2026 VLSI SYMPOSIUM

IEEE/JSAP Symposium on VLSI Technology & Circuits

ADVANCING THE AI FRONTIER through VLSI innovation

2026 Technical Program

June 14-18, 2026
Hilton Hawaiian Village
Honolulu, HI
www.vlsisymposium.org

As of May 17, 2026

IEEE SSCS JSAP Electron Devices Society EIE

This information is current as of April 27, 2026. For more up to date information, please log onto MYS https://vlsi26.mapyourshow.com/8_0/sessions/#/

Contents

Sunday, June 14, 2026

Workshops | 1:00 PM – 5:30 PM

W1 | Advances in Cryo-CMOS: Devices, Circuits and Applications

1:00 PM – 3:15 PM, Tapa 1
Organizers: Quentin Schmidt, TU Delft and Fabio Sebastiano, TU Delft

Near-absolute-zero environments unlock new devices that leverage quantum effects while drastically boosting performance, e.g., by reducing thermal noise, thus enabling performance beyond the intrinsic limits of current room-temperature electronics.

To fully exploit the potential of cryogenic devices and sensors, integrated electronics operating at cryogenic temperatures (cryo-CMOS) is required. However, cryogenic circuit design remains challenging, as transistor behavior at deep-cryogenic temperatures is not yet fully understood or captured in device models.

This workshop highlights recent advances in cryogenic transistor modelling, along with state-of-the-art Cryo-CMOS applications in ultra-sensitive sensing for gravitational-wave detection, high-speed ultra-low-power classical computation using superconducting devices, and spin-based and superconducting-based quantum computing.

- 1:10 PM – W1-1 | **Cryogenic CMOS Devices - Performance, Variability and Defects**, Alexander Gill, imec
- 1:35 PM– W1-2 | **Cryo-CMOS for Gravitational Wave Detection**, Filip Tavernier, KU Leuven
- 2:00 PM– W1-3 | **Hybrid Cryo-CMOS and Josephson Digital Logic for Energy-Efficient High-Performance Computing**, Nobuyuki Yoshikawa, Yokohama National University
- 2:25 PM– W1-4 | **Cryo-CMOS Circuits for the Control and Readout of Silicon Qubits**, Franck Badets, Cea-Leti
- 2:50 PM– W1-5 | **Cryo-CMOS Designs for Superconducting Qubit-Based Quantum Computers: Challenges and Solutions**, Daniel Friedman, IBM

W2 | Embedded Memories in the Sub-2nm Era: SRAM Scaling Perspectives, Alternatives, and 3D Futures

1:00 PM – 3:15 PM, Tapa 2
Organizers: Andreas Burg, EPFL and Jaydeep Kulkarni, UT Austin
In the post-FinFET era, as technology transitions to below 2nm and beyond, embedded memories face increasing pressure. On one hand, workloads like HPC, AI training, and real-time inference demand massive on-chip bandwidth and capacity. On the other hand, conventional SRAM scaling has clearly been approaching physical and architectural limits. This workshop brings together experts from industry and academia to examine the future of embedded memory technologies with a sharp focus on practical and scalable solutions for advanced process nodes.

The session begins by framing the stakes from an AI hardware perspective (NVIDIA), underscoring how memory, is often the bottleneck in large systems. IBM then provides a sweeping technological and historical context, comparing volatile and non-volatile memory types and tracing their evolution. The subsequent presentation from INTEL provides a foundry-perspective on SRAM scaling, SRAM circuit design in cutting-edge nodes, and DTCO, while NXP and RAAAM discuss the potential for high-density logic-compatible non-SRAM embedded memory technologies.

Next, the workshop shifts to the process and integration frontier. IMEC presents the next frontiers in DTCO as well as technology breakthroughs that create new design and integration options for memory. Finally, Stanford University closes with a visionary academic outlook on future memory architectures, and technologies, including

3D integration options and their impact on architectures and logic-memory co-optimization.
In summary, this workshop offers a timely, focused, and high-impact exploration of the trajectory of embedded memory in advanced technology nodes, from established SRAM to emerging alternatives and 3D integration. It aims to engage the VLSI community in identifying and developing paths to revive and continue density scaling for on-chip storage and caches to meet the growing demands of a wide range of applications, including, but not limited to AI. The session will appeal to circuit designers, system architects, process technologists, and researchers.

