Korea Advanced Institute of Science and Technology (KAIST), [2]Ulsan National Institute of Science and Technology (UNIST)

Active backside power delivery networks (BSPDNs) are promising for high-power-consumption systems, but large area required for flying capacitors remains a major obstacle. To address this challenge, we demonstrated, for the first time, Heterogeneous 3D (H3D) integration of a CMOS chip, GaN power switches, and 3D pillar-shaped capacitors for active BSPDNs, with system- and device-level optimization. Thermoreflectance microscopy revealed that the capacitor temperature reached 160 ˚C under active operation, exposing a reliability challenge overlooked in previous on-chip converter studies. A heat-spreading layer reduced the capacitor-layer temperature to 127 ˚C and thermal resistance from 7.50 to 3.60 K·mm/W. Based on the identified system temperature, power-device and capacitor reliability was evaluated. Film and pillar-structure optimization achieved a capacitance density of 138 fF/μm^2 and a projected lifetime of 19.6 years at 3.3 V and 127 ˚C. This study establishes system- and device-level electrothermal co-design as essential for practical implementation of vertically integrated active BSPDNs.

4:35 PM

27-6 4D Electron-Beam Tomography for Failure-Critical Defect Diagnosis and In-situ Healing in Operating Power devices

Xinyi Pei[1], Lin Zhang[1], Yue Guo[1], Guangyang Gu[1], Kaixuan Ni[2], Weiqi Gao[1], Dawei Yan[3], Yi Yang[1], Fangfang Ren[1], Rong Zhang[1], Youdou Zheng[1], Weizong Xu[1], Hai Lu[1], Jiandong Ye[1]

[1]Nanjing University, [2]Nanjing University of Posts and Telecommunications, [3]Jiangnan University

Breakdown-critical defects fundamentally limit the reliability of power devices, yet a critical technological gap remains in noninvasive three-dimensional defect diagnosis and targeted mitigation. Here we report the first demonstration of a correlative four-dimensional (4D) noninvasive diagnosis-and-healing strategy based on graded-energy electron-beam-induced current (EBIC) tomography. By combining graded electron energies with time-resolved electrical response, electrically active but optically invisible defects in operating β-Ga_2O_3 power devices are reconstructed in 3D space and time domain. A 28-nm lateral spatial resolution is achieved, resolving two individual defect cores as near-surface dislocations with lateral dimensions of 148 and 220 nm, respectively. Guided by the defect reconstruction, a site-specific electron-beam healing process is *in-situ* performed, achieving 4.4% breakdown voltage (BV) improvement and stable leakage suppression over 120 h. This work establishes a new paradigm for predictive reliability engineering by enabling closed-loop defect analytics, 4D reconstruction, and *in-situ* targeted healing in operating ultrawide-bandgap (UWBG) power electronics.

5:00 PM

27-7 Two-phase cooling for kW-scale computing chips with a remarkably low liquid flow rate

Rong Fu, Wenzheng Jiang, Chuan Chen, Qidong Wang, Liqiang Cao

Institute of Microelectronics of Chinese Academy of Sciences

An efficient two-phase cooling strategy was developed using a membrane-based heat sink, achieving a remarkably low liquid flow rate of 0.024 L/min per kW of heat dissipation. The strategy adopted an industry-compatible, near-chip cooling configuration and demonstrated superior cooling performance of 1.3 kW per reticle while maintaining the junction temperature below 85 °C.

Session 28: PMA | Advanced GaN Power Devices: Power Delivery, Bidirectional and Low Voltage

2:00 PM, Imperial B

Co-Chairs: Julien Buckley, CEA-Leti and Tetsuzo Ueda, Kansai University

This session showcases work on advanced GaN power devices covering power delivery, driving and low voltage. The first paper by Wei Jia Zhang from Hong Kong University of Science and Technology presents power delivery for high performance computing covering both device and packaging aspects. The second paper by Junjie Yang from Peking University shows a complementary optically triggered 1200V lateral GaN switch. The third paper by Hehe Gong from The University of Hong Kong shows a low voltage p-GaN gate power HEMT for AI processor power delivery. The fourth paper by Yan Cheng from the Shenzhen Pinghu Laboratory presents a low voltage p-GaN gate power HEMT with regrown n+ GaN ohmic contacts for AI data center power delivery. The fifth paper by Hiroyuki Handa from Panasonic Holdings demonstrates GaN-on-GaN monolithic bidirectional switch (BDS) without the need for an additional bias control circuit. The sixth paper by Yuan Qin from University of Science and Technology of China proposes a novel NPN-cap region design that enables 10 kV Multi-Channel GaN Superjunction Monolithic BDS. The seventh paper by Junjie Yang from Peking University demonstrates GaN Monolithic BDS with area-efficient two-stage termination to boost breakdown voltage up to 10kV. The eighth paper by Jiawei Cui from Peking University demonstrates a common-gate GaN-on-Si monolithic BDS to eliminate back-gating effects.

2:05 PM

28-1 Power Delivery for High Performance Computing -- a Devices and Packaging Perspective

Wei Jia Zhang[1], Wai Tung Ng[2], Yuchu Ge[1], Jingyuan Liang[2], Wen Tao Cui[2], Jiupeng Zhang[2], Yipeng Guo[2], Xuan Wang[2], Jiahuizi Huang[2]

[1]Hong Kong University of Science and Technology (HKUST), [2]University of Toronto

Power delivery toward kW per chip has become one of the defining bottlenecks in high-performance computing (HPC). As the supply current escalates (> 1000's A) and voltage shrinks (< 1 V), power conversion efficiency and localized hot spots become the limiting factors in xPU/AI accelerator performance. This paper examines the roles

of power semiconductor devices and packaging have on the power delivery from the high voltage AC power line all the way to the xPU chips. Emerging device technologies such as wide-bandgap (WBG) power devices, advanced FET structures, and high-frequency power stages are reshaping the power delivery architectures. Co-design at the device, package, and system levels is now essential to meet next-generation performance targets. Trends such as on-package and on-die power delivery networks, 2.5D/3D integration, liquid cooling and the thermal-electrical-mechanical trade-offs that govern practical deployment are discussed.

2:30 PM

28-2 Complementary Optically Triggered 1200-V Lateral GaN Switch with R_{SP}-BV Performance Comparable to Electrically Driven Switches

Junjie Yang, Youyi Yin, Haoran Li, Yunhong Lao, Jingjing Yu, Sihang Liu, Han Yang, Xuelin Yang, Wei Wang, Maojun Wang, Bo Shen, Jin Wei
Peking University

Optically triggered power switches are promising solutions for constructing high-voltage conversion systems, as they provide galvanic isolation between low-voltage control signals and high-voltage signals. This work proposes a complementary optically triggered 1200-V lateral GaN power switch (COT-switch) to improve the R_{SP}-BV trade-off and switching performance. The COT-switch achieves several goals. (1) The COT-switch delivers an R_{SP} of 4.1 mΩ·cm^2 and a BV of 1945 V, achieving R_{SP}-BV trade-off comparable to that of electrically driven GaN power switches. (2) During the turn-on transient, the hole barrier collects photogenerated holes to accelerate gate charging process, yielding a turn-on time of only one-eighth that of the control switch without hole barrier. (3) During the turn-off transient, the PD rapidly discharges the gate and the buried hole barrier suppresses the persistent photocurrent. Thus, COT-switch shows a short turn-off time of 1.2 μs, representing best-in-class turn-off speed among high-voltage optically triggered GaN switches.

2:55 PM

28-3 Sub-1-mΩ·mm^2 Low-Voltage (15-V) p-GaN Gate HEMT on 200-mm GaN-on-Si for AI Processor Power Delivery Applications

Hehe Gong[1], Jiaqi He[2], Zineng Yang[1], Hengbo Zhang[1], Xin Yang[1], Han Wang[1], Kai Cheng[2], Yuhao Zhang[1]
[1]THE UNIVERSITY OF HONG KONG, [2]Enkris Semiconductor Inc.

We report 15-V p-GaN gate HEMTs on 200-mm GaN-on-Si with record low specific on-resistance below 1 mΩ·mm^2 and regrown contact resistance of 0.05 Ω·mm, achieving best figure-of-merit, DIBL and PBTI reliability among reported 15-V GaN power devices, showing great potential to improve efficiency in last-stage power delivery for AI processors.

3:20 PM

28-4 Densely Scaled Regrown n$^+$-GaN Ohmic Contacts on 200-mm GaN-on-Si Manufacturing Platform Enabling Ultra-Low $R_{ON,SP}$ and R_{ON}×Q_G in Low-Voltage p-GaN Gate Power HEMTs for Multi-MHz AI Datacenter Power Delivery

Yan Cheng[1,2], David Zhou[1], Yu Shi[1,3], Yu Zhang[1], Yuhao Wang[1], Junjie Zhu[2], Zidong Cai[1], Yuwei Wu[2], Haolin Hu[1], Qiang Gao[1], Zongjie Zhou[2], Longge Deng[2], Xiaoping Wang[1], Chenchang Zhan[3], Yuxi Wan[1], Kevin J. Chen[2,4]
[1]Shenzhen Pinghu Laboratory, [2]Hong Kong University of Science and Technology (HKUST), [3]Southern University of Science and Technology, [4]InnoHK Power Semiconductors and Applications Center (PowerSAC)

This work reports a low-voltage (LV) GaN power-switch manufacturing technology that removes the ohmic-contact floor of scaled E-mode p-GaN gate HEMTs through densely scaled, In-free MOCVD-regrown n$^+$-GaN contacts on a 200-mm GaN-on-Si platform. Selective-area growth into densely populated 0.6-μm trenches at a 1.45-μm cell pitch, despite severe topography of the periodic non-growth ridges, yields R_C of 0.075 Ω·mm with tight wafer-scale uniformity and nearly temperature-independent behavior up to 200 °C. The devices deliver ultralow $R_{ON,SP}$ down to 0.86 mΩ·mm^2 and R_{ON}×Q_G of 2.13 mΩ·nC at V_{GS} of 6 V with BV_{DS} of 35 V. TDGB projects a 10-year lifetime at V_{GS} of 7.42 V at 0.1% failure rate, with stable PBTI and HTRB behaviors at 150 °C. A synchronous buck converter using these devices delivers 12-V/6-V to 1.8-V/1-V conversion at up to 10 MHz with over 90% peak efficiency, validating the technology for next-generation AI-datacenter point-of-load and vertical power delivery.

4:10 PM

28-5 A Current-Collapse-Free GaN-on-GaN Monolithic Bidirectional Switch with Integrated Dual p-n Diodes for Passive Substrate Potential Tracking

Hiroyuki Handa, Masao Kawaguchi, Takashi Kano, Masahiro Ogawa, Naoki Torii, Naohiro Tsurumi, Keiichi Matsunaga, Daisuke Shibata, Yoshio Okayama, Satoshi Tamura
Panasonic

We demonstrate a GaN-on-GaN monolithic bidirectional switch integrating dual vertical p-n diodes for passive substrate potential tracking without an additional bias-control circuit. The diode connected to the higher-potential source passively clamps the conductive substrate within one forward diode drop, suppressing the substrate-potential transient that causes switch-node voltage tails and back-gating-induced dynamic R_{ON} degradation; the roles of the dual diodes interchange with voltage polarity. Extended source-connected p-GaN wells additionally shield the channel and relax gate-edge electric-field crowding. Shielded test structure withstood more than 600 V. In an inductive-load double-pulse switching test at 400 V/4 A with 50 V/ns, the measured substrate potential tracked the higher-potential source without a discernible switch-node voltage tail. The dynamic-to-static R_{ON} ratio remained near unity up to 400 V, demonstrating high-speed bidirectional switching with substrate potential control embedded directly in the power device.

4:35 PM

28-6 10 kV Multi-Channel GaN Superjunction Monolithic Bidirectional Switch

Yuan Qin[1], Ming Xiao[2], Zineng Yang[3], Kai Liu[4], Hengbo Zhang[3], Yuang Cheng[1], Shibing Long[1], Han Wang[3], Kai Cheng[4], Yuhao Zhang[3]
[1]University of Science and Technology of China (USTC), [2]Xidian University, [3]The University of Hong Kong, [4]Enkris Semiconductor Inc.

P-GaN-cap superjunctions (SJs) have increased the breakdown voltage (*BV*) of GaN HEMTs, but their implementation in monolithic bidirectional switches (MBDSs) is hindered by punch-through during full p-GaN depletion. We propose an NPN-cap design enabling SJ operation under either blocking polarity: one PN junction is forward-biased to extract holes, while the adjacent reverse-biased junction suppresses punch-through. Vertically integrating two N-type regions at the p-GaN-cap edges yields a compact hybrid vertical–lateral structure. Implemented in a normally-off dual-cascode multi-channel GaN MBDS, the design experimentally demonstrates, for the first time, a *BV* exceeding 10 kV, together with a low specific on-resistance of 32.8 mΩ·cm^2 and a record power figure-of-merit among GaN MBDSs. The device exhibits symmetric dynamic on-resistance (2.3× the static value) under ±3-kV switching and only a 9% shift after 3-kV, 1000-s HTRB stress, demonstrating its potential for medium-voltage AC power electronics.

5:00 PM

28-7 10-kV GaN Monolithic Bidirectional Switch with Superior Figure of Merit and Low Dynamic R_{ON}

Junjie Yang[1], Jingjing Yu[1], Haoran Li[1], Hao Chang[1], Teng Li[1], Sihang Liu[1], Youyi Yin[1], Yunhong Lao[1], Han Yang[1], Xiaosen Liu[2], Wei Wang[1], Maojun Wang[1], Bo Shen[1], Jin Wei[1]
[1]Peking University, [2]Tsinghua University

Ultrahigh-voltage GaN monolithic bidirectional switches (MBDS) suffer from electric field crowding and dynamic ON-resistance (R_{ON}) degradation. In this work, we propose an E-mode GaN MBDS with area-efficient two stage termination (2ST-MBDS) to boost breakdown voltage (*BV*), and we introduce synchronized electrical and ultraviolet-optical (UV-optical) driving scheme to suppress dynamic R_{ON} degradation. With an L_{GG} of 77 μm, the 2ST MBDS delivers symmetric *BV* over ±10 kV, yielding a power figure of merit (FOM) beyond 3.01 GW/cm^2. Both the *BV* and FOM represent record values among reported GaN MBDS. When turning on the device, a synchronized UV illumination is applied simultaneously with the gate driving signal. The UV illumination accelerates the release of trapped electrons. As a result, the 2ST MBDS exhibits a low dynamic R_{ON}/static R_{ON} ratio of 1.3 after ±3 kV bidirectional stress. The proposed 2ST MBDS with light-assisted de-trapping scheme offers a promising pathway towards ultrahigh-voltage GaN MBDS.

5:25 PM

28-8 Common-Gate GaN-on-Si Monolithic Bidirectional Switch: Pathway to Eliminate Back-Gating Effects

Jiawei Cui, Haoran Li, Sihang Liu, Hao Chang, Jialin Duan, Teng Li, Yunhong Lao, Chen Wang, Junjie Yang, Han Yang, Maojun Wang, Xuelin Yang, Bo Shen, Jin Wei

Peking University

In this work, we proposed a new approach: common-gate GaN-on-Si monolithic bidirectional switch (CG-MBDS). The CG-MBDS configuration boasts the benefits of both common-source monolithic bidirectional switch (CS-MBDS) and common-drain monolithic bidirectional switch (CD-MBDS). The common gate region of the CG-MBDS is shared by both operation directions for voltage blocking, resulting in low $R_{ON,SP}$; the presence of a source contact, which is electrically connected to substrate, fundamentally solves the back-gating effects.

Session 29: Evening Panel | Rewriting the Rules of Semiconductor Manufacturing: Why the Old Manufacturing Model May Not Be Enough

8:00 PM, Continental 4-5

Moderator: Dr. Ian J Cutress (More Than Moore)

Panelists: Benard Sell, Vice President and GM of Logic Technology Development, Intel; Kazunari Ishimaru CTO, Rapidus, Nicholas Kelez, CEO and Chief Technology Officer, xLight

The IEDM evening panel is held as an interactive forum where experts give their views on important industry topics and is designed to foster an open exchange of ideas, with audience participation encouraged. This year's panel will be moderated by Dr. Ian Cutress, CEO and chief analyst at More than Moore. It will focus on the theme, "*Rewriting the Rules of Semiconductor Manufacturing: Why the Old Manufacturing Model May Not Be Enough.*"

Topics:

- Speed, Scale, and the Cost of Faster Learning
- The Foundry Model Meets a More Specialized Future
- Transistors, Tools, and the Factory Model
- The Capital Reality Behind Manufacturing Disruption
- Low Volume, High Value, and the Rise of Specialized Manufacturing
- The 450mm Warning for Every "Obvious" Manufacturing Shift

Wednesday, December 16, 2026

Session 30: PMA | High-Performance RF FETs and Switches

9:00 AM, Continental 1-3

Co-Chairs: Masataka Higashiwaki, Osaka Metropolitan University and Nicholas Dellas, Infineon Technologies

This session collectively highlights major advances in next-generation RF and high-frequency semiconductor technologies across GaN, Ga_2O_3, diamond, and InGaAs platforms. Key themes include overcoming traditional performance bottlenecks through novel device architectures, heterogeneous integration, and materials engineering. Innovations include a GaN RF switch transistor with a p-type charge-control layer achieving record RF switch figures of merit, optically addressable non-volatile GaN RF switches for reconfigurable front ends, hybrid-bonded GaN-on-Si devices integrated with BiCMOS back-end technology, and ultra-low-noise 45-nm GaN HEMTs enabled by quasi-ballistic transport and regrown contacts. Beyond GaN, researchers demonstrated record-low cryogenic 1/f noise in InGaAs HEMTs for quantum computing, the first GaN/AlN p-channel RF transistors delivering >1 A/mm current and RF power amplification, the first practical diamond RF technology platform featuring an 11-W power MOSFET and 3.6-W MMIC with integrated passives, and Ga_2O_3-on-diamond RF MOSFETs achieving record thermal performance and 4.1 W/mm output power at X-band. Together, these works demonstrate how advanced materials integration, thermal management, contact engineering, and novel device concepts are pushing RF electronics toward higher power density, lower noise, improved efficiency, greater reconfigurability, and tighter integration for future 6G communications, radar, quantum computing, and high-performance wireless systems.

9:05 AM

30-1 GaN RF Switch Transistor with a P-Type Charge Control Layer

Keisuke Shinohara[1], Dean Regan[1], Logan Whitaker[1], Joel Wong[1], Eric Regan[1], Chris Hillman[1], Petra Rowell[1], Joshua Bergman[1], Fanming Liu[2], Rongming Chu[2], David Dai[3], Jr-Tai Chen[3], Vivek Mehrotra[1], Miguel Urteaga[1], Berinder Brar[1]
[1]Teledyne Scientific and Imaging, [2]University of Illinois Urbana-Champaign, [3]SweGaN

We report a new GaN-based RF switch transistor technology that sets record $R_{on}{\cdot}C_{off}$ and $F_c{\cdot}BV$ figures of merit, while simultaneously achieving a high knee current density, a large on/off ratio, and excellent C_{off} linearity. This unprecedented performance was achieved by introducing a P-type Charge Control Layer into an AlGaN/GaN HEMT structure (PCCL-HEMT). The vertical p-i-n diode integrated with the HEMT controls the conductivity of the two-dimensional electron gas (2DEG) channel. Below the pinch-off voltage, the entire 2DEG across the source-to-drain region is depleted, yielding minimal off-state capacitance with negligible bias dependence. The extremely low leakage current through the p-i-n diode enables high breakdown voltage even in highly scaled transistors for a low on-resistance. By selectively etching the SiC substrate beneath the active device area, we realized a "GaN-on-Air" device that further reduces C_{off}. In this work, we developed three PCCL-HEMT variants, each tailored to distinct application needs.

9:30 AM

30-2 Isolated Optically-Addressable Control of GaN Non-Volatile RF Switch for Reconfigurable RF FEMs

Yichen Liu, Yuanke Zhang, Longge Deng, Tao Chen, Yan Cheng, Yat Hon Ng, Zheng Wu, Peixuan Bai, Liuqing Gao, Kevin J. Chen
Hong Kong University of Science and Technology (HKUST)

An optically addressable isolated-control architecture is demonstrated, for a GaN non-volatile RF switch (NVS) enabled by bipolar charge trapping at the dielectric/p-GaN interface. A series-connected UV-activated GaN optically-controlled switch at the gate terminal enables selective electrical access for erase-to-OFF operation and subsequent isolation after writing, eliminating the need for a manually disconnecting floating scheme. Contactless program-to-ON is initially achieved by external UV-illumination. A subsequent proposed in-situ photon-assisted scheme, combining direct hole injection with local photon-assisted de-trapping, achieves complete programming in 500 µs. The device delivers an HRS/LRS resistance ratio over 10^8, cycle endurance over 10^6, and RF performance of 1.7-dB insertion loss and 14.7-dB isolation at 5 GHz that is well retained for 1000 s under isolated-gate conditions, together with an input-referred P_{1dB} of 30 dBm with no destructive self-switching. The optically-addressable scheme provides a practical path toward co-packaged isolated control of GaN non-volatile RF switches for reconfigurable RF front-ends.

9:55 AM

30-3 Hybrid-Bonding of GaN-on-Si MIS-HEMTs onto 300 mm BiCMOS BEOL with RF PA Device and Switch Demonstrations

Ismaël Charlet[1], Margot Faure[1], Christophe Dubarry[1], Erwan Morvan[1], Yveline Gobil[1], Tarek Daher[1], Arij Barboura[1], Fanny Morisot[1], Luca Lucci[1], Lucas Nyssens[1], Alexis Divay[1], Julien Delprato[1], Loïc Sanchez[1], Romane Blaise[1], Pascal Chevalier[2], Victor Milon[2], Daniel Gloria[2], Cédric Durand[2], Olivier Valorge[1]
[1]CEA-Leti, [2]STMicroelectronics

This work demonstrates for what we believe is the first time CMOS-compatible die-to-wafer (D2W) hybrid-bonding heterogeneous integration of GaN-on-Si high electron mobility transistors (HEMT), fabricated on a 200 mm wafer, onto coplanar-waveguides (CPW) implemented in a 300 mm BiCMOS back-end-of-line (BEOL). The integration relies on D2W hybrid-bonding, enabling a known-good-die approach and 5-µm-pitch interconnects compatible with sub-1 mm^2 chiplet integration. This demonstrator incorporates (1) a multifinger power amplifier (PA) device achieving ft/fmax values of 83/200 GHz, an output power density of 4.69 W/mm at 28 GHz & VDD=20 V. (2) RF switches achieving a RonCoff of 313 fs and a power-handling capability exceeding 20 dBm in a series configuration. These excellent results highlight the strong potential of hybrid-bonding for next-generation wireless front ends, offering a compact footprint, high integration density, high yield, and a cost-effective heterogeneous integration pathway.