- 1:03 PM– W2-1 | **The Strategic Role of Embedded Memory in Future AI/HPC Platforms**, Mahmut Ersin Singangil, NVIDIA
- 1:25 PM– W2-2 | **Pushing Memory Beyond Scaling Twists Through Innovations**, Rajiv Joshi, IBM
- 1:47 PM– W2-3 | **Challenges and Solutions in SRAM Scaling at Advanced Sub-2nm Nodes**, Fatih Hamzaoglu, Intel
- 2:09 PM– W2-4 | **Breaking the SRAM Bottleneck: High-Density GCRAM and System-Level Memory Innovation for Modern SoCs**, Robert Giterman, RAAAM and Nihaar Mahatme, NXP
- 2:31 PM– W2-5 **| DTCO for SRAM scaling and future opportunities using scaled emerging memories**, Fernando Garcia Redondo, imec
- 2:53 PM– W2-6 | **Disruptive Architectures via Diverse, Dense, and Deeply 3D Integrated Memory + Logic,** Robert M. Radway, University of Pennsylvania and Subhasish Mitra, Stanford University

W3 | Combining Light and Logic: Electronic–Photonic Co-Design for High-Performance Systems

1:00 PM – 3:15 PM, Tapa 3
Organizer: Ruud Oldenbeuving, imec

The convergence of electronics and photonics is transforming computing, communication, and sensing architectures. As AI, cloud, and real-time applications scale, electrical interconnects face limits in power, thermal stability, signal integrity, and latency.

Photonic Integrated Circuits (PICs) offer major gains in bandwidth, energy efficiency, and reach. Silicon photonics now enables high-volume optical transceivers and WDM systems for datacenters and data/telecom, while expanding into LiDAR, biosensing, quantum, and microwave photonics.

Yet PICs depend on Electronic ICs (EICs) for modulation, readout, and processing. Drivers, TIAs, PLLs, DSPs, and AI accelerators form the electronic backbone. Coupling EICs and PICs holistically is essential to maximize performance and efficiency while minimizing cost and complexity. This workshop promotes Electronic–Photonic Co-Design as a foundation for next-generation systems.

- 1:10 PM– W3-1 | **Enabling Integrated Photonics through Circuits, Photonics and Packaging Co-Design**, James Jaussi, Intel
- 1:35 PM– W3-2 | **From Circuit to System: Co-Simulation for High-Performance Electronic–Photonic Designs**, Parya Samadian, Synopsys
- 2:00 PM– W3-3 | **Electronic–Photonic Co-Design of Co-Packaged High Throughput Optical Transceivers**, Mayank Raj, AMD
- 2:25 PM– W3-4 | **Co-Integrated and Co-Designed Electronic–Photonic Transceivers for High-Speed Interconnects**, Xin Yin, Ghent University
- 2:50 PM– W3-5 | **Addressing Key Challenges in Optical Engine Integration Through Advanced Packaging**, Kimin Jun, Samsung

W4 | Design, System and Cross-Technology Co Optimization for Silicon Spin Qubits

3:30 PM – 5:45 PM, Tapa 1
Organizers: Chris Escott, Diraq and Nard Dumoulin, Diraq
Quantum computers have the promise to disrupt several aspects of society by solving problems in chemistry, materials, optimization, and cryptography that are intractable today. However, delivering this promise requires

fault tolerance, which in turn demands an exponential scale-up in the number of qubits compared to today's prototypes [1].

Therefore, building commercially viable quantum computers requires a scale-up which cannot happen without deep collaboration between quantum device innovators and the CMOS manufacturing ecosystem. This session aims to tackle the critical question: how do we make quantum computers more CMOS compatible, and CMOS more quantum computer compatible? By bringing together world leaders in spin qubit technologies, we'll discuss the strategies and frameworks that will enable quantum processors to move from lab prototypes to industrial reality.

Silicon spin qubits are a leading candidate for this scale-up because they leverage the existing CMOS ecosystem [1]. Yet, achieving manufacturability and yield at scale of quantum processors based on spin qubits requires technology co-optimization at the design (DTCO), system (STCO) and cross-platform (XTCO) levels.
[1] N. Dumoulin Stuyck et al., "CMOS compatibility of semiconductor spin qubits," 2024, arXiv. doi: 10.48550/ARXIV.2409.03993.

Topic covered: quantum computing, semiconductor spin qubits, quantum devices, quantum technologies, cryogenic CMOS, DTCO, spin qubit integration.

- 3:40 PM– W4-1 | **A Scalable 300mm Platform for Developing Gate-Defined Silicon Spin Qubit Devices**, Danny Wan, Imec
- 4:05 PM– W4-2 | **QSOI® : a Platform for Large Scale Quantum Integrated Circuits**, Tristan Meunier, Quobly
- 4:30 PM– W4-3 | **System-Level Implementation for Scalable Silicon Quantum Computers**, Goichi Ono, Hitachi
- 4:55 PM– W4-4 | **Millikelvin-CMOS for Quantum Integrated Circuits**, Nikolai Yurttagul, SemiQon Technologies
- 5:20 PM– W4-5 | **Scaling CMOS Spin Qubit Quantum Processors using Superinductors**, Albert L. Gomez Saiz, Quantum Motion

W5 | High-Performance CMOS for DRAM: Enabling Mobile, Graphics, Datacenter, and HBM in the AI Era

3:30 PM – 5:45 PM, Tapa 2
Organizers: Dan Mocuta, Micron Technology

The workshop examines how high-performance CMOS, advanced memory, and memory-centric architectures can meet AI's rising bandwidth, power, and scalability demands across mobile inference, data-center training, and HPC. Mustafa Badaroglu (Qualcomm) highlights the ingredients for scalable distributed AI—more-Moore compute, memory technologies, and scale-up/scale-out fabrics—under tight power, thermal, and footprint limits. Vamsi Alla (AMD) presentation will focus on HBM and the Memory Requirements of Datacenter AI. Mike O'Connor (Nvidia) will highlight key architectural and technology directions, including advanced packaging (2.5D/3D integration), HBM evolution, and opportunities for co-design across devices, circuits, and systems. Dwaipayan Biswas (imec) addresses inference bandwidth bottlenecks, including wide I/O LPDDR, 2.5D/3D scaling (active interposers, HBM-on-GPU), and near bank-level compute. William Kyunghoon Kim (SK Hynix) discusses DRAM optimization via CMOS/circuit co-optimization and a high-performance CMOS roadmap. Ameen Akel (Micron) reviews why workloads are memory-bound and how to scale effective bandwidth within power envelopes.

- 3:40 PM– W5-1 | **More-Moore Compute and Memory Technologies for Distributed AI Scalability**, Mustafa Badaroglu, Qualcomm
- 4:00 PM– W5-2 | **AI Workloads and the Future of DRAM: Bandwidth, Energy, and Integration at Scale,** Mike O'Connor, Nvidia
- 4:20 PM– W5-3 | **3D architectures enabling memory requirements for Datacenter AI,** Vamsi Alla and Zachary Blair, AMD

- 4:40 PM– W5-4 | **AI and DRAM: Optimizing Memory for Machine Learning and Inference**, Dwaipayan Biswas, imec
- 5:00 PM– W5-5 | **High-Performance CMOS Technology as a Key Enabler for Next-Generation DRAM**, William Kyunghoon Kim, SK Hynix
- 5:20 PM– W5-6 | **Memory-Centric Architectures for the AI Era: Pushing the Limits of Bandwidth, Efficiency, and Compute**, Ameen Akel, Micron Technology

W6 | VLSI Device Manufacturability: Improving Semiconductor Yield Through Virtualization

3:30 PM – 5:45 PM, Tapa 3
Organizers: Joseph Ervin, Lam Research

Virtualization enabled by machine learning is being introduced extensively into semiconductor research and development (R&D), but perhaps the most profitable use of this technology will be during yield ramp and high-volume manufacturing (HVM). The semiconductor industry faces severe challenges (such as systematic defects) during yield ramp, with yield ramp failures leading to massive financial losses. Traditional yield improvement methods—focused on isolated issues and wafer-based testing—are slow, costly, and often counterproductive.