10:20 AM

30-4 Record-Low-Noise 45-nm RF GaN HEMTs Enabled by Hybrid-MBE Regrown S/D Contacts and Quasi-Ballistic Transport

SuMin Choi[1], WanSoo Park[2], SeungWoo Son[2], InGeun Lee[2], DongHee Shin[3], YoungKyun Noh[3], SangJin Yoon[4], TaeGyeong Kim[4], TaeWoo Kim[5], KyoungHoon Yang[6], DaeHyun Kim[2]

[1]Electronics Engineering, Kyungpook National University, [2]Kyungpook National University, [3]IVWorks, [4]qsi, [5]Texas Tech University, [6]Korea Advanced Institute of Science and Technology (KAIST)

We present a 45-nm-gate-length InAlGaN/GaN HEMT exhibiting a record-low minimum noise figure below 0.2 dB at Ka-band, while maintaining an associated gain exceeding 10 dB. These outstanding figures of merit are enabled by three key technology enablers: (1) Hybrid-MBE-driven selective *n*+ GaN S/D regrowth with a poly-GaN-free surface on the as-grown active region, providing a manufacturable and scalable integration route with a defect- and dislocation-free regrown interface and an ultra-low contact resistance of 0.02 W×mm; a highly scaled 45-nm T-gate featuring low gate resistance and reduced Miller capacitance; and (3) *quasi-ballistic* carrier transport that suppresses intrinsic channel diffusion noise through reduced carrier scattering and velocity fluctuations. Ballistic transport analysis and physics-based noise modeling consistently explain the measured ultra-low-noise performance. This work establishes the lowest reported NF_{min} from X- to Ka-band among all reported transistor technologies, establishing RF GaN HEMTs as a leading platform for future 6G low-noise front-end systems.

11:10 AM

30-5 Record-Low 1/f Noise at 4 K in Cryogenic InGaAs HEMTs for Quantum Computing via Quantum Well Heterostructure Engineering

Jaeyong Jeong[1], Eunjung Cha[1], Sofia Tugnoli[1], Antonis Olziersky[1], Daniele Caimi[1], Marilyne Sousa[1], Steffen Reidt[1], Bernd Gotsmann[1], Sanghyeon Kim[2], Cezar Zota[1]

[1]IBM, [2]Korea Advanced Institute of Science and Technology (KAIST)

We report the first systematic study of low-frequency 1/f noise (LFN) at 4 K in cryogenic InGaAs HEMTs for quantum-computing amplifiers and interfaces. Cryogenic LFN limits readout signal-to-noise ratio and increases phase and frequency noise in control lines, thereby degrading qubit coherence and readout fidelity. Unlike most transistor metrics, LFN worsens upon cooling from 300 to 4 K. We observe a non-monotonic temperature dependence with a minimum near 200 K, revealing an additional disorder-related mechanism beyond conventional carrier-number fluctuation at cryogenic temperature. By comparing multiple quantum-well designs, we identify channel–barrier interface disorder as the dominant source of excess 4 K LFN. Quantum-well heterostructure engineering suppresses this noise by two orders of magnitude without compromising device performance. The optimized HEMTs achieve the lowest reported transistor LFN at 4 K over a broad operating bias range, with a noise indication factor of 0.14 and flat-band voltage noise below $1\times10^{-12}V^2/Hz$.

11:35 AM

30-6 GaN/AlN p-channel RF transistors with >1 A/mm and C-Band power amplification

Kazuki Nomoto, Zexuan Zhang, Jimy Encomendero, Lei Li, Reet Chaudhuri, Austin Hickman, Naomi Pieczulewski, James Hwang, David Muller, Grace Xing, Debdeep Jena

Cornell University

GaN p-channel field-effect transistors (pFETs) to date carry 10-100X lower current densities than their n-channel counterparts, and have never produced large-signal RF amplification. Here, we show that a highly conductive polarization-induced 2D hole gas (2DHG) at the GaN/AlN heterojunction overcomes these limitations. We report the first GaN/AlN pFETs on single-crystal AlN substrates that deliver room temperature (RT) on-currents exceeding 1 A/mm—matching levels of n-channel transistors—with a record-high speed of f_T/f_{max} = 25/45 GHz, and the first RF power amplification. At 77 K, I_{on} > 2 A/mm and its speed increases to f_T/f_{max} = 36/82 GHz. These results remove a major hurdle towards the realization of new applications in complementary RF electronics, power devices, and cryogenic computation using III-nitride wide-bandgap semiconductors.

12:00 PM

30-7 First Demonstration of an 11-W P_{out} Diamond RF Power MOSFET and a 3.6-W Diamond MMIC Power Amplifier with On-Chip Passives

Xinxin Yu[1,2], Bing Qiao[1,2], Yan Sun[3], Zhengshen Wang[3], Hehe Gong[4], Liu Wang[3], Ruize Wang[3], Shi He[1,2], Ran Tao[1,2], Huaixin Guo[3], Zhonghui Li[1,2], Yuhao Zhang[4], Jiandong Ye[5], Yuechan Kong[3], Tangsheng Chen[3]
[1]Nanjing Electronic Devices Institute, [2]State Key Laboratory of Wide Bandgap Semiconductor Devices and Integrated Technology, [3]National Key Laboratory of Solid-State Microwave Devices and Circuits, [4]The University of Hong Kong, [5]Nanjing University

Diamond RF power technology has long been hindered by current collapse, current scaling challenges in transistors, and the complete absence of on-chip passive components. In this work, we address these critical gaps by demonstrating an ampere-class multi-finger diamond RF power MOSFET with an air-bridge architecture, achieving a record RF output power (P_{out}) of 11.1 W, and robust operation up to 250 °C. Meanwhile, a comparison between large-signal P_{out} and DC power (P_{DC}) confirms minimal current collapse. In addition, we further demonstrate on-chip inductors and capacitors on diamond, as well as their operation capability up to 250 °C. These efforts lead to the demonstration of the first fully-integrated diamond MMIC power amplifier with on-chip matching, which delivers an output power of 3.6 W, representing the first watt-level RF MMIC achieved in ultra-wide bandgap semiconductors. These results provide a viable, industry-relevant pathway for ultra-high-power density and high-temperature RF front-ends.

12:25 PM

30-8 X-band 1.25 A/mm Ga_2O_3-on-diamond RF MOSFETs with Record Low R_T of 1.7 mm·K/W and P_{out} of 4.1 W/mm

Min Zhou[1], Mingjie Xiang[1], Hong Zhou[1], Xi Liu[1], Kun Zhang[1], E Zhou[2], Chao Yang[2], Zhaofu Zhang[2], Chao Yuan[2], Yitao Feng[1], Chenlu Wang[1], Chunfu Zhang[1], Zhihong Liu[1], Han Wang[3], Yuhao Zhang[3], Yue Hao[1], Jincheng Zhang[1]
[1]Xidian University, [2]Wuhan University, [3]THE UNIVERSITY OF HONG KONG

This work reports the first demonstration of heterogeneously integrated, interfacial-layer-free X-band Ga_2O_3-on-diamond RF transistors featuring record-low thermal resistance (R_T) and record-high output power density (P_{out}) among all Ga_2O_3 RF transistors. Utilizing a scalable, exfoliation-based layer-transfer direct bonding process, this integration method eliminates the interfacial amorphous layer typically required by other bonding techniques. This Ga_2O_3/diamond interface, combined with the high thermal conductivity of diamond substrate, drastically enhances heat dissipation, reducing R_T to 1.7 mm·K/W and effectively mitigating the self-heating effect. The fabricated transistors achieve a record top-gate drain current density of 1.25 A/mm and high on/off ratio of 2×10^{11} in Ga_2O_3 devices. Consequently, the devices deliver a P_{out} of 4.1 W/mm and a power-added efficiency (PAE) of 25.8% at 10 GHz. This P_{out} is the highest reported for Ga_2O_3 transistor**s** to date. These results demonstrate a highly promising path for scalable, heterogeneously integrated Ga_2O_3-on-diamond platforms in high-frequency and high-power RF applications.

Session 31: NC | Edge Intelligence and On-Chip Learning

9:00 AM, Continental 4

Co-Chairs: Erika Covi, Technical University of Munich and Stephan Menzel, FZ Julich

This session presents recent advances in edge intelligence and on-chip learning using ReRAM, SOT MRAM, and gain cells. The session covers seven papers. CEA-LETI demonstrates few-update on-chip learning exploiting ReRAM write-noise. Tsinghua University shows supervised forward-only learning on a 2.25 Mb 28 nm ReRAM chip exploiting locality and sparsity. Researchers from the University of Hong Kong demonstrate 28-nm 2T1R RRAM-CIM macro that achieves accurate and efficient forward-transpose matrix-vector multiplication for neural-network training. Researchers from University Paris-Saclay and University of Texas report gradient-guided learning using 28 nm ReRAM fully on-chip. Beihang University demonstrate a four-die 3D-Stacked 8-Mb SOT-MRAM compute-in-memory system for real-time edge Intelligence. POSTECH demonstrates a 2k synaptic array of split-gate write-transistor gain cells with a sub-100 nm gate gap for ultra-fast analog AI training. Peking University reports a hyperspectral image fusion system based on a 40nm 1Mb RRAM CIM chip.

9:05 AM

31-1 Exploiting ReRAM Write-Noise in Gradient Perturbation for Few-Update On-Chip Learning

Matilde Piccoli[1], Adrien Renaudineau[1], Mohammed Akib Iftakher[1], Kamel-Eddine Harabi[1], Mario Pallo[2], Simone D'Agostino[2], Jean-Michel Portal[3], Damien Querlioz[1], Elisa Vianello[2], Louis Hutin[2]
[1]Université Paris-Saclay, [2]CEA-Leti, [3]Aix-Marseille University

On-chip learning is a key requirement for adaptive edge systems, but remains difficult to reconcile with analog ReRAM IMC, where backpropagation requires nonlocal error propagation, activation storage, and precise control of variable devices. We demonstrate, to our knowledge for the first time, write-noise perturbation-based on-chip learning in ReRAM hardware, exploiting device programming variability as the perturbation source for local gradient estimation. Starting from a meta-trained initialization, gradient perturbation enables few-update adaptation using only noisy conductance sampling, forward MAC operations, and a scalar loss signal. Learning is implemented on ReRAM crossbars, reaching the 95% target accuracy on the Omniglot dataset within few updates per layer. We further introduce a program-and-verify-inspired sampling strategy that improves the accuracy/write-cost trade-off while preserving the benefit of intrinsic write noise. These results show that ReRAM programming variability can be used as a functional resource for on-chip learning rather than treated only as a device limitation.

9:30 AM

31-2 Dynamics-Enabled Brain-Inspired On-Chip Learning with Locality and Sparsity on a 28-nm RRAM Synaptic Chip

Jingwen Cai[1], Qi Qin[1], Peng Yao[1], Ze Wang[1], Sheng Hu[1], Keer Liu[1], Mengyun Zhang[1], Xingjun Wu[1], Zhenqi Hao[1], Qingtian Zhang[1], Jianshi Tang[1], He Qian[1], Haixia Yu[2], Chao Yue[2], Fujun Bai[3], Xiping Jiang[3], Bin Gao[1], Huaqiang Wu[1]
[1]Tsinghua University, [2]Tongxin Microelectronics Co., Ltd., [3]Xi'an UniIC Semiconductors Co. Ltd.

Intrinsic RRAM nonidealities constrain the use of conventional backpropagation (BP) for efficient on-chip learning. Drawing on the biological principles of local learning, time-dependent plasticity, and sparse synaptic updates, we present a dynamics-enabled brain-inspired learning framework that accommodates RRAM device characteristics. It incorporates supervised forward-only learning (SFO) to address device nonidealities of state variations and nonlinear updates, relaxation-based post-update plasticity for temporal adaptive learning, and hybrid-granularity sparse update (HGSU) to reduce programming count. We experimentally validate the learning framework on a 28-nm RRAM synaptic chip with a 1152×2048 array. This chip supports precise on-chip forward and backward propagation for learning. Moreover, in a traversability-aware autonomous-driving adaptation task, the fabricated chip achieves 95.30±0.18% accuracy, accelerates learning by 27.7× compared to conventional schemes, and lowers total learning energy by 52.3× compared with the evaluated digital baseline.

9:55 AM

31-3 First Demonstration of 28-nm Symmetric 2T1R RRAM Macro Suppressing Forward-Transpose Mismatch for AI Training and Iterative Solving

Chenchen Ding[1], Haobo Liu[2], Yuancan Li[2], He Xiao[2], Zhicheng Xu[2], Peiyi He[2], Zhengke Yang[3], Haiqiao Hong[2], Taiqiang Wu[2], Wenyong Zhou[2], Zhibin Feng[2], Can Li[2], Longyang Lin[3], Jianshi Tang[4], Bin Gao[4], Zhengwu Liu[2], Hao Yu[3], Wong Ngai[2], Huaqiang Wu[4]
[1]ECE, THE UNIVERSITY OF HONG KONG, [2]THE UNIVERSITY OF HONG KONG, [3]Southern University of Science and Technology, [4]Tsinghua University

We demonstrate, for the first time, STEP, a 28-nm 2T1R RRAM-CIM macro that achieves accurate and efficient forward-transpose matrix-vector multiplication. Symmetric- Transpose current paths make forward and transpose readout identical, achieving a 64.76× lower cell-level mismatch than conventional reverse readout and suppressing errors from asymmetric RRAM barriers and transistor mismatch. Efficient shared periphery lets multiple arrays share a single set of peripheral circuits rather than replicate one per array, avoiding peripheral duplication for concurrent multiplication. Flexible Parallel programming addresses arbitrary cells via independent forward/reverse lines, enabling efficient sparse and irregular weight updates. Validated on an implicit neural representation training task and a MIMO solving task, STEP consistently outperforms conventional RRAM-CIM baselines across three dimensions: it improves computational accuracy by 1.55 dB (INR) and 5.71 dB (MIMO), reduces energy consumption by 19.9% (INR) and 3.21× (MIMO), and cuts programming rounds by 58.3% (INR).

10:45 AM

31-4 Experimental Metropolis-Adjusted Langevin Learning with Fully On-Chip ReRAM-to-ReRAM Sampling

Benjamin Wallace Walker[1], Mohammed Akib Iftakher[1], Adrien Renaudineau[1], Kamel-Eddine Harabi[1], Mathieu Coumba Faye[2], Jean-Michel Portal[3], Thomas Dalgaty[2], Joseph S. Friedman[4], Elisa Vianello[2], Damien Querlioz[1]
[1]Université Paris-Saclay, [2]CEA-Leti, [3]Aix-Marseille University, [4]University of Texas at Dallas

Near-sensor intelligence must adapt to gradual shifts in input data and quantify its uncertainty to maintain robustness post-deployment. Bayesian learning implements both features in a principled manner, but is challenging to realize in ReRAM: existing methods either scale poorly or necessitate repeated precise conductance updates. Here we demonstrate, for the first time, experimental Metropolis-adjusted Langevin algorithm (MALA) learning with ReRAM. Our approach combines gradient information with the randomness of single-shot programming, enabling uncertainty-aware learning. The on-chip circuit allows efficient conditional ReRAM-to-ReRAM sampling without digitization or computer-in-the-loop transfer; a source-cell current and signed gradient current jointly set the target cell's programming distribution. The fabricated 2T1R system achieves 96.8% ensemble accuracy on four-class bird recognition and 0.94 AUROC for non-bird OOD detection. It requires 6.2× fewer programming events than the naïve random-walk Metropolis-Hastings approach, and maintains ensemble performance for 24 h. These results establish a variability-native route to adaptive, uncertainty-aware near-sensor learning.

11:10 AM

31-5 First RISC-V-Controlled Four-Die 3D-Stacked 8-Mb SOT-MRAM Compute-in-Memory System for Real-Time Edge Intelligence

Biao Pan[1], Shouzhong Peng[1], Bohan Xu[1], Zhaochun Liu[1], Haolin Ge[1], Xueyan Wang[1], Jin Luo[1], Hanting Chen[1], Xuan Li[2], Hui Jin[2], Kaihua Cao[2], Zongyang Li[3], Pengrong Lin[3], Hongxi Liu[2], He Zhang[1], Weisheng Zhao[1]
[1]Beihang University, [2]Truth Memory Corporation, [3]Beijing Microelectronics Technology Institute

We demonstrate the first RISC-V-controlled fourdie 3D-stacked 8-Mb spin-orbit torque magnetic randomaccess memory (SOT-MRAM) compute-in-memory (CIM) system. A comprehensive system-technology co-optimization (STCO) framework co-designs field-free Type-y SOT magnetic tunnel junctions (MTJs), 2-Mb CIM circuits, and waferlevel through-silicon-via/redistribution-layer (TSV/RDL) integration. The devices achieve 0.2 ns ultrafast switching and 10^{12}-cycle endurance. Four 2-Mb SOT-MRAM dies and a 40-nm RISC-V controller enable concurrent operation at 2.306 TOPS, 7.19 TOPS/W, and 48.7 GOPS/mm^2. A three-node demonstrator sustains 60-fps inference at 98% accuracy. This demonstrates that HBM-inspired STCO-driven 3D integration enables mature-process CIM to deliver real-time edge intelligence at a level rivaling advanced-node platforms.

11:35 AM

31-6 First Demonstration of a Split-Gate Write-Transistor Gain Cell with a Sub-100 nm Gate Gap and Its 2K Synaptic Array for Ultrafast Analog AI Training

Junyoung Choi[1], Eungcheol Kim[2], Bongjun Ju[2], Ga-eun Heo[2], Seungkun Kim[2], Byoungwoo Lee[2], Woochan Song[2], Jeonghoon Son[2], Beomjun Kim[2], Seyoung Kim[2,2]
[1]Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), [2]Pohang University of Science and Technology (POSTECH)

For the first time, we demonstrate a BEOL-compatible 2K (64×32) synaptic gain-cell array with intrinsic pulse-coincidence selection for analog AI training. Each cell uses a single split-gate write transistor (SGWT) whose two gates connect to row and column lines, enabling an intrinsic AND update only under coincident pulses for stochastic pulse-based training. By eliminating the additional write transistor and the floating intermediate node of 3T0C cells, the architecture reduces storage-node (SN) parasitic capacitance by up to 13.4×. The SGWT was scaled to a sub-100 nm gate gap, the critical dimension governing write current, and demonstrates sub-10 ns switching at ≤0.5 V, update selectivity exceeding 2×10^3, excellent symmetry, and endurance over 1.6×10^6 pulses. The fabricated 64×32 cross-point array demonstrates large-scale integration, while a monolithic 3D (M3D)-integrated oxide-transistor chip validates BEOL integration on Si-CMOS. Hardware-aware simulations using measured switching achieve 98.94% hand-gesture and 90.39% traffic-sign classification accuracy, supporting scalable analog AI training.

12:00 PM

31-7 A 40nm 1Mb RRAM CIM-based Hyperspectral Image Fusion System via Reinforcement Learning-based Programming Scheme and In-situ FP PCA/IPCA

Kexun Li[1], Hao Ai[1], Yading Yi[1], Ao Shi[1], Lianliang Wu[1], Nan Tang[1], Jiaqi Li[1], Yingchen Ji[1], Jinshan Li[1], Zongwei Wang[1], Yulin Feng[2], Lifeng Liu[1], Yimao Cai[1], Peng Huang[1]
[1]Peking University, [2]Beijing Information Science and Technology University

Hyperspectral image (HSI) fusion combines complementary modalities to enhance spatial–spectral information and is well suited for RRAM compute-in-memory (CIM) implementation owing to its high parallelism and energy efficiency. However, practical implementation is hindered by unreliable RRAM multilevel storage and inefficient support for core fusion operations. We demonstrate the first RRAM CIM-based HSI-fusion system using a 40-nm 1-Mb 1T1R chip. A reinforcement-learning-based programming scheme (RiLPS), combining a device-variation-robust policy with physics-guided actions, reduces retention-error variation by 47% versus incremental step pulse programming (ISPP). The system also supports in-situ floating-point principal component analysis (PCA) and inverse PCA through a unified two-dimensional calibration model. On practical datasets, it achieves 36.7-dB PSNR, 3.65 SAM for fusion quality, and 0.75 IoU for downstream tasks, all comparable to software results. Compared with GPU baselines, the system delivers 222× higher energy efficiency and a 3.7× speedup, providing a viable solution for edge HSI processing.

Session 32: EDT | Ferroelectric Devices

9:00 AM, Continental 5

Co-Chairs: Asif Khan, Georgia Institute of Technology and Suraj Cheema, MIT

This session traces ferroelectric memory from the HfO_2-ZrO_2 stack itself, through channel materials scaled from single-crystal silicon to oxide semiconductors to monolayer MoS_2, and onward to array-level NAND operation and reconfigurable compute- and search-in-memory hardware. The first paper in the session, from National Taiwan University, shows record leakage-EOT results on a TiO_2-intercalated HfO_2/ZrO_2 superlattice capacitor, attributing its record capacitance density to enhanced reversible dielectric susceptibility rather than remanent switching. The second paper, from the Institute of Microelectronics of the Chinese Academy of Sciences, shows compelling results on the first 3D ferroelectric NOR built on an ultra-thin vertical single-crystal silicon channel, overcoming the trap-limited transport and leakage of conventional poly-Si architectures. The third paper, from National Taiwan University, shows innovative results on an oxide-semiconductor nanosheet ferroelectric FET in which a bottom polarization control layer deliberately disables switching in the lower ferroelectric, resolving the long-standing erase deficiency of wide-bandgap channels. The fourth paper, from Samsung Advanced Institute of Technology and Yonsei University, shows pioneering results on the first ferroelectric NAND with a two-dimensional monolayer channel, establishing 2D materials as a path to extreme 3D scaling while identifying the direct-growth thermal budget as the key integration challenge. The fifth paper, from Samsung Advanced Institute of Technology and Samsung Semiconductor R&D Center, shows decisive results on array-level low-V_{PASS} ferroelectric NAND, revealing that conventional block erase is fundamentally infeasible in oxide-semiconductor channels and establishing page-level erase and low-voltage program inhibition as viable replacements. The sixth paper, from the National University of Singapore, A*STAR, and the University of Notre Dame, shows remarkable results on a dual-ferroelectric-gate FeFET that can be electrically reconfigured between volatile and nonvolatile modes, unifying reservoir computing and vector-matrix multiplication within a single ultra-low-voltage array. The seventh paper, from Seoul National University, Samsung Advanced Institute of Technology, and Sogang University, shows elegant results on a one-transistor multi-bit ternary CAM using ambipolar tunnel FETs with ferroelectric-assisted charge trapping, enabling tunable approximate search in the most compact CAM cell reported to date.