Device and manufacturing virtualization can provide a solution to this problem. Virtualization can be used to lower the time and cost of physical experimentation and wafer-based development, particularly during yield ramp. The latest advances in ML/AI and virtualization are being quickly adopted by fabs, tool providers, test equipment companies, lithography vendors, EDA developers and others for root-cause identification, process backtracking, and causal modeling of failure modes. Most importantly, ML and virtualization are enabling the development of predictive tools that can be used to understand the impact of varying manufacturing conditions on final product yield. Process retargeting and scenario analysis is possible, along with a quantification of expected yield improvement under proposed conditions. The ability to perform multi-factorial, "in-silico" predictive yield analysis is going to revolutionize the design and manufacturing of semiconductor devices. Knowledge about these tools and how they are being applied is now critically important for industry participants.

In this workshop, semiconductor industry executives and thought leaders will discuss the latest advances in virtualization technologies, and how these are being applied in semiconductor yield ramp and high-volume production. Real life examples of yield improvement through virtualization will be reviewed, including applications in advanced device development, manufacturing, process development, lithography and other critical semiconductor manufacturing phases. The goal of this workshop is to provide participants with knowledge about how industry peers are using virtualization to improve the manufacturability and yield of advanced VLSI devices.

- 3:40 PM– W6-1 | **Accelerating Defect Reduction and Improving Manufacturing Yield using AI/ML and Physics-Based Process Modeling**, Joseph Ervin, Lam Research
- 4:05 PM– W6-2 | **AI Assisted Memory Technology Development**, Pei Lin Pai, Winbond Electronics
- 4:30 PM– W6-3 | **Accelerating Innovation in Memory Semiconductors: Enhancing Productivity Through a Virtual Manufacturing Solution**, Dongyean Oh, SK hynix
- 4:55 PM – W6-4 | **Manufacturability of High-Density 3D Interconnects by Combining Early Si-Work, Machine Learning and Virtual Process Emulation**, Zsolt Tokei, imec
- 5:20 PM– W6-5 | **Prescribing Frontend Manufacturing Excellence with Virtual Modeling**, Linmiao Zhang, Micron Technology

Monday, June 15, 2026

Short Courses | 8:30 AM – 5:00 PM

SCC: Circuits Short Course | AI-Driven Design Acceleration: Learning Across Circuits, Technology, and Yield

Tapa 3

Organizers: Vanessa Chen, Carnegie Mellon and Kenichi Okada, Institute of Science Tokyo

8:35 AM

SCC1 | When Will Analog EDA Have Its ChatGPT Moment?, Boris Murmann, University of Hawaii

Abstract: Analog, RF and mixed-signal IC design has been stubbornly slow and largely manual for many decades. Although a large number of EDA tools for full or partial automation have been developed, there has been limited adoption due to a wide range of issues related to the unstructured and intricate nature of analog design. Will this situation change given the current trajectories and success of AI-driven automation? This presentation will investigate this question based on the experience and perspectives of an analog designer. It will focus on identifying the main showstoppers that must be understood and addressed before AI-driven analog design can thrive.

9:25 AM

SCC2 | AI for Accelerating Memory Development, Tsung-Yung Jonathan Chang, TSMC

Abstract: As semiconductor technology advances to newer nodes, achieving chip design convergence and timely IP delivery within traditional cycle times has become increasingly challenging. This is driven by a complex interplay of factors, including ever-increasing transistor count and memory density, layout-dependent effects, the complexity of DRC rules, and lengthy simulation runtimes. This short course will demonstrate how to leverage AI technology to accelerate memory development. Using SRAM as a primary example, we will illustrate how AI can boost IP development productivity by at least 2X annually.

10:35 AM

SCC3 | Generational Transformation That Is Reengineering VLSI, Aman Joshi, SanDisk

Abstract: This short course/presentation will cover how Sandisk engineering has implemented AI for accelerating VLSI design. It will also cover the AI trends emerging that can accelerate AI adoptions in Enterprises. In addition, the presentation will include AI system design architecture that is accelerating the LLM performance and shed some light on why memory and storage is a critical component of this, fast evolving core AI hardware infrastructure.