9:05 AM

32-1 Ultra-Low EOT = 0.27 nm and J_{leak} = 1 μA/cm^2 of Physical Thickness 4 nm Superlattice HfO_2/ZrO_2 with TiO_2 Intercalation for BEOL eDRAM

Jia-Yang Lee, Kuan-Ting Wu, Mochamad Januar, Kun-Yuan You, Chia-Chun Hung, Min-Hung Lee
National Taiwan University

Proposed HZH/T structure achieves remarkable equivalent oxide thickness (EOT) of 0.27 nm and a leakage current density (J_{leak}) of 1 μA/cm^2, with a physical thickness of 4 nm and a back-end-of-line (BEOL)-compatible post-

metallization annealing (PMA) at 450°C. The superlattice (SL) configuration of HfO_2/ZrO_2 with TiO_2 intercalation offers promising avenues for enhancing permittivity. This enhancement is attributed to the reversible dielectric response facilitated by interfacial susceptibility enhancement with spontaneous polarization, as validated by the unified JA-LGD-TDGL framework. The intercalation of TiO_2 effectively eliminates interface localized oxygen vacancies (Vo) to improve J_{leak} and reliability. Experimentally, reliability is demonstrated with an EOT < 0.29 nm for 10^{12} endurance cycles, and a 10-year lifetime is projected at an electric field of 2.6 MV/cm for time-dependent dielectric breakdown (TDDB). The proposed HZH/T structure meets all power, performance, and area (PPA) criteria for DRAM applications and holds potential to overcome the memory wall in Tera-generation technologies.

9:30 AM

32-2 First Demonstration of 3D FeNOR with a 5-nm Ultra-Thin Single-Crystal Silicon Channel

Ruoyao Ji, Junshuai Chai, Yong Du, Runhao Han, Zeqi Chen, Xiaoyu Ke, Xiaoqing Sun, Bao Zhang, Xin Gong, Hao Xu, Yongkui Zhang, Xiaolei Wang, Qing Luo, Wenwu Wang, Tianchun Ye
Institute of Microelectronics of Chinese Academy of Sciences

In this work, we report the first demonstration of a 3D ferroelectric NOR (FeNOR) featuring a 5nm ultra-thin single-crystal silicon channel. The fully depleted single-crystal silicon channel combined with an engineered air-gap structure enables excellent electrostatic control, achieving an Ion/Ioff ratio of ~10^7 and an ultralow off-state current of ~10^{-13}A. By integrating a 10-nm $Hf_{0.5}Zr_{0.5}O_2$ ferroelectric gate dielectric, the proposed device further enables low-voltage and high-speed program/erase operation. A memory window (MW) of ~0.6V is achieved under ±5V and 50ns PGM/ERS pulses, providing sufficient margin for reliable SLC NOR operation. Furthermore, the device demonstrates a fast read latency of 10ns with excellent retention and read-disturb immunity. This work establishes an effective pathway toward scalable, highperformance 3D NOR flash based on ultra-thin single-crystal silicon channels.

9:55 AM

32-3 Novel In_2O_3/SiO_2 Nanosheets FeFET with Polarization Control Layer Featuring 2.2 V memory window and > 1E10 Endurance

Yuan-Ming Liu, Yu-Shan Wu, Rachit Dobhal, Hsien-Ming Sung, Ching-Ju Tseng, Fan-Hao Shen, Ying-Jung Chen, Yi-Shiang Tsao, C. W. Liu
National Taiwan University

In_2O_3 nanosheets with the thick bottom oxide wrapped around by HZO/TiN to disable the polarization switch in the bottom HZO achieve 2.2 V memory window. The polarization for memory window is solely controlled by top o-phase HZO and is balanced by whole channel charge of electrons and donors for polarization down and up, respectively. The hydrogen in bottom oxide can also passivate the weakly bonded oxygen in In_2O_3 to improve the low threshold voltage for program of 50 mV at 3 V overdrive voltage for 1E4 s and stress time/ΔV_{TL} of 200 s/mV. The program/erase pulse time of 100 ns/500 ns is measured with the endurance exceeding 1E10 cycles, and the retention of 2E3 s at 85 °C. TiN as an etching stop layer and metallic sacrificial layer protects the S/D during the via opening and enhances S/D doping. The process thermal budget is below 400 °C with backend compatibility.

10:45 AM

32-4 First Demonstration of 2D-Channel Ferroelectric NAND Achieving >11 V Memory Window for Extreme 3D Scaling

Kyunghun Kim[1], Minsu Seol[1], Sanghyun Jo[1], Yoonsang Park[1], Seokhoon Choi[1], Kihong Kim[1], Young-Chul Leem[1], Yunseong Lee[1], Jung-Kyun Kim[1], Taek Jung Kim[1], Kyung Mee Song[1], Donghoon Kim[1], Junyoung Kwon[1], Min Seok Yoo[1], Hyunmi Lee[1], Sejin Kim[2], Joohoon Kang[2], Jaewoo Lee[1], Changhoon Jung[1], Duk-Hyun Choe[1], Kyung-Eun Byun[1], Seung-Yeul Yang[1], Yongsung Kim[1], Changhyun Kim[1], Sijung Yoo[1]
[1]Samsung Advanced Institute of Technology (SAIT), [2]Yonsei University

We report the first successful demonstration of two-dimensional (2D) MoS_2-channel ferroelectric NAND (FeNAND), verifying its feasibility for advanced NAND scaling. Our FeNAND with single crystal monolayer MoS_2 achieves a memory window (MW) exceeding 11 V, steep subthreshold swing (SS), and low-power, disturb-free operation owing to its unipolar accumulation mode. Furthermore, we evaluate MoS_2 directly grown on 200-mm wafers, successfully securing an average MW > 10 V with excellent uniformity and systematically address the process-induced challenges of direct growth methodologies. This work presents the first experimental evidence of FeNAND operation with 2D channel materials, while highlighting key reliability considerations for 2D channel integration.

11:10 AM

32-5 Overcoming Array-Level Challenges for Low-Power Oxide-Channel Ferroelectric NAND Flash

Sijung Yoo[1], Hyoseok Kim[2], Junyeong Lim[2], Seogwoo Hong[1], Seokhoon Choi[1], Seung Dam Hyun[1], Jung-Kyun Kim[1], Taek Jung Kim[1], Sanghyun Jo[1], Donghoon Kim[1], Yoonsang Park[1], Kyung Mee Song[1], Kihong Kim[1], Kwang-Hee Lee[1], Sungduk Hong[2], Sung Jin Kim[2], Suhwan Lim[2], Kwangsoo Kim[2], Juhyung Kim[2], Wanki Kim[2], Jaehoon Jeong[2], Sangwook Kim[1], Jinseong Heo[1], Yongsung Kim[1], Duk-Hyun Choe[1]

[1]Samsung Advanced Institute of Technology (SAIT), [2]Samsung Electronics

Low-VPASS ferroelectric NAND (LVP FeNAND), enabled by ferroelectrics combined with oxide semiconductor channels, has recently been proposed as an ultralow-power NAND flash architecture. Here, we extend this concept to array-level operation and examine its practical implications. We show that, despite its substantial power advantages of low-VPASS operation, array-scale implementation introduces challenges in block erase and program inhibition. Through systematic analysis, we identify underlying limitations and demonstrate viable solutions, establishing a pathway to practical implementation of LVP FeNAND architecture.

11:35 AM

32-6 First Hardware Demonstration of a 3D Stackable and Reconfigurable Full Synaptic Dual-Ferroelectric-Gate FeFET Array featuring Record-Low 1.5 V Operating Voltage for Energy-Efficient Computing

Ning Ji[1], Bingjie Dang[1], Chaoming Wu[1], Yi Tong[1], Xiaopeng Li[1], Leming Jiao[1], Xiaolin Wang[1], Duy Hieu Trinh[1], Xinming Li[1], Dong Zhang[1], Kai Ni[2], Jingsheng Chen[1], Xiao Gong[1]

[1]National University of Singapore, [2]University of Notre Dame

For the first time, we propose and experimentally demonstrate an ultra-low-power reconfigurable dual-ferroelectric-gate field-effect transistor (R-Fe^{2}FET). The key contributions are threefold. First, by optimizing the coercive field (E_c) of the HZO ferroelectric capacitor and introducing dual-ferroelectric-gate voltage coupling, a record-low FeFET operating voltage of 0.35 V is achieved in a device with an ultra-scaled channel length (L_{CH}) of 35 nm, enabling sub-fJ nonvolatile write energy. Second, the proposed R-Fe^{2}FET can be electrically reconfigured through the mode-gate bias, allowing the same device to operate either as a volatile-memory element for reservoir computing or as a nonvolatile multilevel-memory element for vector-matrix multiplication. Third, based on this device concept, we experimentally demonstrate a reconfigurable FeFET array with a record-low array operating voltage of 1.5 V. The fabricated array successfully performs speech recognition with ~89.4% ten-class accuracy using 5-bit weights, while consuming 125 fJ/spike for volatile reservoir operation and 0.25 fJ/spike for nonvolatile in-memory.

12:00 PM

32-7 A Compact 1T Multi-Bit Ternary CAM Using Ambipolar Tunnel FETs with FE-Assisted Charge-Trap Flash for Tunable Approximate Search

Minjeong Ryu[1], Yoonsang Park[2], Jae Seung Woo[1], Joo Hyeon Jeon[3], Sung In Cho[3], Duk-Hyun Choe[2], Sijung Yoo[2], Woo Young Choi[1]

[1]Seoul National University, [2]Samsung Advanced Institute of Technology (SAIT), [3]Sogang University

A novel ultra-compact one-transistor (1T) multi-bit ternary content-addressable memory (TCAM) based on the ambipolar tunnel FET (TFET) with ferroelectric (FE)-assisted charge-trap-flash (CTF) operation is proposed and experimentally demonstrated. The proposed 1T multi-bit TCAM cell offers two key advantages: localized charge trapping enables the don't-care state "X" by selectively modulating the threshold voltage (V_{th}) for either p- or n-type conduction, while FE-assisted CTF enables balanced program/erase (P/E)-driven tunability. The fabricated 1T cells encode 7 distinguishable states through balanced P/E operation, with 0.5-V interlevel V_{th} windows and > 10-year retention. Leveraging this capability, principal component analysis (PCA)-based hashing enables bit-significance-aware X-bit masking for tunable approximate search. Array-level 5-way 5-shot associative retrieval validates that tunable match tolerance improves k-nearest-neighbor (kNN) classification accuracy from 93.85 % to 98.46 %. Moreover, cluster-prototype row compression reduces the effective task-level footprint to 2.4 F^2/bit, demonstrating the potential of the proposed design for compact and adaptive in-memory search.

Session 33: ALT | Advanced Device and Process Innovations

9:00 AM, Continental 6

Co-Chairs: Yusuke Oniki, Rapidus and Wookhyun Kwon, Samsung

As the semiconductor industry enters the angstrom era, sustaining performance, power, area, and cost improvements increasingly depends on innovations that extend beyond conventional device scaling. New transistor architectures, alternative channel materials, advanced gate-stack engineering, and novel process integration schemes are becoming essential to enable future generations of logic, memory, and power delivery technologies. This session, Advanced Device and Process Innovation, brings together a distinguished set of contributions that address these challenges from multiple perspectives, highlighting breakthroughs that will shape the evolution of advanced CMOS platforms. A major theme of the session is the continued advancement of Complementary FET (CFET) technology and the exploration of germanium-based channels for enhanced device performance. Intel presents a germanium nanoribbon PMOS technology featuring a reliable and aggressively scaled gate stack aimed at low-voltage CMOS applications. National Cheng Kung University reports the first demonstration of hetero-oriented n-Ge(111)/p-Ge(110) Ge CFETs, providing a compelling pathway toward highly scaled SRAM implementations. Complementing these material and device innovations, imec delivers insights into the scaling behavior and performance trade-offs of both single-level and three-dimensional stacked nanosheet FET architectures, offering an important perspective on future CFET integration strategies. Another key focus of the session is gate-stack and threshold-voltage engineering, which remain critical challenges for deeply scaled CMOS technologies. Hanyang University introduces a silicate-mediated volumeless dipole reversal technique that enables work-function-metal-free multi-threshold-voltage integration for CFET platforms. AIST presents a novel volumeless threshold-voltage control approach based on selective oxidation of work-function metals in gate-all-around CMOS devices. Together, these innovations point to new directions for achieving flexible and manufacturable threshold-voltage tuning while meeting increasingly stringent scaling requirements. The session also highlights advances in process technology and system-level device integration. imec demonstrates junction-anneal-free n-type GAA nanosheet FETs utilizing very-low-temperature Si:P source/drain formation, addressing one of the key thermal-budget challenges for future monolithic and three-dimensional integration schemes. Georgia Tech presents a vertical channel-all-around power transistor capable of improving both power density and thermal characteristics for on-chip power delivery, an increasingly important requirement as logic systems continue to scale and demand greater energy efficiency. Finally, the session broadens the scope of advanced transistor innovation beyond traditional logic applications. imec reports the world's first demonstration of a 40 nm gate-oxide FinFET with more than 30 V breakdown voltage, providing a promising device platform for future 3D NAND flash memory peripheral circuits and illustrating how process innovation can enable new application spaces while leveraging advanced CMOS manufacturing infrastructure. Collectively, the papers in this session demonstrate that the future of semiconductor technology will be driven not by a single breakthrough, but by the convergence of innovations in materials, device architecture, gate-stack design, and process integration. We believe these presentations will provide valuable insights into the technological building blocks required for the next generation of high-performance, energy-efficient computing and memory systems.

9:05 AM

33-1 Germanium Nanoribbon PMOS with Reliable and Scaled Gate Stack for Low-Voltage CMOS Technology

Wouter Mortelmans[1], Ashish Verma Penumatcha[2], Chelsey Dorow[2], David Kohen[2], Raseong Kim[2], Tiwari Vishal[2], Pratyush Buragohain[2], Rahul Ramamurthy[2], Jeffrey Hicks[2], John J Plombon[2], Adedapo A Oni[2], Andrey Vyatskikh[2], Thomas Hoff[2], Thoe Michaelos[2], I-Cheng Tung[2], Emmannuel Khora[2], Shashi Bhushan Sinha[2], Shane M Harlson[2], David Boss[2], Eric Mattson[2], Siddharth Chouksey[2], Ashish Agrawal[2], Nazmul Arefin[2], Scott B Clendenning[2], Tristan A Tronic[2], Kevin P O'Brien[2], Matthew V Metz[2], Jessica M Torres[2], Jack Kavalieros[2], Myung-Hee Na[2], Uygar E Avci[2]

[1]Foundry Technology Research, Intel Corporation, [2]Intel Corporation

The semiconductor industry has identified CFET technology as next innovation needed to continue CMOS scaling. The CFET architecture can enable hybrid-channel integration through vertically stacked Ge P- on Si NMOS devices. To realize Ge P-FET, a gate-oxide that meets performance, reliability and thickness targets is key. We present a Ge PMOS reliability study on scaled MOS capacitors and nanoribbon transistors. We use Si-cap and GeON approaches to demonstrate gate-stacks that meet CMOS technology requirements. The Ge MOSCAP with oxynitrided Si-cap

was scaled to record-thin EOT of 8.7Å with NBTI ~100x lower compared to Si. Thermally grown GeO_x is known to be a poor passivation layer, but our GeON MOSCAP matches Si NBTI at EOT of 4.8Å which is 6Å thinner compared to Si. Using GeON, we built a Ge nanoribbon transistor with high effective mobility of 230cm^2/Vs at scaled EOT of 6.4Å with 0.6x iso-field NBTI of Si
nanoribbon devices.

9:30 AM

33-2 Silicate-Mediated Volume-Less Dipole Reversal Enabling WFM-Free Multi V_T CFET Integration

Jaewon Chung[1], Sangkuk Han[2], Wonjae Choi[2], Haesoo Jang[2], Sangmyun Lim[2], Kyungwook Park[2,3], Sangwoo Jeong[2], Soyoung Park[2], Gwangyeon Lee[2], Yejin Han[2], Changhwan Choi[2]
[1]Department of Semiconductor Engineering, Hanyang University, [2]Hanyang University, [3]Samsung Electronics Co., Ltd.

We propose a novel volume-less Si flip dipole layer that reverses the interfacial dipole polarity of both AlO_x and LaO_x, enabling WFM-free multi-V_T tuning without the thickness penalty of conventional flip layers. Rather than growing a SiO_2 film, the volume-less Si reacts with oxygen in the underlying dipole oxide to form an interfacial silicate, reversing the dipole direction while preserving EOT scalability. The flip yields a negative V_T shift for AlO_x (nFET) and a positive shift for LaO_x (pFET) at ~50mV/step. Owing to spatial separation of the dipole from the channel, it suppresses remote Coulomb scattering, giving up to ~2× higher mobility, lower D_{it}, and improved SS. The silicate-based flip adds a negligible oxygen-vacancy penalty, enhancing BTI and TDDB reliability. Combined with the dipole-first scheme, a 7-level multi-VT operation is demonstrated for both n- and p-type devices, offering a scalable, low-thermal-budget, WFM-free multi-VT solution for CFET integration.

9:55 AM

33-3 First Demonstration of n-Ge (111)/p-Ge (110) Hetero-Oriented Ge CFETs for 6T SRAM bit-cell with Advanced Ru/Mo MOL Contacts

Fang-Jui Chu[1], Kuan-I Wu[2], Wen-Hsin Chang[3], En-Ching Chen[1], An-De Xu[4], Che-Hsun Tsai[4], Chien-Yu Liao[2], Hao-En Che[2], Yi-Lin Shih[2], Min-Hui Chuang[5], Chieh-Yang Chen[5], Yu-Wen Xu[2], Szu-Huan Hsu[5], Yu-Chi Wang[2], Kuan-Ju Chou[5], Tzu-Chieh Hong[2], Nein-Chih Lin[2], Ta-Chun Cho[6], Chien-Ting Wu[6], Chien-Hsueh Chiang[7], Guang-Li Luo[6], Kun-Lin Lin[6], William Cheng-Yu Ma[4], Yao-Jen Lee[2], Wen-Fa Wu[6], Tatsuro Maeda[3], Yiming Li[2], Yeong-Her Wang[1]
[1]National Cheng Kung University, [2]National Yang Ming Chiao Tung University, [3]National Institute of Advanced Industrial Science and Technology (AIST), [4]National Sun Yat-sen University, [5]Institute of Communications Engineering, National Yang Ming Chiao Tung University, [6]Taiwan Semiconductor Research Institute (TSRI), [7]Dept. of Electronic Engineering, National Yunlin University of Science and Technology

This work demonstrates the first hetero-oriented Ge CFET 6T SRAM bit cell integrating a top n-Ge (111) channel and a bottom p-Ge (110) channel. Carrier-specific orientation engineering yields peak electron and hole mobilities of 738 and 296 cm^2/V·s, respectively, enabling balanced n/p drive in a vertically stacked footprint. Ru and Mo are introduced as low-size-effect MOL contacts, with Mo improving device current and achieving an inverter gain of 47.1 V/V. Static butterfly curves and dynamic write, hold, and read waveforms experimentally verify full 6T SRAM operation. A measurement-calibrated circuit framework further links MOL resistance to array-level delay, predicting up to 28.9% reduction for Mo versus conventional Co. These results establish hetero-orientation and MOL-metal co-optimization as key enablers for scalable Ge CFET SRAM.

10:20 AM

33-4 Insights into Ultra-Scaled Single-Level and 3D-Stacked Nanosheet FETs

Anabela Veloso[1], Bjorn Vermeersch[1], Geert Eneman[1], Ruben Asanovski[1], Jeroen E. Scheerder[1], Paola Favia[1], Daniel Casey[2], Conor Cullen[2], Clement Porret[1], Thomas Dursap[1], Felix Seidel[1], Claudia Fleischmann[1,3], Ritam Sarkar[1], Arka Dutta[1], Thomas Chiarella[1], Andrea Pondini[1,3], Rami Khazaka[2], James Myers[1], Naoto Horiguchi[1]
[1]imec, [2]ASM, [3]KU Leuven

We report on ultra-scaled, tight pitch p-type nanosheet (NS) FET-based devices built using low-temperature (LT) (<350°C) epitaxy for the S/D regions, providing in-depth insights into device performance by correlating it with process choices/control and layout effects, and with statistically sound DC and LF-noise data supported by physical analyses (e.g., (S)TEM, EDS, APT). Compressively strained channels for both single-level and CFET's top NSFET devices are measured after a similar S/D epi process, with epi grown only from the NS sidewalls for the latter.

A novel scheme to improve the thermal performance of NSFETs without impacting the channel strain is also proposed, combined with heat spreader layers in CFET (their material and position in stack evaluated). With feasibility of dual LT S/D epi electrically demonstrated, introduction of a backside (BS)-adapted S/D configuration for PMOS is shown to ease heat removal from the channel(s) via the BS contacts.

11:10 AM

33-5 Vertical Channel-All-Around Power Transistor Enabling Enhanced Power Density and Thermal Characteristics for On-Chip Power Delivery

Song-Hyeon Kuk, Kad Dokwan Kook, Chengyang Zhang, Hung-Chun Chou, Shimeng Yu, Satish Kumar, Suman Datta
Georgia Institute of Technology

To reduce routing I^2R loss which degrades the power efficiency of the system-on-chip (SoC) package, on-chip point-of-load fully-integrated-voltage-regulators (FIVRs) ha been proposed. However, their power transistors simultaneously require OFF-state breakdown voltage over 5 V, low specific on-resistance ($R_{on,\ sp}$), high power density, low thermal resistance (R_{th}) and high thermal capacitance (C_{th}), to minimize conduction loss and self-heating-induced temperature rise. Here, we demonstrate high-power-density (4910 kW/mm^2) and thermally resilient amorphous-oxide-semiconductor power transistor by transitioning from planar transistor (Planar) to vertical channel transistor (VCT) and vertical channel-all-around transistor (VCAAT), which scales the physical device footprint while simultaneously suppressing temperature rise. VCAATs exhibit low R_{th}, low transient temperature rise and ripple with high C_{th} due to their enhanced heat spreading capability, which were characterized by thermo-reflectance measurement and electro-thermal simulations. These results establish VCAATs as compact and thermally beneficial power switch for on-chip FIVR monolithically integrated into active backside power delivery network.

11:35 AM

33-6 Junction Anneal-Free n-type GAA Nanosheet FETs Featuring Very-Low Temperature (<450°C) Si:P S/D

Thomas Chiarella[1], Ritam Sarkar[1], Pierre Eyben[1], Niels Claessens[2], Brendan Marozas[2], Rami Khazaka[2], Clément Porret[1], Anjani Akula[1], Thomas Dursap[1], Christian Banti[1], Parth Ashish Rathi[1], Xiuju Zhou[1], Pallavi Puttarame Gowda[1], Ali Abdelgawad[1], Farid Sebaai[1], Anne Vandooren[1], Jishnu Ganguly[1], Maria Chistiakova[1], Efrain Altamirano-Sánchez[1], Alexis Franquet[1], Rita Tilmann[1], Hiroaki Arimura[1], Andrea Pondini[1], Min-Soo Kim[1], Lucas Petersen Barbosa Lima[1], Naveen Reddy[1], Jérôme Mitard[1], Steven Demuynck[1], Serge Biesemans[1], Naoto Horiguchi[1]
[1]imec, [2]ASM

This work presents the benefit of using very low temperature (VLT <450°C) phosphorus-doped epitaxially grown S/D junction (Si:P), gate all around (GAA) nanosheet transistors (NSFET) without junction anneal, down to 48 nm gate pitch and 16 nm gate length. The comparison to reference high temperature (670°C) co-flow with 850°C rapid thermal anneal (RTA) and the impact of S/D junction proximity below the gate are discussed for bottom dielectric isolated (BDI) junctions. The validation of this feature on nanosheet devices paves the way for cFET devices (complementary FETs) where compatibility of the nFET thermal budget with the underlying pFET is of prime importance. To our knowledge, this is the first demonstration of scaled n-type GAA NSFETs at 48 nm CPP using <450°C Si:P S/D without junction activation anneal, achieving matched Ion/Ioff performance and approximately 2× lower contact resistivity than the RTA-based reference.

12:00 PM

33-7 Novel Volumeless Threshold Voltage Control in Gate-all-around CMOS FET Enabled by Selective Oxidation of Work Function Metal

Kenzo Manabe, Hiroyuki Ota, Yukinori Morita, Misako Morota, Kazuya Uejima, Ryutaro Nishino, Atsushi Yagishita, Naoto Kumagai, Chia-Tsong Chen, Masanaga Fukasawa, Shutaro Asanuma, Toshihiro Kamei, Yuuki Ishida, Naoya Okada, Yuji Kasashima, Takahiro Goya, Toshifumi Irisawa, Fuminori Ito, Takashi Matsukawa, Yoshihiro Hayashi
National Institute of Advanced Industrial Science and Technology (AIST)

In this work, we demonstrate a volumeless multi-threshold-voltage (V_{th}) approach based on selective oxidation of an n-type work-function-metal (WFM) for stacked gate-all-around (GAA) CMOS FETs. Selective oxygen introduction into TiAlC realizes effective work functions (EWFs) suitable for both NMOS and PMOS without increasing the WFM thickness, enabling PMOS V_{th} tuning while preserving the NMOS V_{th}. The proposed scheme achieves V_{th} modulation

without mobility degradation and also reduces PMOS gate leakage, consistent with an EWF-induced increase in the electron barrier height. These results demonstrate selective WFM oxidation as a promising integration-friendly solution for volumeless multi-V_{th} control in aggressively scaled GAA CMOS technologies.

12:25 PM

33-8 World's First Demonstration of 40 nm Gate-Oxide FinFET with ≥30 V Breakdown for 3D-NAND Flash Memory

Sandipta Roy[1], Ritam Sarkar[1], Philippe Matagne[1], Joao Bastos[1], Jishnu Ganguly[1], Romain Ritzenthaler[1], Yangyin Chen[2], Satoru Mayuzumi[2], Lin Hou[2], Eiichi Fujikura[2], Mohan Dunga[2], Vladimir Machkaoutsan[3], Shamin Houshmand Sharifi[3], Myunghun Woo[4], Sangmin Kang[4], Woojin Nam[5], Takashi Maeda[6], Stephan Brus[1], Emmanuel Dupuy[1], Vincent Brissonneau[1], Jelle Vande Weeghde[1], Tanushree Sarkar[1], Johan De De Backer[1], Andrea Mingardi[1], Min-Soo Kim[1], Jerome Mitard[1], Alessio Spessot[1], Naoto Horiguchi[1]
[1]imec, [2]SanDisk, [3]Micron Technology, Inc., [4]Samsung Electronics, [5]SK hynix, [6]Kioxia Corporation

For the first time, a functional high-voltage (HV) FinFET transistor for 3D-NAND Flash memory has been experimentally demonstrated. The fabricated FinFET achieves an off-state breakdown voltage exceeding 30 V under optimized process conditions with fin widths scaled down to 20 nm and 40 nm SiO_2 gate oxide with poly-Si gate. With identical junction conditions, the FinFET achieves higher I_{on} at the same footprint compared with the planar reference device. The FinFET also demonstrates improved body-bias immunity, reducing the V_{thn} shift by ~2 V under a 25 V body bias. The off-state breakdown voltage remains nearly independent of fin width, indicating junction-limited breakdown behavior. These results demonstrate the capability of FinFETs as compact HV word-line pass transistors for 3D-NAND, enabling significant area scaling at iso-performance compared with current state-of-the-art planar devices.

Session 34: MS | Multiscale, AI & Advanced Simulation Methodologies

9:00 AM, Continental 7-9

Co-Chairs: Yan Zhang, MIT Lincoln Lab and Xi-Wei Lin, Synopsys

AI, machine learning, and physics-based modeling are converging to accelerate semiconductor innovation across devices, materials, memories, reliability, and interconnects. This Modeling & Simulation session presents five papers covering the following topics: (1) an autonomous AI-driven compact-model generation framework that converts measurement data into SPICE-ready transistor models; (2) a physics-aware graph-learning approach that captures disorder-induced variability in amorphous oxide transistors; (3) a probabilistic digital twin that enables rapid optimization of RRAM programming policies with large hardware gains; (4) a defect-to-device leakage simulation framework linking spectroscopy, quantum defect physics, and Monte Carlo transport; and (5) an experimental and theoretical investigation of topological semimetals showing unconventional resistivity scaling that could enable future nanoscale interconnects. Together, these papers illustrate a common trend toward combining first-principles physics, experimental validation, and AI-powered automation to reduce development time, improve predictive accuracy, and accelerate next-generation semiconductor technology deployment.

9:05 AM

34-1 A Fully Automatic Device Compact Modeling Framework Combining Symbolic Physics-KAN and LLM Co-operative Multi-agent System

Yu Huang[1], Zikang Mei[2], Hao Gong[2], Runchuan Ma[2], Wei Tang[2], Xiaojun Guo[2]
[1]School of Integrated Circuits, Shanghai Jiao Tong University, [2]Shanghai Jiao Tong University

Device compact modeling is essential, yet it is bottlenecked by the trade-off between physical interpretability and automation. This work proposes a multi-agent co-operative compact modeling framework that combines a physical kernel with Kolmogorov-Arnold network (KAN) correction. The physical kernel supplies baseline parameters for trainable B-spline corrections that are symbolically distilled into closed-form Verilog-A formulas, while multiple LLM agents automate the full modeling cycle with structured outputs that prevent hallucination. The approach is demonstrated on fabricated dual-gate TFTs of IGZO, organic, and a-Si, producing a SPICE-ready model (RMSE 0.073 dec, model-to-SPICE deviation <0.003%) for IGZO from scratch in under 2 hours, with fewer parameters than equivalent MLP models, and generalizing to the other materials with comparable accuracy.

9:30 AM

34-3 Unveiling Spatial Disorder-Modulated Transport in Amorphous Oxide-Semiconductor FETs: From 3D TCAD to Physics-Aware Graph Learning for Variability Prediction

Wei Zhang[1], Juntong Chen[1], Kai Chen[1], Yue Jiet Chong[1], Zixuan Wang[1], Jianze Wang[1], Zhen Wu[1], Tao Du[2], Chen Sun[1], Ning Ji[1], Yan Song[2], Ye Lu[2], Xiao Gong[1], Shengman Li[1], Xuanyao Fong[1]
[1]National University of Singapore, [2]Fudan University

This work combines transfer characteristics from 250 IGZTO devices, localizeddisorder-resolved three-dimensional TCAD, and physics-aware graph learning. A stochastic Voronoi channel assigns each finite correlation region randomized donor concentration, band mobility, shallow-donor states, and conduction-band-tail states. After calibration to a representative ID-VG curve, the ensemble reproduces the measured distributions of threshold voltage (VTH), subthreshold swing (SS), and transition-region width (VTR). Electron-density and conduction-band maps reveal a gate-driven evolution from isolated pockets, through a few critical source-to-drain paths, to extended conduction. The same regions are encoded by a self-supervised Voronoi graph neural network (VO-GNN), with model inputs separated into topology, geometry, transport, and AOS trap-state features. With 100 labeled TCAD devices, VO-GNN achieves R^2 = 0.992/0.988/0.928 and mean absolute errors of 1.2 mV/1.83 mV/dec/6.83 mV for VTH/SS/VTR, respectively, while preserving ensemble statistics and substantially reducing batched evaluation time relative to TCAD.

9:55 AM

34-4 Physics-Guided Probabilistic Digital Twin for AI-Enabled Real-to-Sim-to-Real Optimization of RRAM Array Programming Policies

Zhixing Jiang[1], Zichang Liu[1], Mukun Cui[1], Jianshi Tang[1], Yuyao Lu[1], Ruofei Hu[1], Jiadong Zhang[1], Junchen Li[1], Yuan He[1], Kaimeng Liu[1], Chengxiang Ma[1], Yibei Zhang[1], Qi Qin[1], Ze Wang[1], Yanbo Su[1], Junyang Zhang[1], Yijia Fan[1], Haixia Yu[2], Chao Yue[2], Xingjun Wu[1], Yudeng Lin[1], Peng Yao[1], Bin Gao[1], He Qian[1], Huaqiang Wu[1]
[1]Tsinghua University, [2]Tongxin Microelectronics Co., Ltd.

High-precision and fast programming remains a long-standing bottleneck limiting the practical deployment of analog RRAM computing-in-memory systems. We present a physics-guided probabilistic digital twin for AI-enabled Real-to-Sim-to-Real optimization of RRAM array programming policies. Using vast wafer-scale experimental data, the twin employs a Transformer encoder and physics-regularized two-component Gaussian-mixture decoders to model history-conditioned, non-Gaussian voltage-transition and relaxation distributions. Across 200 hardware-measured policies excluded from training, the twin predicts array-level average programming pulse number (*$\overline{PN}$*) and relative deviation (*RD*) with R^2=0.949/0.951, close to the experimental repeatability. The model is fully validated across **four RRAM technologies (130/40/28/12nm)**, where cross-process adaptation requires ~50× fewer target-domain samples than training from random initialization. With two hardware-calibration rounds, policy search offers an estimated >2000× speedup over hardware-only optimization. Final hardware measurements verify that the distilled 32-leaf decision-tree policy advances the *$\overline{PN}$*-*RD* frontier, reducing *$\overline{PN}$* by ~45% compared to typical delay-verify scheme and extending stability time by ~300×.

10:45 AM

34-5 Bridging the Defect-to-Device Gap: Experimentally Resolved Multielectron Trapping Enables Field–Temperature Monte Carlo Leakage Prediction

Mingyan Wang[1], Yixu Yao[1], Junjie Zhou[2], Ke Wei[1], Xinhua Wang[1], Xinyu Liu[1], Guanhua Yang[1], Shiyou Chen[2], Ling Li[1], Sen Huang[1]
[1]Institute of Microelectronics of Chinese Academy of Sciences, [2]Fudan University

Predictive dielectric-leakage modeling has long lacked a continuous path from quantum defect theory and defect spectroscopy to device-scale Monte Carlo transport. Here, we establish the first direct quantum-defect-to-device simulation chain. A bias-selective deep-level transient spectroscopy (DLTS) protocol resolves two successive electron-pair captures in SiNx dielectric, while spin-resolved hybrid-functional calculations assign them to the K-center $K^+ \rightarrow K^- \rightarrow K^{3-}$charge ladder. The resulting charge-state graph and quantum nonradiative-multiphonon (NMP) kernel drive a defect-resolved capture–emission–recapture state machine coupling kinetic Monte Carlo to trajectory-resolved ensemble Monte Carlo. The framework reproduces the complete measured field–temperature surface, outperforms an otherwise matched classical Marcus NMP kernel, and decomposes crossovers among competing leakage mechanisms. Converting experimentally resolved quantum defect kinetics directly into device

current, the framework establishes a transferable foundation for predictive leakage and reliability simulation across trap-limited dielectrics.

11:10 AM

34-6 Unconventional Resistivity Scaling in Topological Semimetals for Future Nanoscale Interconnects

Haotian Su[1], Hyeonhu Bae[2], Sydney Fultz-Waters[1], Young Suh Song[1], Anh Tuan Hoang[1], Sam Vaziri[3], Meera Patel[4], Andrew J Mannix[1], Milan Pesic[4], Xinyu Bao[3], Shan X Wang[1], Binghai Yan[2], Eric Pop[1]
[1]Stanford University, [2]Pennsylvania State University, [3]Taiwan Semiconductor Manufacturing Company (TSMC), [4]Applied Materials

As we enter the angstrom node era, interconnect dimensions must scale alongside those of transistors. Conventional metals, like Cu and Ru, display resistivity (ρ) increase from surface and grain-boundary scattering as their size is reduced. Topological semimetals (TSMs), quantum materials with surface states that are topologically protected against disorder, offer a promising alternative for ultra-scaled interconnects. Here, we *experimentally* show unconventional resistivity scaling for TSM NbP, reaching ~30 μΩ·cm at room temperature and 1.5 nm thickness. Importantly, ρ scaling occurs both on sapphire and on *amorphous* SiO_2 substrates, establishing a back-end-of-line-compatible TSM option with wafer-scale uniformity. We also use electronic structure *calculations* to study the robustness of surface states in 1.2-6 nm thin TSM films and realistic interconnect dimensions with amorphous dielectrics. Finally, we compare several TSM classes, including Weyl, triple-point, and chiral semimetals, revealing distinct topological origins and scaling mechanisms that provide fundamental insights into size-dependent transport in TSMs.

Session 35: RSD | Memory Technologies and Compute-in-Memory

9:00 AM, Imperial A

Co-Chairs: Gabriele Navarro, CEA-Leti and Grace Chung, Samsung

This session brings together eight articles covering reliability studies across a wide range of memory technologies and Compute-in-Memory (CIM) architectures. The first four articles address endurance degradation, disturbance mechanisms, retention, radiation tolerance, and long-term stability, along with practical mitigation and recovery techniques. Highlights include new approaches for optimizing the power-reliability tradeoff in 3D NAND, understanding disturbance mechanisms in 3D DRAM, extending the endurance of oxide-channel FeNAND, and improving the reliability of RRAM. The final four articles focus on CIM architectures and intelligent memory systems. Going beyond device-level demonstrations, several articles combine techniques at the device, array, circuit, algorithm, and system levels. Key contributions include self-correcting and fault-aware RRAM-CIM schemes, drift-resilient PCM systems for edge applications, and a 3D FeNOR CIM macro designed for KV-cache acceleration in edge LLMs. What makes this session particularly interesting is the strong connection between device physics and system-level performance. Rather than simply identifying physical reliability limits, the articles propose architectural, coding, recovery, or control schemes to overcome them, often validated using fabricated arrays, test chips, or physical hardware. Overall, this session highlights how advances in device physics, reliability engineering, circuit design, and system co-design converge to enable energy-efficient, resilient, and intelligent memory for next-generation AI and high-performance computing.

9:05 AM

35-1 RowHammer and ColumnDisturb in 3D DRAM: New Vulnerability and Bitline Slice-Tracking Mitigation

Kiseok Lee, Faaiq Waqar, Sungwon Cho, Zixuan Pan, Suman Datta, Shimeng Yu
Georgia Institute of Technology

3D DRAM is a promising technology for sustaining DRAM scaling beyond the sub-10 nm node. However, its vertically integrated architecture introduces new dynamic disturbance mechanisms driven by wordline coupling and floating-body effects. This work establishes a device-to-system vulnerability analysis framework that uncovers a strong interaction between the RowHammer and ColumnDisturb phenomena. By comprehensively modeling all leakage mechanisms under realistic high-aspect-ratio etch process variations, we predict RowHammer and ColumnDisturb thresholds of approximately 57k and 120k activations, respectively—both significantly below the maximum activation count within a conventional 64 ms refresh window. To address this challenge, we propose a selector-based bitline slice-tracking mitigation scheme that reduces the counter area from 3% to 0.1% while effectively

suppressing disturbance accumulation. Cycle-accurate Ramulator simulations show that the proposed approach provides reliable protection with less than 1% performance overhead, making it a practical solution for next-generation 3D DRAM systems.

9:30 AM

35-2 Revisiting Threshold Voltage Placement in 3D NAND Storage: An Energy-Reliability Co-Optimization

GIUK KIM[1,2], Kyungmin Lee[2], Taehyeon Song[2], Sizhe Ma[1], Yixin Qin[1], Moonyoung Jung[3], Suhwan Lim[3], Kwangyou Seo[3], Wanki Kim[3], Daewon Ha[3], Vijaykrishnan Narayanan[4], Myounggon Kang[2], Kai Ni[1]
[1]University of Notre Dame, [2]University of Seoul, [3]Samsung Electronics, [4]Pennsylvania State University

We reveal how memory-window (MW) position governs trade-offs among energy, READ/PASS disturb, and program inhibition in 3D NAND, and identify a co-optimized point through experiments and modeling. We develop: (i) a 3D structure-aware energy model incorporating MW-dependent voltages and array parameters, (ii) an analytical disturb model capturing MW position and gate-stack charge dynamics, and (iii) a PGM-inhibit scheme enabling self-boosting at V_{PASS} =1 V under X-, Y-, and XY-mode conditions. The framework identifies an optimum TLC MW centered at -5.2 V, reducing ISPP and READ energies by 97% and 96%, respectively. This MW is experimentally realized, while the disturb model yields structural criteria for suppressing READ disturb. Moreover, when conventional global self-boosting fails at V_{PASS} of 1 V, the proposed scheme restores reliable inhibition with negligible PGM disturb under worst-case conditions. This work establishes a unified energy-disturb-inhibition framework for efficient and reliable future NAND storage.

9:55 AM

35-3 Endurance Failure Pathways in Oxide-Channel Ferroelectric NAND: Switching-Accelerated Breakdown and Trap-Mediated Memory Window Degradation

Jung-Kyun Kim, Taek Jung Kim, Yoonsang Park, Yumin Kim, Seokhoon Choi, Yunseong Lee, Moonil Jung, Kwang-Hee Lee, Sanghyun Jo, Kihong Kim, Donghoon Kim, Kyung Mee Song, Kyunghun Kim, Young-Chul Leem, Sangwook Kim, Seung-Yeul Yang, Yongsung Kim, Duk-Hyun Choe, Sijung Yoo
Samsung Advanced Institute of Technology (SAIT)

We investigate the endurance failure mechanisms of oxide-channel ferroelectric NAND (FeNAND) for reliable operation. While the depolarized erase (ERS) state reduces polarization-induced cycling degradation unlike conventional Si-channel FeNAND, oxide-channel FeNAND still exhibits two critical endurance limits: catastrophic breakdown and trap-mediated memory window (MW) degradation. In this work, we distinguish these two failure modes and identify the underlying mechanisms. Based on this understanding, we propose an operation-based strategy to recover the MW. These results clarify the reliability bottleneck and provide guidelines for reliable oxide-channel FeNAND.

10:20 AM

35-4 DART: A Differential Antiparallel RRAM Technique with Enhanced Sensing Margin for Ultra-High Reliability eNVM

Jingwei Sun[1], Haoran Wang[1], Zongwei Wang[1,1], Yingchen Ji[1], Shuai Hu[1], Chengyu Liu[1], Zhixing Cai[1], Yujie Chen[1], Lin Bao[1], Ling Liang[1], Chen Li[2], Shoumian Chen[2], Yuhang Zhao[2], Yimao Cai[1]
[1]Peking University, [2]ICEM

Mixed-criticality systems require differentiated reliability, which conventional fixed-reliability eNVM cannot provide. We propose a reconfigurable differential antiparallel RRAM technique (DART). Its differential mode provides a 52.7-μA sensing margin, supporting 20-year-equivalent retention at 150°C and stable sensing during heavy-ion single-event effect (SEE) testing. Stuck-at-fault (SAF) detection and reconfiguration further enable 10^6-cycle endurance. DART freely switches between differential mode for critical data and independent 1T1R mode for ordinary data, improving memory efficiency by at least 24%. Its antiparallel switching completes complementary programming in one step, reducing write latency by at least 50% at comparable reliability and enabling reliable, efficient mixed-criticality eNVM.