11:25 AM

SCC4 | From Theory to Real IC: Practical AI-Driven Floor Planning at MediaTek, Tai-Lai Tung, MediaTek

Abstract: As integrated circuit (IC) design scales and complexity reach unprecedented heights, traditional methodologies struggle to meet demands for faster schedules, improved power, performance, and area (PPA), and reduced redesign risks. In this talk, I share MediaTek's real-world journey in developing and deploying AI-powered solutions for floor planning—charting a path from initial motivation to practical impact. We explore why MediaTek chose to focus first on AI-driven floor planning, targeting bottlenecks such as manual effort, iterative design cycles, and quality assurance. Our methods leverage reinforcement learning, graph neural networks, and deconvolution architectures to automate block placement and aspect ratio selection, enabling the creation of high-quality floorplans in hours rather than days. Throughout the session, I will share lessons learned, practical tips for teams adopting AI in IC design, and insights into data requirements, model training, and integration with existing design flows.

Lunch Break

1:15 PM

SCC5 | Opportunities and Limitation in AI Assistant for Analog IC Design, Mike Shuo-Wei Chen, USC

Abstract: Analog integrated circuit design has traditionally been viewed as an art work and requires a mostly manual design process. In recent years, there are resurgent interests in making analog circuit design easier and faster. For example, there has been circuit architecture development that makes synthesizable analog circuit block become more feasible than ever. Meanwhile, AI/ML has been transforming our society in many ways, providing unprecedented convenience to our daily life. Many successful examples have been shown for image, video and language processing and generation. Can AI/ML similarly benefit analog IC design? In this talk, I will overview past and ongoing efforts in facilitating analog IC design from design exploration, simulation, to tapeout. I will also discuss various design examples and silicon prototypes to examine their effectiveness and limitations.

2:05 PM

SCC6 | Machine Learning in Diagnosis, Shawn Blanton, Carnegie Mellon University

Abstract: As process nodes shrink to the atomic scale, achieving high-volume yield increasingly relies on a fundamental understanding of failure mechanisms. While repetitive hard macros like SRAM have historically served as the 'canaries in the coal mine,' the primary challenge has shifted to logic. Due to the unpredictable physical topologies generated by evolving Place-and-Route algorithms, localizing and mitigating logic failures is now the gatekeeper for producing high-yielding, reliable chips. In this talk, the application of machine learning (ML) in diagnostic processes within the context of the Advanced Chip Testing Laboratory is explored. The talk is partitioned into three stages: pre-diagnosis, during-diagnosis, and post-diagnosis. Pre-diagnosis involves activities conducted before diagnosis deployment, such as traditional diagnostic ATPG and DFT enhancements. ML applications at this stage include test reordering, optimizing test response data collection, and predicting diagnosis outcomes to better allocate computational resources. During-diagnosis encompasses algorithms like cause-and-effect analysis and path tracing, supported by technologies such as fault simulation. ML here is used to learn dynamically during diagnosis, exemplified by models like k-nearest neighbors that evolve on-chip to improve diagnostic accuracy. Post-diagnosis covers activities after diagnosis execution, primarily focusing on volume diagnosis to enhance resolution—the ability to identify failure points within the netlist or layout. ML's role in post-diagnosis is significant due to the availability of extensive, well-structured data from previous diagnoses, physical failure analyses, and fault simulations. The bulk of the talk emphasizes post-diagnosis, with specific focus on improved localization and behavior identification to improve overall diagnostic effectiveness.