11:10 AM

35-5 A Cross-Layer Co-design Reliability Framework Featuring Residue-Checked Self-Correction with 100% VMM Accuracy for Deterministic RRAM-Based Compute-in-Memory Achieving Lossless Image Compression

Yuxiang Jia[1], Qiyu Wang[1], Yiyang Chen[1], Ao Shi[1], Hao Ai[1], Yading Yi[1], Siyuan Chen[1], Shiyang Li[1], Kexun Li[1], Lianliang Wu[1], Nan Tang[1], Guangjin Li[1], Lei Sun[1], Yulin Feng[2], Lifeng Liu[1], Peng Huang[1]
[1]Peking University, [2]Beijing Information Science and Technology University

RRAM-based Compute-in-Memory (CIM) offers substantial energy-efficiency and latency advantages through highly parallel in-situ computing. However, post-programming relaxation, conductance variation and long-term drift can collectively lead to vector–matrix multiplication (VMM) errors, hindering the extension of CIM beyond error-tolerant workloads. This work presents a device-array-algorithm co-design reliability framework with three key advances: 1) An Adaptive Stabilization Programming Scheme (ASPS) is proposed to dynamically identify device-specific stabilization pulse-pair amplitudes through response-guided parameter search during programming, suppressing post-programming relaxation and reducing device-to-device variation by 62.6% compared with conventional ISPP. 2) For the first time, a Residue-Checked Self-Correction (RCSC) is proposed to verify and correct VMM results through a real-time modulo-3 residue check, maintaining 100% VMM accuracy throughout 10^5 s. 3) Lossless image compression is experimentally demonstrated on a fabricated 28nm 576Kb RRAM-CIM chip for the first time, achieving lossless reconstruction with zero VMM RMSE and zero MSE.

11:35 AM

35-6 Comprehensive Reliability Enhancement for High-Precision RRAM Compute-in-Memory via Fault-Aware Hadamard Transform and Low-Rank Compensation

Jichang Yang[1], Xinyuan Zhang[1], Jia Chen[2], Meng Xu[1], Hegan Chen[1], Yan Zeng[1], Jizhong Jiang[1], Songqi Wang[1], Yi Li[1], Zhongrui Wang[3], Yi Li[2], Han Wang[1]
[1]THE UNIVERSITY OF HONG KONG, [2]Huazhong University of Science and Technology, [3]Southern University of Science and Technology

Precision-sensitive edge AI demands reliable matrix computation within tight energy budgets, making RRAM CIM a promising substrate. However, sparse stuck-device errors cause severe local distortions, while programming noise, post-write relaxation, and retention drift introduce distributed conductance errors. Peaked target conductance distributions further complicate mapping, whereas hardware-aware retraining is data-dependent and full-array correction incurs precision-proportional overhead. We present HT-LRC, a training-free reliability framework combining structured Hadamard transform (HT) and low-rank compensation (LRC). Fault-map-guided HT maps compatible target conductances to stuck-device locations and smooths the target distribution, reshaping the residual for compact SVD-based LRC. An analog LRC branch corrects dominant residual components without task data or intermediate digitization. On a fabricated RRAM-CIM prototype, HT-LRC restores digital-equivalent unconditional and conditional 3D point-cloud generation, reduces generation error by up to 78%, and achieves 31-fold lower energy than a GPU, providing a low-overhead route to lifecycle RRAM-CIM reliability.

12:00 PM

35-7 A 2-Mb Drift-Robust PCM Chip for Privacy-Sensitive Edge Intelligence: Intermediate-Field Slope Readout for Multilevel Storage and Column-Level Self-Reference for Compute-in-Memory

Longhao Yan[1], Xinjie Wen[1], Ruiheng Wang[1], Zelun Pan[1], Xi Li[2], Zhitang Song[2], Yuchao Yang[1,1,1,1]
[1]Peking University, [2]Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences

Phase-change memory (PCM) combines dense nonvolatile storage with energy-efficient compute-in-memory (CIM), but structural-relaxation-induced conductance drift narrows multilevel read margins and degrades analog CIM accuracy. We present a 2-Mb carbon-doped $Ge_2Sb_2Te_5$ (C-GST) PCM chip fabricated in a 40-nm foundry CMOS process with two workload-matched drift-robust readouts. For multilevel storage, intermediate-field slope readout extracts the state-dependent ln – slope from two voltages, retaining 93% of the initial slope response at 10^7 s, compared with 43% under conventional low-field Ohmic readout. For repeated CIM, single-voltage column-level self-reference normalizes differential charge by the simultaneously accumulated total charge, suppressing common-mode scaling and reducing feature-scaling error from 38.6% to 2.6% at 10^7 s without dedicated drift-tracking cells. A home-assistance demonstration achieves 0.781% raw-key BER and 100% fall recall at 10^7 s, validating reliable credential recovery and drift-robust local perception.

12:25 PM

35-8 First Demonstration of 3D FeNOR CIM Macro with Device-Array-System Co-Design for Reliable KV Cache Accelerator in Edge LLM

Zhiyuan Ning[1], Yuejia Zhou[1], Weiqiin Huang[1], Guangnan Dai[2], Runteng Zhu[1], Hanyong Shao[1], Mingxiang Yu[1], Xiaojian Xu[1], Jinghao Yang[1], Wenpu Luo[1], Zhiqi Liu[1], Wenshuai Yao[1], Yuancheng Yang[3], Zhiliang Xia[3], Zongliang Huo[3], Yihan Zhang[2], Kechao Tang[1], Ru Huang[1]

[1]Peking University, [2]Hong Kong University of Science and Technology (HKUST), [3]Yangtze Memory Technologies Co., Ltd. (YMTC)

In edge LLM, 3D NOR-type FeFET (FeNOR) based CIM offers comprehensive solutions for the capacity, endurance, and data-movement bottlenecks of KV cache. However, device variation undermines reliable KV append and attention computation. Through device–array–system co-design on a 4×32×32 3D-FeNOR array, we demonstrate a reliable and low-latency CIM macro for Append-and-Attention in KV cache. 3D-FeNOR with an Al2O3 interlayer achieves an enlarged memory window and thereby a 3× lower state-overlap error ratio, with >107 cycles endurance and >10 years extrapolated retention. A novel CIM-Native Write is employed for accurate append, while reducing latency by 15.4× compared to regular verification. We further propose Paired-Differential Encoding and Reliability-Aware Entry Mapping to suppress bit- and token-wise error accumulation, increasing context-output fidelity from 3.55% to 88.48%. The 3D FeNOR CIM deployed on a PCB shows 99.65% raw ADC accuracy, while end-to-end evaluation indicates up to an 894× accuracy–energy–latency FoM improvement via the combined optimizations.

Session 36: ODI | Advanced Image Sensors

9:00 AM, Imperial B

Co-Chairs: Keita Yasutomi, Shizuoka University and Jiwon Lee, POSTECH

This session includes seven papers highlighting the latest developments in CMOS image sensor (CIS) technologies. The first paper by Y. Nonaka et al. from Sony presents an advanced hybrid Cu-Cu chip-on-wafer (CoW) stacking technology for a versatile heterogeneous integration platform. The second paper by H. Kim et al. from Samsung Electronics showcases a 0.6 μm pitch 200 MP image sensor achieving 1.55 e- read noise in 4x4 binning mode for high-SNR low-light imaging. The third paper by Salvo et al. from Meta presents a 3-layer stacked AI image sensor featuring 512-GOPS processing capability and a large on-chip memory for wearable devices and robotics. The fourth paper by Zhao et al. from GalaxyCore presents a 0.64 μm single-wafer stacked CIS with an advanced deep trench isolation (ADTI) structure. Next, in an invited paper, Tsuru et al. from Kyoto University introduce a wide variety of sensor architectures for event-driven X-ray detection with sub-10 μs temporal resolution and excellent spectroscopic performance. The sixth paper by Hayashi et al. from Canon proposes an 8-memory charge-domain global shutter 6.4 μm CIS enabling temporally aligned HDR imaging and full-area dual-pixel phase detection auto focus (PDAF). Lastly, K. Aoyama et al. from Sony demonstrate a single-exposure HDR CIS featuring the smallest reported lateral overflow integration capacitor (LOFIC) pixel at 1.45 μm pitch with low read noise.

9:05 AM

36-1 Scalable CMOS Image Sensor Architecture Utilizing Advanced Chip-on-Wafer Stacking

Yasuhiro Nonaka, Junichiro Fujimagari, Kan Shimizu, Kenya Yoshimura, Takashige Doi, Hitoshi Okano, Yusaku Sugimori, Yoshihisa Kagawa, Kentaro Akiyama

Sony Semiconductor Solutions Corporation

We developed a scalable CMOS image sensor (CIS) architecture utilizing advanced hybrid Cu-Cu chip-on-wafer (CoW) stacking. The proposed architecture integrates multiple logic chips on the backside of a sensor wafer, enabling future high-functionality CIS architectures and heterogeneous integration without increasing pixel-chip area. Two key enabling technologies were developed: leakage suppression technology for logic chips with a 3-μm-thick Si layer and a gap-fill planarization process that reduced bonding-surface topography from approximately 1000 to 200 nm. A prototype device demonstrated 6-μm-pitch hybrid Cu-Cu bonding and successful operation of stacked logic chips. The architecture increased ADC parallelism from 4 to 12 while maintaining CoW flexibility and cost advantages. Imaging evaluations confirmed no characteristic degradation in the gap-fill regions. These results demonstrate the feasibility of the proposed scalable CIS architecture for future high-functionality and heterogeneous CIS architectures.

9:30 AM

36-2 The Smallest Dark Random Noise in the 4x4-Binning Mode of 200-Megapixel CMOS image sensors

Hyungchae Kim, Seunghwan Lee, Junoh Kim, Daehyung Lee, Youngbin Park, Tae Hyuk Kim, Seungwon Yang, Kieyoung Woo, Nanhee Kim, Jinsuk Huh, Yun Hyeok Kim, Eun Sub Shim, Hyeongdong Jo, Sang-In Bae, Hyeonseok Han, Kyeong-Jae Byeon, Taesub Jung, Jungbin Yun, Howoo Park, Kyungho Lee, Haechang Lee
Samsung Electronics

This work presents a 200-MP sensor featuring world-leading noise characteristics. By implementing advanced dual-VTG technology and an optimized signal-transfer circuit to minimize charge trapping and parasitic capacitance, respectively, we achieved a dark random noise of 1.55 e– in 4x4- binning mode—the lowest reported value among 200-MP sensors to date. Additionally, a pixel-fixed-pattern-noise of 0.5e– was achieved in the same mode; this, together with the reduced random noise, decreased the multi-frame averaged total noise by 47%, resulting in a 2 dB SNR increase under 0.5-lux illumination. These results highlight the critical role of reducing both dark random noise and pixel-fixed-pattern-noise in low-light imaging.

9:55 AM

36-3 A 50M 0.64μm Pixel CMOS Image Sensor with Monolithic Homogeneous Integration Using an Advanced Deep Trench Isolation Structure

Stanly Zhao[1], Weihong Yang[2], Hugo Zhang[1], Xia Tang[1], Haowen Zhao[2], Yang Lan[1], Kun Huang[1], Xiaojun Guo[2]
[1]GalaxyCore Inc., [2]Shanghai Jiao Tong University

This work presents a monolithic homogeneous integrated backside-illuminated CMOS image sensor (CIS) with an advanced deep trench isolation (ADTI) structure. Process and layout co-optimization relieve epitaxial stress, enabling direct growth of defect-free single-crystalline silicon over deep trenches. This continuous device-quality layer accommodates pixel arrays and logic transistors while retaining deep trench isolation (DTI) structures. The 0.64-μm-pitch pixel achieves a full-well capacity (FWC) of 7000 e^-, a read noise of 0.84 e^-, and a random telegraph signal (RTS) noise of 2 e^- rms, yielding a single-exposure dynamic range of 78.4 dB. A 50 MP monolithic integrated CIS chip is successfully demonstrated.

10:20 AM

36-4 Whidbey: A 13MP, 0.81μm, 512GOPs, 3-Layer Stacked AI Image Sensor for Wearables and Robotics

Barbara De Salvo[1], Raffaele Capoccia[1], Tsung-Hsun Tai[1], Song Chen[1], Korina Fotopoulou[1], Aparna Ramachandran[1], Zhao Wang[1], Kleber Stangherlin[1], Stefano Ambrogio[1], Ziyun Li[1], Syed Shakib Sarwar[1], Jieyu Lin[1], Shuwei Lin[1], Ennio Salemi[1], Gianmarco Ottavi[1], Vikas Natesh[1], Jiajia Wu[1], Jonathan Thier[1], Erick Roane[1], Chiao Liu[1], Chin-Chun Wang[2], Chih-Hao Lin[2], Wei-Fan Chou[2], Hideyuki Fukuhara[2], Kazuya Mori[2], Rimon Ikeno[2]
[1]Meta Reality Labs, [2]Brillnics

In this work, we present Whidbey, a novel 3-layer (45nm/40nm/16nm) monolithic AI image sensor. Whidbey has a 4096x3200 resolution with a 0.81μm pixel, and supports both color rolling-shutter (RS) mode for point-of-view capture and monochrome global-shutter (GS) mode for computer vision. It integrates high-performance AI compute (512 GOPs) and large on-chip memories (10MB SRAM, 4MB eMRAM). With the Whidbey software devkit, we demonstrate extended on-sensor use cases, such as wake gesture, hand/person tracking, continuous learning, and a novel video neural tokenizer. Whidbey's AI sensor solution is unprecedented in today's semiconductor landscape and will shape the future of AI wearables and humanoid robots.

11:10 AM

36-5 Trigger-Output Event-Driven X-ray Astronomy SOI Pixel Sensor "XRPIX"

Takeshi Go Tsuru[1], Tomonori Ikeda[2], Takayoshi Kohmura[3], Shunsei Anzai[3], Hikaru Sato[3], Moeno Nakatani[3], Yuusuke Uchida[4], Kouichi Hagino[5], Takuto Narita[5], Yasushi Fukazawa[6], Yusuke Suda[6], Masaki Hashizume[6], Hideaki Katagiri[7], Mika Kagaya[7], Ikuo Kurachi[8], Yasuo Arai[9], Takehiro Takayanagi[9], Koji Mori[10], Ayaki Takeda[10], Hiromasa Suzuki[10], Taishu Kayanoki[10], Yusuke Nishioka[10], Sho Nagayama[10], Koki Sugimoto[10], Ryo Takagi[10], Maito Indo[10], Tomitaka Tanaka[10], Tsubasa Sakamoto[10], Maya Shimoto[10], Yume Hayano[10], Haruno Takamoto[10], Yudai Iiba[10], Keita Ogata[10], Shota Homma[10], Keisuke Mori[10], Akari Ogura[10], Ryoma Setoguchi[10], Chikato Chinen[10], Koya Yamasaki[10], Kenta Yamada[10], Kumiko Nobukawa[11], Reo Matsui[11], Shotaro Sato[11], Yuto Takeshima[11], Fuga Matsumura[11], Kazuki Singu[11], Takaaki Tanaka[12], Yoshiyuki Onuki[13], Hiroumi Matsuhashi[13], Riki Sato[13], Kazuki Fujikawa[13], Kenji Shimazoe[13], Setsuo Sato[13], Kanta Nagai[13], Hikaru Sekiguchi[13], Macoto Fujimoto[13], Hiroyuki Uchida[14], Yusuke

Komura[14], Hikaru Uebayashi[14], Kou Sakamoto[14], Rinto Kawada[14], Woojoo Lee[14], Masayoshi Nobukawa[15], Mizuki Uenomachi[16], Bing Hoang Dang[16], Shoji Kawahito[17], Keiichiro Kagawa[17], Toshifumi Imamura[18], Takafumi Ohmoto[18], Masahiro Ono[18]
[1]Physics, Science, Kyoto University, [2]National Institute of Advanced Industrial Science and Technology (AIST), [3]Tokyo University of Science, [4]Japan Aerospace Exploration Agency (JAXA), [5]Riken, [6]Hiroshima University, [7]Ibaraki University, [8]D&S Incorporated, [9]High Energy Accelerator Research Organization (KEK), [10]University of Miyazaki, [11]Kindai University, [12]Konan University, [13]The University of Tokyo, [14]Kyoto University, [15]Nara University of Education, [16]Institute of Science Tokyo, [17]Shizuoka University, [18]A-R-Tec Corporation
XRPIX is a monolithic active pixel sensor for next-generation X-ray astronomy. Using Silicon-on-Insulator (SOI) technology, it integrates a thick, fully depleted high-resistivity sensor substrate with a low-resistivity CMOS circuit layer. XRPIX achieves temporal resolution better than 10 us via an asynchronous self-trigger while maintaining excellent spectroscopy performance. To optimize charge collection efficiency (CCE) and suppress dark current, a Pinned Depleted Diode (PDD) structure was introduced. Furthermore, "Digital XRPIX" integrates ADCs, DACs, and an on-chip pattern generator. This paper reviews the device evolution, pixel circuitry, CCE, spectroscopy, and digital functionalities for future space applications.

11:35 AM

36-6 8-Memory Charge-Domain Global Shutter CMOS Image Sensor Architecture for Smooth High-Dynamic-Range Video Capture

Shuhei Hayashi, Yusuke Onuki, Hajime Ikeda, Kento Watanabe, Yusuke Yamashita, Mahito Shinohara, Takeru Sakamoto, Shunya Kamioka, Saki Takagi, Naoki Ishikawa, Takenori Kobuse, Junji Iwata, Yasushi Matsuno, Katsuhito Sakurai, Takeshi Ichikawa
Canon Inc.
This paper presents a Super 35 mm-format, 6.4 μm-pitch, eight-memory charge-domain global-shutter FSI CMOS image sensor for smooth high-dynamic-range video capture. The proposed charge-domain GS pixel architecture provides in-pixel memories suited for time-multiplexed exposure, enabling incident-light control, temporally aligned HDR imaging, dead-time-free global-shutter operation, and full-area dual-pixel phase-detection autofocus within a single frame. A serial two-memory configuration between each photodiode and floating diffusion enables continuous exposure independent of readout timing by separating exposure and readout memories. The prototype achieves a 94 dB dynamic range and simultaneously acquires HDR image signals and full-area PDAF information. Time-multiplexed exposure suppresses highlight clipping and HDR motion artifacts while preserving natural motion blur, demonstrating the effectiveness of the proposed charge-domain architecture for large-format video applications.

12:00 PM

36-7 1.45-μm-Pitch CMOS Image Sensor Achieving 0.57 e^-rms RN and 109 dB Single-Exposure Dynamic Range with 300 μV/e^- Conversion Gain

Kohei Aoyama[1], Tetsuya Uchida[1], Tomohiko Kawamura[1], Shinobu Asayama[1], Masahiro Yamamoto[1], Yuto Furuya[1], Junya Omori[1], Takuro Naramura[1], Daisuke Uehara[2], Shunsuke Ashikari[2], Youhei Chiba[2], Takaaki Minami[2], Daisuke Kawasaki[2], Eiichi Tauchi[2], Yuya Sakurada[2], Takashi Abe[1], Yasushi Tateshita[1]
[1]Sony Semiconductor Solutions Corporation, [2]Sony Semiconductor Manufacturing Corporation
In this study, we developed a non-shared triple conversion gain lateral overflow integration capacitor (TCG-LOFIC) CMOS image sensor for security cameras with the smallest reported LOFIC pixel pitch (1.45 μm). Although achieving a 1.45-μm-pitch TCG-LOFIC pixel has been difficult because of the limited layout space for pixel transistors, we addressed this challenge using two technologies: (I) full trench isolation (FTI) on one side of the pixel transistor and (II) a conformal-doping-less structure in the upper FTI region. The sensor achieves a 109 dB single-exposure dynamic range, 0.57 e^-rms random noise with a maximum conversion gain of 300 μV/e^-, and 50% quantum efficiency with 80% modulation transfer function at 850 nm.

Session 37: MT | Oxide Semiconductor Memories

2:00 PM, Continental 4

Co-Chairs: Katherine Chiang, TSMC and Shoichi Kabuyanagi, Kioxia

Oxide semiconductor, due to its wide bandgap, ALD processability, and BEOL compatibility, has become the promising channel material for novel device architecture to enable continuous 3D scaling. This session captures the breadth of that progress across four key fronts: monolithically 3D-stackable DRAM, embedded and high-density DRAM, 3D NVM, and P-type oxide semiconductor, highlighting their potential for AI, SRAM replacement, and high-capacity data storage. The session is organized into two parts. In the first part, Samsung starts with a 3D-stackable gain-cell eDRAM based on IGZO vertical-channel transistors for AI applications. SEL then presents a 3-tier stacked 2T1C memory with crystalline indium oxide, developed as a potential SRAM replacement. Next, Imec presents stacked-transistor 2T0C eDRAM, featuring DTCO-optimized read latency. Finally, a new architecture called 1T Insertion-Gate Memory will be demonstrated by NTU for 3D DRAM application. In the second part, IMECAS presents the first 3D-stackable ferroelectric NOR memory based on GAA IGZO Nanosheet FeFET, featuring 10^{11}-cycle endurance, a 5.2 V memory window, and 3-bit-per-cell operation. POSTECH then demonstrates the first IGZO-based vertical electrochemical NAND device capable of 5-bit-per-cell operation at sub-7 V. The session concludes with state-of-the-art p-type SnO transistors from TSMC.

2:05 PM

37-1 Fully Logic BEOL Compatible and Monolithically 3D Stackable Capacitor-less Gain-Cell eDRAM enabled by single-gated IGZO VCTs for AI applications

Daewon Ha, K. Lee, D. Yeon, J. Park, Y. Lee, K.J. Moon, K. Yoo, S. Lee, S. Hong, C. Im, Y. Lim, S. Park, K. Kim, M. Kim, I. Shin, M. Terai, J.H. Bae, J. Seo, H. Lee, J. Kim, W. Jung, Y.S. Tak, H. Yoon, J. Lee, S.I. Cho, M.J. Hong, T. Kim, W. Lee, M.H. Cho, K. Park, J. Jeong, S.W. Park, S.H. Lee, P. Yun, Y. Son, S. Hyun, S.J. Ahn, Jaihyuk Song
Samsung Electronics

For the first time, we present monolithically 3D stackable gain-cell eDRAM based entirely on single-gated IGZO vertical channel transistors (VCTs), embedded between 80nm pitch metal layers with full compatibility of logic BEOL. A buried borderless contact (BBC), uniquely enabled by IGZO VCT structure, significantly reduces the number of contacts and metal lines from 7 to 5 and from 5 to 3, respectively. Fabricated gain-cells with a metal-insulator-metal (MIM) capacitor of 1.8 fF achieves a data retention time of >500s at 25°C and 40s at 85°C, $>10^{12}$ cycling endurance, and 10 years N-/PBTI lifetime. Combined with 3 bit/cell MLC operation and IGZO width scalability down to 7 nm make it integrated with smaller pitch metal layers, leading to a drastic increase in memory density.

2:30 PM

37-2 Crystal Indium Oxide FET with Extremely Low Off-State Leakage and Monolithic 3D 3-tier 200-MHz 2T1C Memory with Reduced Thickness Toward SRAM Replacement

Hiroki Inoue[1], Taichi Ebihara[1], Takanori Ito[1], Kota Yamamoto[1], Takeya Hirose[1], Kazuma Furutani[1], Yusuke Kuwahara[1], Tsutomu Yamamoto[1], Daisuke Okabe[1], Kikuo Miyata[1], Yoshinori Ando[1], Ryota Hodo[1], Yasuhiro Jinbo[1], Tsutomu Murakawa[1], Takanori Matsuzaki[1], Tatsuya Onuki[1], Yuto Yakubo[1], Kenichi Okazaki[1], Toru Nakura[2], Shunpei Yamazaki[1]
[1]Semiconductor Energy Laboratory, [2]Fukuoka University

We demonstrate a memory test chip featuring three-tier monolithically stacked 2T1C cell arrays on Si CMOS. Each 2T1C cell array utilizes crystal indium oxide FETs (crystal IO FETs) and a coplanar capacitor formed on the same plane. By using coplanar capacitors, the manufacturing process is simplified to achieve a 50% reduction in tier thickness. Device characteristics across all tiers exhibit V_{TH} of 0.326–0.447 V, μ_{FE} of 13.2–15.9 cm^2/Vs, and I_{ON} of 114–165 μA/μm under zero back-gate voltage. The I_{OFF} of the crystal IO FET is estimated at 160 zA/μm. All three tiers achieve R/W cycle frequencies of 150 MHz and reach 200 MHz with elevated voltage. Data retention exceeds 99.7% after 3600 s without refresh. Regarding write endurance, a 100% pass rate is maintained after 10^{11} write cycles. Under optimized conditions, the μ_{FE} measures 82.3 cm^2/Vs and the R/W cycle frequency is estimated at 504 MHz.

2:55 PM

37-3 Highly Scaled (Lg=25nm) Stacked-Transistor IGZO 2T0C eDRAM on 300-mm Wafers with DTCO-Optimized Read Latency

Arvind Sharma[1], Daisuke Matsubayashi[1], Charwak Apte[2], Hubert Hody[1], Yiqun Wan[1], Jie Li[1], Rishabh Kishore[1], Harold Dekkers[1], Alexandru Pavel[1], Arvis Jomerts[1], Hyungrock Oh[1], Russell Schreiber[2], John Wuu[2], Subhali Subhechha[1], Marie Garcia Bardon[1], Gouri Sankar Kar[1], Fernando Garcia Redondo[1], Dwaipayan Biswas[1], Attilio Belmonte[1]

[1]imec, [2]Advanced Micro Devices, Inc. (AMD)

We demonstrate highly scaled (down to Lg = 25 nm) IGZO 2T0C eDRAM with stacked transistors on 300-mm wafers and position it as a cost-effective embedded memory solution for future logic nodes. The fabricated stacked 2T0C bitcells achieve median retention time > 1000 s while maintaining Wtr/Rtr performance of Ion > 10 µA/µm. Through DTCO, we demonstrate that the integration of these bitcells in N7 BEOL enables ~44 Mb/mm2 macro density (> A14 SRAM density [1]), achieving ~5 ns / 2 ns read latency at 25 °C / 85 °C (~2x over baseline), < 2 ns write latency, and excellent disturb immunity for a 32 KiB macro.

3:20 PM

37-4 A Novel 1T Insertion-Gate Memory for High Density 3D-Stackable DRAM Applications

Yu-Shan Wu[1], Feng-Min Lee[2], Yu-Yu Lin[2], Hsu-An Chou[1], Yuan-Ming Liu[1], Yi-Shiang Tsao[1], Yu-Tsung Liao[1], Ching-Ju Tseng[1], Fan-Hao Shen[1], Ying-Jung Chen[1], Hsien-Ming Sung[1], Chun-I Lu[3], Kun-Jen Hsueh[3], Chih-Chao Yang[3], Ming-Hsiu Lee[2], C. W. Liu[1]

[1]National Taiwan University, [2]Macronix International Co., Ltd., [3]Taiwan Semiconductor Research Institute (TSRI)

A 3D-stackable low-voltage 1T Si–IGZO insertion-gate memory cell is demonstrated for the first time with intrinsic read/write selectability and zero-bias hold. Different from conventional floating-gate memory relying on high-field tunneling, the IGZO insertion gate serves as a storage node and is programmed by low-voltage WWL/WBL pulses. The stored SN charge shifts the effective V_T of the underlying Si pFET read transistor, enabling distinct I_{RBL} read states with a sensing margin exceeding 10^6. The cell is written at V_{WWL}= 0.7 V and held/read at V_{WWL}= 0 V, achieving over 5000 s retention. WBL-programmable multi-state readout is demonstrated by applying different WBL programming voltages. WBL-controlled inhibit operation selectively suppresses the Si read path without destructive disturbance of the stored states, providing 3T-like select functionality within a physically 1T cell. TCAD simulations further evaluate the potential scalability of the insertion-gate memory toward scaled structures. These results support low-voltage, high-density, and 3D-stackable memory applications.

4:10 PM

37-5 First Demonstration of Shared-BL/SL 3D FeNOR Memory Based on GAA IGZO Nanosheet FeFET, Fabricated by Channel-last Process, Featuring 10⊃1]⊃1] Endurance, 5.2V MW, and 3-bit/cell Operation

Shipeng Wang[1], Ruifeng Tang[2], Kaiping Zhang[2], Yang Yang[2], Xuedong Zhao[2], Chuanke Chen[2], Yu Liu[2], Jiebin Niu[2], Shengjie Zhao[2], Pengfei Jiang[2], Wei Wei[2], Di Geng[2], Chunmeng Dou[2], Qing Luo[2], Ling Li[2], Ming Liu[2]

[1]State Key Laboratory of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics of Chinese Academy of Sciences, [2]Institute of Microelectronics of Chinese Academy of Sciences

We experimentally demonstrate a 4-layer 3D-stackable ferroelectric NOR memory based on channel-last gate-all-around nanosheet (GAA NS) IGZO FeFETs with shared bitline (SBL) and shared source line (SSL) configuration. The proposed architecture is realized through one-step lithography and a channel-last process, enabling compact vertical FeNOR integration with reduced process complexity and thermal-budget protection of the IGZO channel. Benefiting from strong GAA electrostatic control and the NS IGZO channel, the devices exhibit an on/off current ratio above 10^6 with ultralow off-state current. The Al:HfO_2 ferroelectric layer further suppresses gate leakage and enables endurance beyond 10^{11} cycles together with a large memory window of 5.2 V. Reliable high-density memory operation is then verified by 3-bit/cell programming, disturb immunity, and read-crosstalk suppression. Based on the demonstrated 3D FeNOR scheme, a storage density of 18.28 Gb/mm^2 is projected, highlighting its potential as a storage-class memory candidate bridging the density-latency gap between DRAM and NAND flash.

4:35 PM

37-6 First Demonstration of IGZO-Based Vertical Electrochemical NAND (V-EcNAND) Achieving 5-bit/cell at 40 nm Channel with Sub-7 V Operation

Hyejin Kim[1], Minho Kim[1], Junghoon Park[2], Wanki Kim[2], Daewon Ha[2], Seyoung Kim[1]
[1]Pohang University of Science and Technology (POSTECH), [2]Samsung Electronics
We demonstrate a three-dimensional vertical NAND based on an IGZO-channel electrochemical field-effect transistor (V-EcNAND), in which the threshold voltage is modulated by reversible ion exchange between the channel and an ion reservoir layer through the electrolyte, governed by a low-activation-energy ion-hopping mechanism. By decoupling pass-bias-induced channel conduction from program-induced ion migration, the device achieves sub-7 V operation (PGM/ERS) without read disturbance or degradation of transfer characteristics. The memory window was enlarged and ion transport was facilitated through the optimization of the channel, electrolyte, and reservoir layers. A vertical EcFET with a 40 nm channel length, fabricated on an 8-inch wafer, further exhibits reliable 32-state (5-bit) operation, 10^{10}-cycle endurance, and stable retention, all achieved at a low thermal budget (< 200°C). Furthermore, string-level operation confirmed that unselected cells remain undisturbed under program bias, validating NAND compatibility. These results establish the V-EcNAND as a low-voltage candidate for next-generation high-density NAND technology.

5:00 PM
37-7 BEOL Compatible P-type SnO Transistors with Field-Effect Mobility > 5 cm²/V·s and I_{on}/I_{off} > 10^5
Cheng-Chen Kuo, Xiaheng Huang, Chengxing He, Shu-Wei Chang, Yu-Chuan Shih, Cheng-Hsien Wu, Sam Vaziri, Elia Ambrosi, Hengyuan Lee, Chun-Chen Wang, Saverio Ricci, Songrui Hou, Isha Datye, Paulina Rincon, Gerben Doornbos, Chun-Chieh Lu, Georgios Vellianitis, Viraj Nilesh Madhiwala, Yu-Cheng Su, I-Che Lee, Chen-Feng Hsu, Wei-Yen Woon, Mark van Dal, Yu-Ming Lin, Xinyu Bao, Min Cao
Taiwan Semiconductor Manufacturing Company (TSMC)
This work presents high-performance, p-type SnO thin-film transistors (TFTs) featuring PVD sputtered channel material. These high-quality p-type SnO films exhibit a high-purity SnO phase, a near-unity Sn:O ratio, and sub-nanometer surface roughness. Back-gated SnO TFTs were fabricated with channel lengths down to 150 nm and dual-gated TFTs were demonstrated. The optimized SnO TFTs deliver outstanding electrical characteristics, including field-effect mobility (μ_{FE}) > 5 cm²/V·s, I_{on}/I_{off} >1.5×10^5, threshold voltage (V_{th}) ~ −0.5 V, subthreshold swing (*SS*) of <200 mV/dec, and a low hysteresis <100 mV. Furthermore, we evaluate the potential of these p-type SnO TFTs in n-p 2T0C gain cells, achieving a promising read-to-leakage selectivity (I_{read}/I_{leak}) of ~3000. Finally, CMOS compatible SnO integration process was demonstrated on 300 mm wafers.

Session 38: EDT | Emerging Memories

2:00 PM, Continental 5
Co-Chairs: Soo Gil Kim, SK hynix and Suhwan Lim, Samsung
This “EDT | Emerging Memories” session features seven papers. First, a defect-coupled dipole engineering strategy enables wafer-scale multi-Vth IGZO integration for a low-voltage 2T0C DRAM with long retention and multi-bit storage. Next, a vertical-channel GAA IGZO high-voltage device for 3D-NAND Flash periphery achieves a record-high breakdown voltage. This is followed by an EC-VNAND string transistor that implements three-variable analog multiplication for hardware-native Gated Linear Attention computation at the device level. Subsequently, a BiCS-compatible GOMET architecture based on an IGZO channel achieves an ultrahigh-density 3D RRAM array. Next, field-free pSOT-MRAM writing via z-spin on a 300-mm wafer shows strong immunity to magnetic field disturbances, followed by a VCMA-MRAM array with reconfigurable TRNG and PUF functionality unified for zero-trust security. Finally, a WSe2-based mirror-coupled FET reduces the transistor count of a bistable latch, achieving an ultra-compact SRAM cell.

2:05 PM
38-1 First-Principles-Guided Defect-Coupled Dipole Engineering for Wafer-Scale Monolithic Multi-V_{th} IGZO Integration Enabling 0.75-V 2T0C DRAM with >2 ks Retention and 3-bit Storage
Xufan Li[1,2], Yanyu Yang[2], Tong Yu[2], Xiaoyue Feng[2], Xiaoyi Zhang[2], Shaotang Sun[2], Peiqi Guo[2], Tianhao Liao[2], Yupeng Lu[2], Gaobo Xu[2], Huaxiang Yin[2], Rui Chen[2], Jiawei Wang[2], Lingfei Wang[2], Ling Li[2], Ming Liu[2]

[1]State Key Laboratory of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics of Chinese Academy of Sciences, [2]Institute of Microelectronics of Chinese Academy of Sciences
Voltage compatibility with advanced low-voltage logic is a central requirement for embedded IGZO DRAM. However, low-voltage 2T0C operation remains constrained by limited independent threshold-voltage optimization for retention and readout. Here, we establish a first-principles-guided defect-coupled dipole engineering strategy for wafer-scale monolithic multi-threshold-voltage (multi-V_{th}) IGZO integration, validated in a 0.75-V 2T0C DRAM. Within the HfO_x/AlO_x/IGZO gate-stack, ALD oxidant chemistry and selectively patterned nanoscale AlO_x interlayers co-tune dielectric defects and interfacial dipoles, yielding an on-die V_{th} separation of up to 0.7 V while preserving comparable field-effect mobility. A ΔV_{th} of 0.5 V is selected for dual-V_{th} 2T0C cells on a 200-mm wafer, enabling: **(1)** >2 ks retention at V_{DD} = 0.75 V; **(2)** ~10 mV WWL-induced storage-node voltage drop; and **(3)** 3-bit storage with eight distinguishable states maintained for >250 s. This strategy enables low-voltage, capacitorless, multi-bit IGZO DRAM, providing a wafer-scale route to controllable V_{th} engineering for BEOL integration.

2:30 PM

38-2 Ultra-small Footprint Vertically Stacked IGZO-based GAA High-Voltage Device for Future 3D-NAND Flash and High Bandwidth Storage Periphery

Shankha Mukherjee, Roman Izmailov, Devin Verreck, Wei Zheng, Jie Li, Kruti Trivedi, Tathagata Paul, Yongbin Jeong, Aakash Patel, Alexandru Pavel, Christophe Lorant, Geert Van den Bosch, Maarten Rosmeulen
imec
As 3D-NAND Flash advances toward higher density and bandwidth, large-area high breakdown voltage (BV) transistors in program and read paths increasingly limit further peripheral footprint and latency scaling. In this work, for the first time, we report a CMOS BEOL and 3D-NAND compatible, IGZO-based vertical-channel gate-all-around (GAA) device that simultaneously delivers a high OFF-state BV and ON-state drive current ($I_{D,ON}$) at a substantially reduced footprint. By replacing poly-Silicon with IGZO in a multi-gate GAA architecture, we improve the OFF-state BV ~4×, achieving a BV > 27 V. We engineer the drain contact and ungated extension region to realize an $I_{D,ON}$ of ~1 μA/μm without sacrificing BV. Further gate pitch scaling to 45 nm boosts $I_{D,ON}$ to ~100 μA/μm while maintaining a BV ~18 V. Finally, we validate device scalability down to an ultra-scaled channel hole radius of 35 nm and demonstrate operation in an array for the first time.

2:55 PM

38-3 EC-VNAND String Enabling Three-Variable Analog Multiplication for Hardware-Native Attention

Yumin Kim, Uicheol Shin, Byonggwon Song, Yongwook Kim, Hyun Jae Song, Sunho Kim, Seonghun Jang, Garam Park, Seung-Yeul Yang, Yongsung Kim, Min-Hyun Lee
Samsung Advanced Institute of Technology (SAIT)
We present an electrochemical (EC) RAM vertical NAND string transistor that realizes three-variable analog multiplication — ID ∝ G × V_{WL} × V_D — in a single device stack, enabling hardware-native attention computation aligned with the kernelized similarity and temporal gating methods of Gated Linear Attention (GLA) for large language model (LLM) inference. A planar transistor ECRAM device fabricated on an 8-inch wafer demonstrates a memory window of 4.87 ± 0.24 V (CV (σ/μ) = 4.99 %, n = 35) and endurance exceeding 10^5 cycles with an ion barrier layer (IBL). A three-gate vertical transistor ECRAM (VTr-EC) string validates series-channel current summation across independently programmed wordlines (WLs), with measured currents matching harmonic predictions (R^2 = 0.925, N = 2500). GLA simulation incorporating the EC-VNAND conductance model achieves lower perplexity (PPL) than softmax attention on standard natural language processing (NLP) benchmarks, establishing the first device-physics-based attention kernel without dedicated algorithmic circuits.

3:20 PM

38-4 GOMET: A BiCS-Compatible Gate-Operated Memristor-Embedded Transistor Topology for Scalable 3D RRAM Enabling 828 Mb/mm² Density in 32-Tier Array

Yi Ding, Mingcheng Shi, Yuyan Wang, Yuan He, Ruofei Hu, Yuankun Li, Jiaming Li, Bowen Shen, Benjamin Yang, Zhixing Jiang, Yiming Zhou, Zhen Zhu, Xurui Zhu, Wen Sun, Qingtian Zhang, Bin Gao, He Qian, Jianshi Tang,

Huaqiang Wu
Tsinghua University
Scaling 3D RRAM demands both high vertical density and cell-level selectivity, yet conventional 1T1R cells cannot be directly integrated into bit cost scalable (BiCS) stacks because shared intermediate electrodes electrically shunt neighboring cells. Here we demonstrate GOMET, a BEOL-compatible gate-operated memory cell with a recessed C-shaped IGZO channel serving simultaneously as the access transistor and RRAM electrode. By forming a stack-grown Ta/TaOx switching interface, GOMET preserves interface quality while enabling selector-free individual cell operation. Fabricated devices achieve an on/off ratio exceeding 10^7 and 2-bit operation below 2.2 V. An experiment-calibrated TCAD-to-SPICE DTCO projects a 32-tier implementation with 828 Mb/mm^2 density through co-optimization of geometry, process, electrothermal behavior, and parasitics. System-level evaluation further demonstrates 27.5× higher throughput density than the baseline, highlighting GOMET as a scalable BiCS-compatible 3D embedded memory architecture.

4:10 PM

38-5 First 300mm Wafer Demonstration of Field-Free pSOT-MRAM Writing via Z-Spin with Immunity to Tesla-Level Magnetic Field Disturbances

Mengxi Wang[1], Shengru Chen[1], Yi Xu[1], Yingmei Zhu[1], Xuejian Tang[1], Chao Wang[1], Shiqi Liu[1], Shasha Wang[2], Zhenghui Ji[2], Xingyu Zhu[2], Dinggui Zeng[2], Yihui Sun[2], Guchang Han[1], Tiejun Zhou[1], Shikun He[2], Bo Liu[1]
[1]State Key Laboratory of Spintronics Devices and Technologies, [2]Zhejiang Hikstor Technology Co., Ltd.
For the first time, we demonstrate spin-orbit torque (SOT) driven field-free perpendicular magnetization switching (FFS) using robust z-spin (z-polarized spin current) that exhibits immunity to magnetic field disturbances up to 1 T. The spin polarization direction is determined by the uncompensated moments orientations at the antiferromagnet (AFM)/non-magnet (NM) interface, which can be firmly fixed at the interface after field annealing, leading to high robustness of z-spin against magnetic field disturbances. The observed results are obtained from 100 nm magnetic tunnel junction (MTJ) arrays fabricated on a 300 mm wafer. FFS with a sub-nanosecond (0.3 ns) pulse width is achieved with no performance degradation over a wide range of operating temperatures. Furthermore, introducing the voltage-controlled magnetic anisotropy (VCMA) effect can further lower the switching current density by more than 30%. This work provides a highly reliable, efficient, and robust FFS scheme, offering a feasible pathway toward the commercialization of SOT-MRAM technology.

4:35 PM

38-6 Ultra-Fast (<5 ns) and Energy-Efficient (<80 fJ/bit) VCMA-MRAM Integrating Unified TRNG-PUF for Zero-Trust Security

Xianzhe Chen[1], Jie Yu[1], Xiaohan Huang[1], Haitao Du[2], Zheng Zhu[3], Chin-chung Chen[3], Guanyi Li[1], Chengrui Xie[1], Xinxu Lin[1], Andrew Lee[3], Kaiwen Shen[1], Yifan Gao[1], Yingfen Wei[1], Hao Jiang[1], Xumeng Zhang[1], Ming Wang[1], Di Wu[3], Haozhe Zhu[1], Zhongguang Xu[2], Albert Lee[3], Qi Liu[1], Ming Liu[1]
[1]Fudan University, [2]University of Science and Technology of China (USTC), [3]Suzhou Inston Technology Co., Ltd.
Spin-based hardware root-of-trust primitives are emerging as promising solutions for next-generation security systems, leveraging their low power, high endurance, and non-volatility. In this work, we demonstrate a unified true random number generator (TRNG) and physical unclonable function (PUF) based on a voltage-controlled magnetic anisotropy magnetic random access memory (VCMA-MRAM) array. The unified TRNG-PUF security system selectively configures different blocks within the mini-array for PUF or TRNG operation. In TRNG mode, the VCMA-MRAM array achieves sub-2.5-ns random bit generation with 81.83 fJ/bit energy consumption and >10^{12}-cycle endurance, while passing NIST SP800-90B and SP800-22 randomness evaluations. In PUF mode, the same array achieves a native BER of 1.0% at 25°C and an inter-Hamming distance of 49.99% without additional stabilization. The unified system leverages PUF for secure key storage and enhances hardware encryption security through TRNG with power-consumption of <80 fJ/bit.