3:15 PM

SCC7 | Delivering High-Quality Chips at Rapid Speed: The Future of Chip Design with Rapidus DMCO/Raads, Koki Tsurusaki, Rapidus

Abstract: With the explosive progress of AI, autonomous driving, and High-Performance Computing (HPC) in recent years, cutting-edge logic semiconductors are required to achieve higher levels of PPA (Power, Performance, and Area) than ever before. In the 2nm generation and beyond, the complexity of design has increased dramatically due to intricate physical phenomena, including lithographic limits, physical variability, and thermal issues. In traditional semiconductor development flows, Design and Manufacturing are siloed. Challenges on the manufacturing side often surface only after the design is finalized, leading to extensive design revisions (iterative processes). This disconnect has become the primary factor hindering Time to Market (TTM). The Core Strategy: DMCO To overcome these challenges, Rapidus has placed Design-Manufacturing Co-Optimization (DMCO) at the core of its strategy—a concept that moves toward a highly integrated fusion of design and manufacturing. DMCO transcends the conventional DFM (Design for Manufacturing) framework. It is a methodology that dynamically feeds back physical characteristics of the manufacturing process and equipment variability into the design phase to optimize both simultaneously. The chip design solution embodying this strategy is Raads (Rapidus AI-Agentic Design Solution), currently under development by Rapidus. Raads: AI-Agentic Design Solution In addition to the standard Reference Design Flow typically provided by foundries, Raads provides an "Agentic" design infrastructure where AI agents autonomously judge and execute each stage of design, verification, and optimization. In this presentation, I will provide an overview of these two technologies and explain how they contribute to shortening the time to market.

4:05 PM
SCC8 | Agent AI in EDA: Navigating the General Challenges of Autonomous Design and Optimization, Erick Chao, Siemens
Abstract: The EDA industry is at the forefront of leveraging Agent AI to tackle increasingly complex chip design and verification challenges. This presentation explores the transformative potential of autonomous agents in EDA, from intelligent design space exploration to adaptive optimization. We will delve into the core technical hurdles that arise when deploying Agent AI in such intricate domains. The discussion will generalize these challenges, demonstrating their relevance

SCT: Technology Short Course | Technologies Shaping the Future as Key Enablers for AI

Tapa 2
Organizers: Anabela Veloso, imec and Yoshiaki Kikuchi, Sony Semiconductor Solutions

8:35 AM
SCT1 | Opportunities for Future Advanced Logic and Technology Scaling, Dechao Guo, IBM
Abstract: Advanced logic technology scaling has powered semiconductor technology evolution for more than five decades—and it will remain essential for driving progress in the decades ahead. Yet the breakthroughs that will sustain future scaling are different from those that defined earlier generations. This short course will highlight the multidimensional innovations poised to shape the next era of logic technology, offering insights into the device, interconnect, material, and system level advances that will define future nodes.

9:25 AM
SCT2 | 3D Augmented Dimensional Scaling with Design-System Technology Co-Optimization, Dwaipayan Biswa, imec
Abstract: In the ever-evolving technology landscape, the advent of denser technologies has become a catalyst for unprecedented innovation. Technology scaling enable higher compute density, memory density/bandwidth and hence is a key factor towards scalable system architecture design. In the more than Moore era, system-technology co-optimization (STCO) is a promising paradigm for leveraging the synergy between emerging technology and application-driven architectures to achieve higher efficiency and performance at cost parity. As artificial Intelligence (AI) and Machine Learning (ML), continue to advance and become more integrated into various industries, there is a growing need for joint optimization across the stack – workloads-architecture-design and technology. In this talk, we will discuss scaling bottlenecks in four system domains – Mobile; Cloud CPU servers; Datacenter GPUs; Edge Inference accelerators and the role of emerging technology to mitigate these bottlenecks. 3D integration and advanced packaging offer great opportunities to unlock CMOS scaling – new device architectures, BEOL boosters, backside technology, functional partitioning and heterogeneous memory stacks, all of which will play a key role in next generation system architectures. Through system disaggregation, technology specialization and deep scaled assembly methods, the CMOS2.0 era will be instrumental in the future of dimensional scaling.

10:35 AM
SCT3 | Heterogeneous Integration with Advanced 2.5D/3D Technologies, Chih Hang Tung, TSMC
Abstract: The role of System integration and packaging in semiconductor industry has transformed in recent years. Wafer Level System Integration enables system scaling complementing Moore's Law transistor scaling. This short course focuses on technologies enablement by 2.5D and 3D chip level interconnection. Technologies and challenges from electrical, structural, and thermal are discussed. Industry solutions are covered. Novel and new heterogeneous integration, including photonics are also discussed for their potential future roles in system Power, Performance, Area, and Cost (PPAC) scaling.