5:00 PM

38-7 Mirror-Coupled Field-Effect Transistor for Ultra-Compact SRAM

Xiao-Yu Sun[1], Chen Pan[2], De-he Kong[3], Pengfei Wang[3], Zhenhua Wu[4], Xing-Jian Yangdong[3], Zhiqiang Zeng[3], Weiqi Dang[3], Shi-Jun Liang[3], Feng Miao[3]

[1]Institute of Brain-inspired Intelligence, School of Physics, Nanjing University, [2]Nanjing University of Science and Technology, [3]Nanjing University, [4]Zhejiang University
We report a mirror-coupled field-effect transistor (M-CFET) based on 2D WSe_2 films to breach the postMoore SRAM scaling wall imposed by routing complexity. By completely eliminating conventional independent gates, the M-CFET features vertically stacked dual channels where the source/drain electrodes of each layer mutually serve as the electrostatic split gates for the opposing layer. Exploiting the intrinsic ambipolar transport of WSe_2, a state-interlocking mechanism is implemented, reducing the transistor count of a conventional bistable latch from 4 to 2. Vertical monolithic 3D integration of M-CFETs and Si nFETs demonstrates an ultracompact SRAM cell with stable write-hold characteristics. Benefiting from the state-interlocking mechanism, the SRAM cell completely eliminates complex metal interconnects. This achieves an ultra-compact SRAM cell area of 0.00756 μm^2 under the 28-nm technology node, achieving a 50% area reduction compared to conventional CFET technology.

Session 39: FS | System-Level Optimization Combining Logic, Memory and Packaging

2:00 PM, Continental 6
Co-Chairs: Ali Khakifirooz, Intel and David Chu, TSMC
As physical scaling slows, unlocking next-generation AI and high-performance computing (HPC) performance demands a radical, cross-domain co-design from silicon to rack-scale architecture. In this session, industry and academic experts from IBM, Intel, Nvidia, Marvell, SK Hynix, and UCLA present holistic system-level optimizations designed to break the memory and thermal limits. Presentations will span system-level architectural innovations, integration of optical interconnects for scalable computing, and the critical thermal-mechanical strategies to manage extreme power densities and structural stress in advanced heterogeneous packages.

2:05 PM
39-1 IBM Power's System Architecture: Composable Compute & Memory from Chiplets to Multi-Rack Pods
William Starke
IBM
In this paper, we explore several themes and approaches for optimizing silicon architectures targeting large scale, highly reliable, coherent, shared memory, multi-processor enterprise systems. In an era where semiconductor scaling has slowed, leading to the emergence of chiplet-based silicon architectures; where imbalances between compute elements and memory elements have grown; where specialized compute accelerators have become mainstream; and where composable pools of computational and memory resources blur the definitions of systems verses clusters; there are increasing opportunities for innovative differentiation.

Demonstrated in the context of current and upcoming IBM Power systems, we highlight the maximization of silicon surface, the minimization of latency radius, and the balancing of density, with a strong focus on reliability, efficiency, cost optimization, and flexibility. From leveraging chiplet architectures to refactor socket and system topologies, to partitioning function across silicon boundaries for intra-socket composability, to optimized signaling architectures for multi-socket and multi-system composability, to leveraging hierarchy in the memory architecture to balance compute and memory elements, we show how employing these approaches can result in highly differentiated systems.

2:30 PM
39-2 Memory Optimizations for Datacenter
Rajat Agarwal, Mahesh Natu, Arjun Kripanidhi, Binuraj Ravindran
Intel Corporation
Agentic AI systems execute multi-step plan-act-observe loops in which GPUs accelerate reasoning while the CPU control plane manages orchestration, tool execution, and retrieval state. This shift exposes two distinct memory constraints in the datacenter: a bandwidth bottleneck from hot, frequently reused data such as key-value (KV) cache and CPU-native small-model inference, and a capacity bottleneck from large, often-idle orchestration, sandbox, and retrieval state. No single memory tier satisfies both. We present three optimizations that match memory architecture to each bottleneck: (1) three-dimensional DRAM-on-logic stacking for bandwidth-per-watt scaling; (2) Compute Express Link (CXL)-based memory tiering and pooling, including a hardware-managed Flat Memory Mode

that sustains 97.3% of native DDR5 performance at 24%–39% lower memory cost at iso-capacity; and (3) inline hardware compression that cuts memory footprint by up to 33% with no application changes at iso-throughput.

2:55 PM

39-3 System-Level Co-Optimization of Compute, Memory, and Advanced Packaging for AI GPUs

Mahmut Ersin Sinangil, Sudhir Kudva, David Nellans, Walker J. Turner, C. Thomas Gray
Nvidia

Artificial intelligence (AI) factories are emerging as a new class of accelerated-computing infrastructure whose scaling depends on co-optimizing compute, memory, advanced packaging, power delivery, and cooling. This invited paper examines how workload evolution drives coupled requirements for bandwidth, capacity, locality, energy efficiency, power integrity, and thermal management. Using LLM decode as an example, roughly 91% of energy per token is spent moving data, and package power density directly bounds achievable signaling bandwidth and usable 3D compute power. We argue that future gains require platform-level co-design across memory hierarchy, 2.5D/3D integration, chiplet fabrics, rack-to-die power delivery, and near-package cooling, and outline cross-layer directions for balanced systems from silicon to rack scale.

3:45 PM

39-4 Photonic Interconnect for Scalable Memory in Accelerated Computing

Dan Oh, Ankur Aggarwal, Suresh Pothukuchi, Jelena Pesic, Saket Chadda, Trung Diep
Marvell Semiconductor

The rise of agentic AI, with workloads consuming up to 1,000x more tokens per inference than traditional LLMs, has intensified the Memory Wall in accelerated computing. KV-cache growth rapidly exceeds co-packaged HBM capacity, while CXL-based disaggregated memory over PCIe suffers from copper loss and latency penalties that stall real-time agentic interactions. This paper presents Photonic Fabric, a silicon photonics-based interconnect platform enabling scalable, low-latency disaggregated memory for agentic AI accelerators. Key technologies include PFLink (14.4 Tbps, >50 m reach), OMIB (optical multichip interconnect bridge), and a Photonic Fabric Memory Appliance (PFA) providing >32 TB shared memory at 115 Tbps and ~100 ns RDMA latency. A 56 Gb/s optical link was validated on a 5 nm CMOS test vehicle. System-level simulations predict 60-90% power savings and up to 5.6x throughput improvement over NVIDIA DGX-H100 systems.

4:10 PM

39-5 Addressing Thermal and Mechanical Challenges in Advanced Packaging for High Bandwidth Memory

Jaesik Lee[1], Younghoon Hyun[2], WonHae Kim[1], Yeontaek Hwang[2], Daewoong Lee[2], Sangyong Lee[2], Jongkyu Moon[2], Juheon Yang[2], Kyoungmoo Harr[2], Minseok Jang[2], Jinwoo Choi[1], Heejin Lee[2], Ki-Ill Moon[2], Kangwook Lee[2]
[1]SK Hynix America, [2]SK Hynix

We report package development challenges associated with High Bandwidth Memory 4(HBM4) using a Test Vehicle: achieving 12-layer stacking and investigating stress interaction between HBM and advanced packaging technologies that have been used for Logic and HBM integration. Material and process innovations overcome challenges, such as die warpage and gap fill, caused by decreasing die thickness and reducing gap height. Micro-bump joint integrity for entire 12-layer stacks after reliability confirmed with daisy chain resistance and Confocal Scanning Acoustic Microscopy (C-SAM) analysis. The results indicated that test vehicle achieved void-free and high micro-bump integrity. Package interaction study is conducted with thermo-mechanical simulation methodology using identical floor plan. Higher stress is observed both at the HBM stacks (base die & core die) and at extremely low-k (ELK) layer when assembling on advanced substrate due to high coefficient of thermal expansion (CTE) mismatch. ELK stress at advanced substrate was as high as 2.5x compared to Silicon interposer technology. Finally, we propose a novel thermal solution with integrated cooling element concurrently integrated with DRAM core dies in a HBM, which provides 20-30% junction temperature reduction in a thermal simulation.

4:35 PM

39-6 Advanced Thermal Management for Next-Generation Logic, Memory, and Packaging Integration: Multi-Scale Modeling, Materials, and Cooling Technologies

Tiwei Wei[1], Feifan Xie[1,2], Ye Yang[1], Zhenliang Pan[1], Keyu Wang[3], Zhengwei Chen[3], Akshat Patel[3], Gopinath Sahu[3], Ketan Yogi[3]

[1]University of California, Los Angeles, [2]KU Leuven, [3]Purdue University

Thermal management is a major bottleneck for next-generation heterogeneous integration as chip power densities continue to increase. This work presents a comprehensive thermal management strategy that integrates experimentally validated Cu–diamond thermal vias, Cu/Sn–diamond microbumps, diamond heat spreaders, and single-phase/two-phase direct liquid jet cooling within a multiscale, multiphysics thermal modeling framework. The cumulative impact of these technologies is evaluated for representative backside power delivery network (BSPDN) logic and 3D HBM-on-GPU architectures. For a BSPDN chip with an average power density of 817.9 W/cm^2, the proposed thermal roadmap reduces the peak temperature by 86.74°C, from 144.27°C to 57.53°C. For a 1.61 kW 3D HBM-on-GPU package, the proposed technologies reduce the peak temperature from 120.1°C to 39.4°C under dual-side liquid cooling and from 148.06°C to 74.71°C using top-side liquid jet cooling with bottom-side air jet cooling. These results provide thermal roadmap for future 3D systems through the co-design of advanced thermal materials, heterogeneous packaging, and cooling technologies.

Session 40: MS | Emerging Devices, Ferroelectrics & Nonvolatile Memory

2:00 PM, Continental 7-9

Co-Chairs: Raseong Kim, Intel and Pei-Ying Du, Macronicx

In this session, we have 6 papers on emerging devices, ferroelectrics, and nonvolatile memory. The work by D. Xu et al. (University of Waterloo) presents a comprehensive physics-based compact model for monolithic gallium nitride bi-directional switch devices. In the paper by S. Chen et al. (Southeast University), they propose compact models for SiC devices covering a wide range of temperature from 77 K to 738 K. The work by Y. Hou et al. (The Hong Kong University of Science and Technology) proposes a physics-based compact model for field-free voltage-controlled magnetic tunnel junctions to enable design technology co-optimization flow for probabilistic computing system. The paper by X. Sun et al. (National University of Singapore) proposes the oxide semiconductor dual-gate polarization-boosted charge trapping flash to overcome the long-standing erase limitation. In the work by G. Kim et al. (University of Notre Dame), they develop a unified Monte Carlo-based Landau-Khalatnikov framework that captures polarization switching across different ferroelectric classes. In the paper by Hao-Yu Lu et al. (National Cheng Kung University), they demonstrate a nanoscale single-transistor non-volatile ternary content-addressable memory cell based on the ferroelectric independent-double-gate FinFET.

2:05 PM

40-1 MVSG_BDS: A Physics-based Compact Model for Monolithic GaN Bi-directional Switches

Daiyao Xu[1], Rebecca Fang[1], Johan Alant[1], Kaiman Chan[2], Tim Merkin[2], Ujwal Radhakrishna[2], Lan Wei[1]

[1]University of Waterloo, [2]Texas Instruments

This paper presents a comprehensive physics-based compact model for monolithic gallium nitride bi-directional switch (MBDS) devices. The model is developed using a modular methodology, enabling flexible adaptation to both common-drain (CD) and common-source (CS) MBDS configurations, and customization for different field-plate designs. With proven computational symmetry and robustness, the model correctly describes the channel current and gate leakage under different modes (e.g., diode mode and FET mode) of MBDS operations. Physical effects critical for MBDS (such as substrate effects, trapping effects, temperature dependence, multi-terminal electrostatic coupling) are included. A systematic MBDS-specific parameter extraction is also developed with multi-temperature calibration. Model validation against industry CD and CS MBDS devices, design of cyclo-inverter, and technology benchmarking against uni-directional GaN are shown, validating the accuracy, effectiveness and usefulness of the proposed model. The model has been adopted for industry process design kit (PDK) development.

2:30 PM

40-2 From 77K to 738K: Compact Modeling of SiC Integrated Devices over a Wide Temperature Range towards Power ICs in Aerospace Application

Shuqiang Chen, Yong Gu, Mengyao Zhao, Zhenya Wang, Yiwen Hou, Aimiao Liu, Li Ding, Sheng Li, Jiaxing Wei, Wangran Wu, Weifeng Sun, Jie Ma, Long Zhang, Siyang Liu
Southeast University

Compact models of SiC CMOS and high-voltage LDMOS are developed over 77-738 K for aerospace power IC design. A 300 K core model is extended by incorporating wide-temperature resistance with field-assisted ionization, threshold-voltage shift, temperature-dependent interface-trap charge occupation, and interface-trap-induced Coulomb-scattering mobility. The resistance model captures carrier freeze-out and high-field drain-current recovery at cryogenic temperatures, while the coupled threshold-voltage, interface-trap, and mobility models reproduce the nonmonotonic current variation at high temperatures. The proposed models are validated against measurements of 20 V SiC CMOS and 800 V SiC LDMOS devices over the full temperature range. Circuit-level validation further demonstrates accurate prediction of inverter switching-point shifts, two-stage amplifier gain variation, and CMOS-driven LDMOS output response. The models provide a robust foundation for reliable SiC integrated-circuit design under extreme-temperature aerospace environments.

2:55 PM

40-3 Monte-Carlo Modeling and Experimental Characterization of Hybrid Dynamics for Field-Free Voltage-Controlled Magnetic Tunnel Junctions

Yaoru Hou[1,2], Yantong Di[3], Xuezhao Wu[4], Hongjie Yin[3], Haoran Du[3], Zeying Ding[3], Yuesheng Li[4], Di Wu[4], Jiayong Wang[4], Zheng Zhu[4], Zhihua Xiao[1,2], Hao Cai[3], Qiming Shao[1,2]
[1]Hong Kong University of Science and Technology (HKUST), [2]AI Chip Center for Emerging Smart Systems, [3]Southeast University, [4]Shanghai University

Voltage-controlled magnetic tunnel junctions (VC-MTJs), exhibiting hybrid dynamics of steady-state and probabilistic switching, are promising for combinatorial optimization solvers. To bridge the gap between device physics and large-scale circuit system, we propose a physics-based compact model for field-free VC-MTJs to enable design technology co-optimization (DTCO) flow for probabilistic computing system. By employing a Néel-Brown model, we explicitly map process-induced multidirectional stray fields to dynamic switching perturbations validated by sLLG simulation. Calibrated via experimental measurements from VC-MTJ array, the proposed model captures hybrid dynamics of VC-MTJ and supports large-scale Monte Carlo simulations, achieving high coverage of device-to-device variations. Furthermore, integrating this Verilog-A model into a 28-nm CMOS PDK, we validate DTCO flow to design a probabilistic computing system featuring a 256-kb high-density VC-MTJ p-bit array. Crucially, our model speeds up post-layout simulation runtime in the Cadence Virtuoso by 65× compared to sLLG-based models.

3:45 PM

40-4 First Demonstration of Oxide-Semiconductor Dual-Gate Polarization-Boosted Charge-Trapping Flash: Comprehensive Modeling and Experimental Validation

Xiaoqing Sun[1], Xinming Li[1], Dong Zhang[1], Kai Ni[2], Chen Sun[1], Xiao Gong[1]
[1]National University of Singapore, [2]University of Notre Dame

We report the first oxide-semiconductor (OS) dual-gate (DG) polarization-boosted charge-trapping flash (PB-CTF), which overcomes the erase limitation of unipolar OS channels through a novel channel-clamping mechanism. We establish a comprehensive dynamic model that self-consistently couples multiphase ZrO_2 antiferroelectric polarization switching, electron tunneling, and charge trapping/detrapping. The model captures the coupled polarization and charge evolution during program/erase operations, predicts trap occupancy under varying biases, and provides physics-based design guidelines. Guided by the model, we experimentally demonstrate a 10.3-V clockwise memory window under ±13-V operation, with 6.9 V maintained after 10^5 cycles and projected 10-year retention. This combined modeling and experimental study establishes a comprehensive methodology for high-performance, BEOL-compatible nonvolatile memories toward future 3D NAND and monolithic 3D integration.

4:10 PM

40-5 Monte Carlo-based Unified Ferroelectric Model for Fluorite & Wurtzite, and Their Proximity Coupling Effect

GIUK KIM[1,2], Erdem Ozdemir[3], Yixin Qin[1], Sizhe Ma[1], Ian Mercer[3], Clive Randall[3], Won-Tae Koo[4], Hyun-Soo Kim[4], Vijaykrishnan Narayanan[3], Susan Trolier-McKinstry[3], Kai Ni[1]

[1]University of Notre Dame, [2]University of Seoul, [3]Pennsylvania State University, [4]SK Hynix

Fluorite, perovskite, and wurtzite ferroelectrics (FEs) exhibit distinct polarization switching characteristics, but existing models remain material-specific, limiting unified analysis of heterogeneous FE stacks. Here, for the first time, we develop a unified Monte Carlo-based Landau-Khalatnikov (MC-LK) framework that captures stochastic and time-dependent P switching across FE classes. Using domain-resolved LK energy barriers and coercive-field distributions, the model is validated against measurements of fluorite HZO and wurtzite AlBN. We further extend the framework to wurtzite/fluorite bilayers and reveal proximity-coupled mutual switching reinforcement between FE layers. When incorporated into a FeFET model, this interaction combines a lower operation voltage than a wurtzite-only stack with a larger coercive field and memory window than a fluorite-only stack. Device-level modeling predicts an MW of ~9 V with a flash-compatible gate-stack thickness of ≤ 20 nm at a moderate operation voltage of 11.5 V, bridging material-level switching physics and device-level memory design.

4:35 PM

40-6 Novel 3D-Scalable Single-Transistor Non-Volatile TCAM with Ferroelectric Independent-Double-Gate FinFET

Hao-Yu Lu[1], Chen-Hung Yang[1], Hong-Min Lee[1], Shang-Lun Chung[1], Shih-Hsuan Hsu[1], Yong-Jyun Yang[1], Alptekin Vardar[2], Chao-Hung Wang[1], Yao-Jen Lee[3], Yeong-Her Wang[1], Thomas Kämpfe[2], Darsen Lu[1]

[1]National Cheng Kung University, [2]Fraunhofer Institute for Photonic Microsystems, [3]National Yang Ming Chiao Tung University

We demonstrate a nanoscale single-transistor non-volatile ternary content-addressable memory (nvTCAM) cell based on the ferroelectric independent-double-gate FinFET (FeIGFinFET), with two independently switchable storage gates on the two sidewalls of one shared fin. TCAD simulation shows that the two gates are programmed independently when fin width is about 30 nm or larger, enabling ternary storage within a single device. Furthermore, with symmetric but separated gates, back-gate biasing easily suppresses leakage, reduces read disturbance, and enables flexible Vth tuning to achieve a mismatch/match current ratio of 10^7 and beyond, avoiding normal sub-threshold current and gate-induced drain leakage at the same time. A cross-linked BSIM-IMG compact model is calibrated to measured FeIGFinFET data. SPICE simulation is performed to investigate both the NOR-type TCAM and a novel NAND-type TCAM, which is extendible to a highly scalable 3D-NAND architecture.

Session 41: RSD | Reliability of Next-Generation Transistors

2:00 PM, Imperial A

Co-Chairs: Mario Lanza, National University of Singapore and Stephen Ramey, Intel

This session collectively explores the reliability and advanced characterization of next-generation semiconductor transistors and technology, focusing on understanding degradation mechanisms to ensure robust device performance for future electronic applications. The session begins with two papers on monolayer WSe_2 p-channel FETs that establish critical reliability benchmarks, with one performing the first BTI investigation on highly scaled 300 mm-integrated devices to define a safe operating window, and the other conducting the first on-the-fly BTI study to quantify fast and slow charge trapping in devices with sub-1 nm dielectrics. This is followed by an investigation into cryogenic InGaAs/InP high-electron-mobility transistors (HEMTs) for quantum computing that finds that while bias temperature instability (BTI) is suppressed at 4K, hot-carrier degradation remains a significant concern. Next, a paper on GaN HEMTs utilizes two-dimensional thermal mapping to experimentally visualize device-level hotspots, providing crucial spatial data for reliability-aware electrothermal design. This is followed by a study on ferroelectric transistors (FeFETs) that reveals a paradoxical memory window widening phenomenon driven by soft breakdown, challenging conventional fatigue limits. Finally, the session wraps up with an invited paper covering the reliability of heterogeneous 3D packaging integration.

2:05 PM

41-1 First On-the-Fly BTI Study of 1L-WSe_2 Dual-Gate pFETs with Sub-1 nm EOT toward Reliable Low-V_{min} Advanced Transistors

Yi-Cheng Lo[1], Rittik Ghosh[2], Yu-Wei Hsu[3,4], Tzu-Chuan Su[3], Nien-En Chiang[4], Ting-Hua Wei[3], Ruei-Yu Chen[3], Hung-Li Chiang[3], Wei-Ning Chang[5], Chao-Cheng Lin[5], Yu-Sheng Lai[5], Chenming Hu[3], Chih-I Wu[4], Theresia Knobloch[2], Tibor Grasser[2], Tsung-En Lee[3]

[1]College of Electrical and Computer Engineering, National Yang Ming Chiao Tung University, [2]TU Wien, [3]National Yang Ming Chiao Tung University, [4]National Taiwan University, [5]Taiwan Semiconductor Research Institute (TSRI)

Future advanced 2D CMOS architectures, including nanosheet and monolithic CFET transistors, require BEOL-compatible all-ALD gate dielectrics with CET under 0.9nm (EOT of below ~0.5nm), reduced gate-stack height, and excellent reliability. However, the reliability of aggressively scaled 2D pFET gate stacks remains largely unexplored. Here, we present the first On-The-Fly (OTF) bias-temperature instability (BTI) study of monolayer (1L) WSe_2 dual-gate (DG) pFETs with sub-1nm EOT and channel lengths (L_{CH}) ranging from 1µm to 100nm. By combining sweep-rate hysteresis, measure–stress–measure (MSM), and 100-µs OTF sensing, together with non-radiative multi-phonon (NMP) modeling and effective single-defect decomposition (ESiD), fast and slow trapping mechanisms are quantitatively identified. The devices exhibit small hysteresis over a wide sweep-rate range, NBTI operation up to 125°C, threshold-voltage shifts below 40mV under worst-case NBTI stress (10 MV/cm, 1ks, R.T.), and competitive hot-carrier reliability. These results establish a BEOL-compatible sub-1nm 2D gate dielectric platform for reliable low-V_{min} operation in advanced transistors.

2:30 PM

41-2 First BTI Investigation of 300 mm-Integrated 50 nm-CGP Monolayer WSe_2-Channel pFET Devices

Marwane Ait Ouaicha[1], Tien Dat Ngo[1], Goutham Arutchelvan[2], Tom Schram[1], Quentin Smets[1], Luca Panarella[1], Ben Kaczer[1], D. Heh[2], Ming-Yang Li[2], C.-H Yao[3], Vina Faramarzi[3], Mark Van Dal[2], Etienne De Poortere[3], Iuliana Radu[2], Cesar Javier Lockhart de la Rosa[1], Nadine Collaert[1], Bogdan Govoreanu[1]

[1]imec, [2]Taiwan Semiconductor Manufacturing Company (TSMC), [3]ASML

We report the first BTI assessment of 300 mm, EUV-patterned monolayer WSe_2 pFETs scaled to 50 nm contacted gate pitch. The devices integrate bottom W-based S/D contacts with $Al_2O_3/HfO_2/TiN$ topgate-stack, featuring ~1.9nm CET and D_{it} of ~5×10^{11} $cm^{-2}eV^{-1}$. We underscore maximum overdrive voltage, $|V_{ov}|$, of 1.1V/25°C and 0.41V/125°C for extrapolated lifetime of 10 years - a first for fully manufacturable 2D-channel technology. NBTI stress induces negative V_{th} shift, identifying hole trapping in gate stack as the dominant mechanism. The instability follows two distinct regimes: at low/moderate stress biases, below ~|−3 V|, DV_{th} displays near-complete recovery, consistent with fast border-trap charging/release. At stress biases above this threshold, recovery becomes incomplete over the monitored time, revealing additional bulk high-κ charge trapping and interface-state generation. This process accentuates with increasing temperature. These results establish viable safe-bias window for scaled-CGP WSe_2 pFETs and identify BTI as a key reliability validator for manufacturable 2D-channel technologies.

2:55 PM

41-3 Cryogenic BTI and HCD in low-noise InGaAs/InP HEMTs for quantum computing readout

Cezar Zota[1], Ludovica Cosi[1], Eunjung Cha[2], Siegfried Karg[1], Marilyne Sousa[1]

[1]IBM, [2]PSI

In this work, hot carrier degradation (HCD) and bias temperature instability (BTI) in cryogenic low-noise III-V HEMTs at 4 K is studied for the first time. Cryogenic III-V HEMTs are key components in low-noise amplifiers (LNAs) at the readout stage in quantum computers. Degradation of such components is an important issue that impacts both readout performance and system reliability. We find that, at 4 K, BTI responses freeze out, while HCD remains significant. Fixed-bias HCD primarily causes ΔV_T degradation through fast and slow trapping kinetics. Cumulative bias-ladder HCD instead causes degradation of all performance metrics, indicating an important role of full bias trajectories. We also compare quantum wells with increasing barrier thickness and show that thicker barriers provide significantly improved reliability, e.g. from -15% to -3% $g_{m,peak}$ degradationt. Overall, we present the first report on cryogenic degradation and mitigation strategies in low-noise HEMTs, key components in many quantum computer architectures.

3:45 PM

41-4 Two-Dimensional Thermal Mapping of GaN HEMTs for Reliability-Aware Device Design

Chae Yeon Han[1], Chan Jik Lee[1], Sun Woo An[1], Nahyun Rheem[1], Yoon-Je Suh[1], Bertrand Parvais[2], Sachin Yadav[2], Alireza Alian[2], Uthayasankaran Peralagu[2], Hao Yu[2], Ying-Chun Kuo[2], Lisa Tondelli[2], Arturo Sibaja-Hernandez[2], Sanghyeon Kim[1]

[1]Korea Advanced Institute of Science and Technology (KAIST), [2]imec

Reliability-aware electrothermal design-technology co- optimization of GaN HEMTs requires compact-model inputs that capture localized thermal stress beyond spatially averaged temperature. However, conventional transistor thermometry cannot sufficiently probe two- or even higher-dimensional heat distribution. Here, frontside/backside thermoreflectance microscopy (TRM) visualizes two-dimensional temperature distributions in GaN HEMTs fabricated on imec's CMOS-compatible 200-mm GaN-on-Si RF platform. The measured TRM maps experimentally resolve device-level heat distributions previously inferred mainly from averaged thermometry or simulations, including hotspot localization, field-plate-dependent hotspot migration, and pre-breakdown thermal signatures correlated with failure-prone regions. By resolving where local thermal stress forms and evolves under device- and operation-dependent conditions, this TRM-based approach provides spatial reliability inputs for compact modeling, TCAD calibration, and reliability-aware GaN HEMT design.

4:10 PM

41-5 Beyond Fatigue: Paradoxical Memory Window Widening in MFMIS Oxide-Semiconductor FeFETs Driven by Ferroelectric Soft Breakdown and Channel Depletion Barrier

Byeongjun Bae[1,2], Seongwon Choi[1], Jaemin Yeom[1], Changhyeon Park[1], Jung-Kyun Kim[3], Younjin Jang[3], Hyerim Hong[3], Na-Rae Han[3], Kyooho Jung[3], Jee-Eun Yang[3], Sijung Yoo[3], Seung-Geol Nam[3], Duk-Hyun Choe[3], Sangwook Kim[3], Woo Young Choi[1]

[1]Seoul National University, [2]Device Solutions Division, Samsung Electronics, [3]Samsung Advanced Institute of Technology (SAIT)

While ferroelectric (FE) fatigue conventionally limits the endurance of FeFETs, this work achieves paradoxical memory window (MW) widening in MFMIS oxide-semiconductor FeFETs (OS-FeFETs) after fatigue. We reveal that this anomalous behavior is driven by the combined effects of FE soft breakdown (SBD) and OS depletion barrier. FE SBD induces charge injection into the floating gate. Within the MFMIS OS architecture, the work function difference aligning the FG Fermi level below the OS depletion-barrier peak and the lack of holes prevent both electron emission and charge neutralization. This enables stable charge storage that drives MW widening. To establish design guidelines for MW widening, we developed a current-balancing model that defines the boundary where the Fowler-Nordheim tunneling current matches the FE SBD leakage current, confirming the fabricated device operates within the MW widening regime. These findings extend beyond traditional endurance limits, providing a framework to design high-endurance OS-FeFETs for next-generation non-volatile memory.

4:35 PM

41-6 Reliability Characterization of Next-Generation 3D Heterogeneous Integrated Systems

SyuFong Li

APTQR-1, TSMC

This work introduces TSMC's next-generation System-on-Integrated-Chips (SoIC) 3D stacking technology, focusing on comprehensive reliability qualifications. By leveraging TSMC's cutting-edge System-on-Chip (SoC) process nodes and advanced packaging solutions, this innovative 3D technology extends Moore's Law to meet the rigorous demands of next-generation High-Performance Computing (HPC) applications. We demonstrated the robust process and package reliability of the SoIC structure, confirming its readiness for high-volume manufacturing

Session 42: SMB | Smart Sensors and Bio-Integrated Interfaces

2:00 PM, Imperial B

Co-Chairs: Wan-Thai Hsu, Soundskrit and Tsuyoshi Minami, University of Tokyo

This session covers recent developments in smart sensors and bio-integrated interfaces, highlighting advancements in flexible materials, neuromorphic hardware, and multidimensional imaging. The first part focuses on Smart Sensors, including self-powered soft X-ray detection, multidimensional polarization sensing, and ultrathin graphene sensors for continuous plant monitoring. The second part addresses Bio-Integrated Interfaces, with

presentations on piezoelectric active-matrix sensing for continuous blood-pressure monitoring, multilayer intracortical probes for neural recording, and soft-polymer ECRAMs for in-sensor neuromorphic computing. Together, these papers present advances in sensing devices and their integration into monitoring and bioelectronic systems.

2:05 PM

42-1 Dual-Mode AlGaN/GaN Double-Channel Soft X-Ray Detector With Back-to-Back Schottky Contacts and an Interelectrode Trench for Self-Powered Monitoring and High-Contrast Imaging

Xiuyi Bian, Shunjie Yu, Haoyu Wang, Wenjie Zhang, Binglin Li, Menghui He, Shibing Long, Shu Yang
University of Science and Technology of China (USTC)

A dual-mode AlGaN/GaN double-channel (DC) soft X-ray detector incorporating back-to-back Schottky contacts and an interelectrode trench is demonstrated. At zero bias, carrier separation assisted by the Schottky depletion field enables self-powered detection with a responsivity (R) of 4.5×10^{-2} A/W, a photo-to-dark current ratio ($PDCR$) of 3.1×10^{2}, and rapid rise and decay times (τ_r and τ_d) of 12 and 8 ms, respectively. Under 550-eV irradiation at 20 V, the R and $PDCR$ are increased to 1.8×10^{2} A/W and 2.6×10^{5}, respectively. This enhancement is attributed to field-assisted carrier transport and photoinduced lowering of the back barrier beneath the trench and the Schottky barrier. Bias-tunable dual-mode operation is further demonstrated in a soft X-ray imaging system, enabling continuous self-powered monitoring at zero bias followed by high-gain, high-contrast imaging under applied bias once a signal is detected. This imaging demonstration validates the potential of the proposed operating scheme for dynamic cellular imaging.

2:30 PM

42-2 An Amphibious Multidimensional Polarization Photosensor for 360° Omnidirectional Underwater Full-Stokes Imaging in Turbid Water Environment

Wei Chen, Menglong Wang, Chun Zheng, Yang Li, Yuying Liu, Yuanmin Luo, Zhixiang Gao, Haiding Sun
University of Science and Technology of China (USTC)

We report the first encapsulation-free amphibious polarization photosensor capable of self-powered 360° omnidirectional full-Stokes sensing and enhanced imaging in turbid water. The device integrates double-sided InGaN nanowire-on-Si with a chiral polyvinyl alcohol-cellulose nanocrystals (PVA-CNCs) quasi-solid hydrogel. The band structure of InGaN is intentionally tailored to the 450 nm effective underwater light transmission, delivering a competitive self-powered responsivity of 59.75 mA/W with a response time of 10 μs, while the PVA-CNCs provide stable operation underwater and selective linear/circular polarization filtering, enabling intrinsic underwater full-Stokes polarization sensing (linear-polarization ratio of 1.8; circular asymmetry factor of -0.21). Double-sided nanowire integration extends these capabilities to 360° omnidirectional underwater coverage. Device arrays further enable high-contrast full-Stokes imaging and scattering-suppressed polarization imaging in turbid underwater scenes, providing improved target visibility over typically intensity-only imaging. This work establishes a compact platform for robust full-Stokes imaging underwater.

2:55 PM

42-3 Ultrathin Transparent Graphene Leaf Sensor for Continuous Physiological Monitoring of Plants

Ning Liu[1], Utkarsh Misra[2], Jean Anne C. Incorvia[2], Deji Akinwande[2]
[1]Department of Electrical and Computer Engineering, University of Texas at Austin, [2]University of Texas at Austin

This work presents the first leaf-integrated device that directly and continuously monitors leaf water status, circadian rhythms, and midday stomatal closure without disrupting normal leaf physiology, rather than indirectly inferring these processes from the surrounding microenvironment. The ultrathin, transparent, breathable, and stretchable graphene leaf sensor (GLS) uses the leaf as an active sensing-circuit component to directly detect its physiological dynamics. Its figure-8 ring geometry and bilayer graphene electrodes conform to diverse leaf surfaces and accommodate growth while maintaining measurement stability. Graphene's high optical transparency and a breathable micro-hole architecture minimize interference with light transmission, gas exchange, and transpiration, preserving photosynthesis and respiration during long-term operation. Stable recordings over ten days captured gradual and transient physiological dynamics under changing environmental conditions. The platform also supports leaf-gated conductance modulation, demonstrating potential for integrated physiological sensing and bio-interfaced computing.

3:45 PM

42-4 Heterogeneously Integrated 1T1C Piezoelectric-DGTFT Active Matrix for Continuous BP Monitoring

Delang Lin[1,1], Rongsheng Chen[1], Wei Huang[2], Xiaocheng Wang[2], Man Chun Tseng[1], Fion Sze Yan Yeung[1], Hoi Sing Kwok[1]

[1]Hong Kong University of Science and Technology (HKUST), [2]South China University of Technology

This work presents a heterogeneously integrated wearable 8×8 1T1C active matrix for continuous blood pressure monitoring. Each pixel integrated a piezoelectric capacitor with a three dimensional channel InSnZnO dual-gate thin film transistor. The dual-gate functional multiplexing enabled a single transistor to perform both signal amplification and matrix addressing, thereby simplifying the sensing pixel architecture. The array exhibited uniform electrical characteristics, bias tunable pressure sensitivity, and stable long term pulse wave acquisition. A model pretrained on the MIMIC-III-PPG dataset achieved SBP/DBP mean absolute errors of 3.84/1.97 mmHg. After adaptation to pulse waves acquired by the proposed active matrix and deployment on an ESP32-S3, the system achieved SBP/DBP errors of -0.06±4.60/-0.05±2.67 mmHg and enabled MCU-based real-time continuous BP tracking.

4:10 PM

42-5 A Multilayer Intracortical Probe with Polymer-Only Recording Sites and Interconnects

Qian Wan[1], Bingchen Zhao[1], Lexuan Yang[1], Xiaofeng Luo[1], Wei Ma[2], Yilin Wang[2], Zhihong Li[1,1], Zhitong Zhang[3], Yu-qing Zheng[1,1]

[1]Peking University, [2]Xi'an Jiaotong University, [3]Zhongguancun Academy

High-density intracortical probes are essential for brain-machine interfaces and systems neuroscience, yet many conventional silicon/metal probes remain mechanically rigid. Here, we report a high-density all-polymer intracortical probe with a metal-free implanted region, enabled by poly(ionic liquid)-assisted locking and reorganization **(**PILLAR**)**-PEDOT conductors and wafer-level multilayer polymer microfabrication. PILLAR-PEDOT achieves a conductivity of 2,979 S cm^{-1} after post-treatment and supports micrometer-scale patterning, multilayer routing, and polymer encapsulation without metallic conductors in the implanted region. The process defines 30-μm recording sites and 25/15-μm line/space interconnects and enables a three-layer probe with 30 implanted recording sites and an areal density above 285 sites mm^{-2}. Separate 16-channel probes implanted in mouse cortex recorded multichannel LFPs and sortable single-unit activity. Electrode impedance remained within a comparable range, and sortable single-unit activity was recorded over 10 weeks. These results establish a scalable all-polymer platform for high-density intracortical interfaces with polymer-only implanted conductors.

4:35 PM

42-6 First Sub-Attojoule All Analog Inference Demonstration Using Biocompatible Soft Polymer ECRAMs for Neuromorphic Computing

Philip Varkey[1], Utkarsh Misra[2], Nicholas Zogbi[2], Jack Guarnery[2], Samuel Liu[3,4], Dmitry Kireev[5], Deji Akinwande[2], Jean Anne Incorvia[2]

[1]Microelectronics Research Center, University of Texas at Austin, [2]University of Texas at Austin, [3]National Institute of Standards and Technology, [4]University of Colorado Boulder, [5]University of Massachusetts Amherst

Bio-integrated intelligence requires computing hardware that is soft, energy-efficient, and biologically compatible. Here, we demonstrate an individually addressable 4 × 1 array of graphene–Nafion electrochemical random-access memory (ECRAM) devices for ultralow-energy analog inference. The soft devices exhibit bi-directional analog conductance modulation, conductance-update linearities of 0.716/0.833 for potentiation/depression, retention exceeding 500 s, and stable operation over $5 * 10^4$ cycles. Ion-mediated modulation at the graphene–Nafion interface enables an area-normalized write energy as low as 192 zJ/$μm^2$ and a device-level read energy of 127 nJ. The devices were independently programmed as analog weights in a four-input linear classifier, where voltage-encoded inputs and direct current summation performed the multiply–accumulate operation in hardware. The measured array achieved over 86% accuracy on a banknote-authentication task, with a mean deviation of 85.2 μA from the expected output current. These results advance biocompatible soft ECRAMs from individual memory elements to functional analog-computing hardware for future bio-integrated intelligence.

IEDM Committee List

Executive Committee

General Chair
Meng-Fan (Marvin) Chang
National Tsing Hua University &
TSMC

Technical Program Chair
Jungwoo Joh
Texas Instruments

Technical Program Vice Chair
Rihito Kuroda
Tohoku University

Publications Chair
Srabanti Chowdhury
Stanford University

Publications Co-Chair
Gaudenzio Meneghesso
University of Padova

Publicity Chair
John Paul Strachan
Forschungszentrum Jülich

Publicity Co-Chair
Sandy Liao
TSMC

Courses Chair
Kang-ill Seo
IBM

Courses Co-Chair (Short Courses)
Geert Eneman
imec

Courses Co-Chair (Tutorials)
Uygar Avci
Intel

Focus Session and Special Events Chair
Olga Blum
Sandia National Laboratories

Focus Session & Special Events Co-Chair
Dechao Guo
IBM

Virtual Arrangements Chair
Elena Gnani
University of Bologna

Asian Arrangements Chair
Vita Pi-Ho Hu
National Taiwan University

Asian Arrangements Co-Chair
Sangbum Kim
Seoul National University

European Arrangements Chair
Michael Waltl
TU Wien

European Arrangements Co-Chair
Daniele Ielmini
Politecnico di Milano

Treasurer
Jan Hoentschel
GLOBALFOUNDRIES

Advanced Logic Technology

Chung-Hsun Lin, *Chair*
Intel

Mocuta Anda
Micron

Vincent Chang
TSMC

Gabriela Dilliway
Qorvo

Naoto Horiguchi
imec

Lin Hu
Nvidia

Wookhyun Kwon
Samsung

Kwan-Yong Lim
Qualcomm

Tetsu Ohtou
Tokyo Electron

Yusuke Oniki
Rapidus

Shay Rebh
IBM

Changhwan Shin
Korea University

Emerging Device and Compute Technology

Maud Vinet, *Chair*
Quobly

Benoit Bertrand
CEA-Leti

Suraj Cheema
Massachusetts Institute of Technology

Ching-Tzu Chen
IBM

Albert Cheng
TSMC

Saptarshi Das
Penn State

Michelly De Souza
Centro Universitario FEI

Asif Khan
Georgia Institute of Technology

Soo Gil Kim
SK Hynix

Suhwan Lim
Samsung

Lian-Mao Peng
Peking University

Gaurav Thareja
Applied Materials & Stanford

Jens Trommer
Namlab

Memory Technology

Johannes Mueller, *Chair*
GlobalFoundries

Laura Bégon-Lours
ETH Zurich

Attilio Belmonte
imec

Mattia Boniardi
ST Microelectronics

Katherine Chiang
TSMC

Shoichi Kabuyanagi
Kioxia

Kamal Karda
Micron

Ali Khakifirooz
Intel

Sun-Ghil Lee
Tokyo Electron

Kyunghwan Lee
Samsung

Kai-Shin Li
National Tsing Hua University (NTHU)

Tomoya Saito
Renesas

Kim Tae Kyun
SK Hynix

Yanli Zhang
Infineon

Modeling and Simulation

Dongyean Oh, *Chair*
SK Hynix

Aryan Afzalian
imec

Pei-Ying Du
Macronicx

Albert Fayrushin
Micron

Gianluca Fiori
University of Pisa

Mohammad Hasanuzzaman
IBM

Changwook Jeong
UNIST

Raseong Kim
Intel

Gengchiau Liang
NYCU

Xi-Wei Lin
Synopsys

Yan Zhang
Massachusetts Institute of Technology Lincoln Lab

Neuromorphic and Novel Computing

Seyoung Kim, *Chair*
POSTECH

Valeria Bragaglia
IBM Zurich

Erika Covi
Technical University of Munich

Jun Deguchi
Kioxia

Tuo-Hung (Alex) Hou
National Yang Ming Chiao Tung University / TSRI

Mingu Kang
University of California, San Diego

Ravi Kumar
Infineon

Ming-Hsiu (Eric) Lee
Macronix

Stephan Menzel
Forschungszentrum Jülich

Xin Si
Southeast University of China

Dmitri Strukov
University of California, Santa Barbara

Qiangfei Xia
University of Massachusetts Amherst

Optoelectronics, Displays, and Imaging Systems

Geunsook Park, *Chair*
Apple

Michael Cich
Google

Gauri V. Karve
imec

Jiwon Lee
POSTECH

Stefan Lischke
IHP microelectronics

Pierre Malinge
STMicroelectronics

Minseok Oh
Anduril Industries

Chao Shen
Fudan University

Haiding Sun
University of Science and Technology of China

Yoshiaki Tashiro
Sony

Masayuki Uchiyama
Omnivision

Daniel Van Blerkom
SWIRLabs

Keita Yasutomi
Shizuoka University

Power, Microwave/Mm-Wave and Analog Devices/Systems

Zhikai Tang, *Chair*
Texas Instruments

Shamima Afroz
Northrop Grumman

Marina Antoniou
University of Warwick

Julien Buckley
CEA-Leti

Nicholas Dellas
Infineon

Donald Gajewski
Wolfspeed

Masataka Higashiwaki
Osaka Metropilitan University

Jose Jimenez
Qorvo

Christian Koller
Infineon

Hann-Hue Tsai
National Applied Research Laboratories (NARLabs), Taiwan

Tetsuzo Ueda
Kansai University

Jin Wei
Peking University

Yuhao Zhang
University of Hongkong

Reliability of Systems and Devices

Huimei Zhou, *Chair*
IBM

David Chu
TSMC

Grace Chung
Samsung

Christine Hau-Riege
Marvell

Reika Ichihara
Kioxia

Mario Lanza
University of Singapore

Gabriele Navarro
CEA-Leti

Kai Ni
University of Notre Dame

Kenji Okada
Rapidus

Stephen Ramey
Intel

Sachin Sapatnekar
University of Minnesota

Kechao Tang
Peking University

Tianhao Zhang
Synopsis

Sensors, MEMS, and Bioelectronics

Roozbeh Tabrizian, *Chair*
University of Florida

Taeko Ando
Ritsumeikan University

Azadeh Ansari
Georgia Institute of Technology

Humberto Campanella
Silicon Austria Labs GmbH

Savannah Elsner
Columbia University

Sophie Giroud
CEA-Leti

Wan-Thai Hsu
Soundskrit

Ming-Huang Li
National Tsing Hua University

Tsuyoshi Minami
University of Tokyo

Gayathri Pillai
Indian Institute of Science, Begaluru

Jeronimo Segovia
Texas Instruments