11:25 AM
SCT4 | Material and Process Integration Innovations and Solutions Enabling AI Architectures Fueling the AI Era, Milan Pesic, AMAT

Abstract: The current AI revolution is increasingly driven by power-hungry compute and memory devices. In addition, moving data between compute and memory decreases performance and increases power consumption. This short course explores pivotal role of material and process advancements and innovations specifically designed for the ongoing evolution of compute and memory technologies (specifically DRAM). In DRAM scaling, key focus areas include high-aspect-ratio (HAR) capacitor etching and improvements in capacitor and electrode materials. Efficient capacitor scaling requires higher-k, lower-leakage film stacks. As scaling limitations arise, emerging processes such as wafer bonding in VCT (4F2), integration challenges of high-aspect ratio 4F2 access device and 3D DRAM will be explored. Continuous innovation in processes and tools—including film deposition, etching, wet processing, wafer bonding, and thinning—is crucial to support these advancements. The second portion of the short course is centered around major performance bottlenecks and performances across key transistor modules—superlattice, channel, source–drain, contact, and interconnects. A holistic view of the intrinsic performance and capacitance limits within each element and link these constraints to emerging solutions enabled by novel materials and process innovations will be provided. The discussion is framed around today's hGAA technologies and the transition toward CFET architectures, where these innovations become even more critical. Finally, we will discuss materials-driven approaches to engineering interconnect fabrics aimed at reducing the RC delay between compute blocks.

Lunch Break

1:15 PM

SCT5 | Beyond 6F2 – Scaling Frontiers and the Evolution towards Future DRAM Architecture, Jin-Woo Han, Samsung

Abstract: As traditional DRAM scaling approaches the physical limits of the 6F2 cell structure, the industry faces an inflection point. This course explores the critical transition towards next-generation DRAM architecture designed to overcome the traditional limits. It analyses the layout shift from 6F2 to 4F2 and the eventual migration to stacked architecture 3D DRAM. Attendees will gain a comprehensive understanding of the challenges and enablers of such paradigm shifts for sub-10 nm era.

2:05 PM

SCT6 | Outlook, Challenges, and Opportunities of Emerging Non-volatile Memory Technologies, Masatoshi Yoshikawa, KIOXIA

Abstract: The rapid expansion of artificial intelligence (AI) is fundamentally reshaping the memory hierarchy across both cloud and edge platforms. Current generative AI, primarily based on large language models, increasingly involves workloads that require the movement of large datasets. This trend reveals the limitations of conventional von Neumann architecture. As cloud AI demands large-capacity and high-bandwidth memory for GPUs, high-bandwidth memory (HBM) has emerged as a commercially viable on-chip solution, consisting of 3D chip-stacked DRAM with through-silicon vias (TSVs). As AI continues to evolve, the focus shifts from training to inference. The next phase of AI is expected to center on edge AI, particularly in endpoint IoT, consumer devices, and industrial applications. Edge AI requires always-on and instant-on memory systems characterized by low power consumption and low latency. While advancements in AI are driving changes in computing architecture, there is a focused effort to reduce energy consumption related to the movement of large datasets between CPUs/GPUs and memories by placing the memory closer to these processing units. Moreover, the concept of "Computing in Memory (CiM)", where computations are performed within the memory itself, has attracted significant attention for its potential to enhance computational efficiency, paving the way for practical applications. In this presentation, we will review emerging non-volatile memories (NVMs) not merely as simple replacements for existing memory technologies such as SRAM and DRAM but also as critical enablers of an AI-driven memory hierarchy. We will first explore the current landscape and advancements in emerging NVM technologies, focusing on spin-transfer torque magnetoresistive RAM (STT-MRAM), ferroelectric RAM (FeRAM), resistive RAM (ReRAM), and phase-change memory (PCM), including selector-only memory (SOM). While these technologies have achieved a degree of industrial maturity, they continue to face intrinsic challenges such as reliability, power consumption, and scalability, all of which must be addressed to broaden their range of applications. Finally, we will discuss

developmental pathways for NVMs that can leverage the opportunities presented by next-generation AI and innovative computational systems.

3:15 PM
SCT7 | Oxide Semiconductors for BEOL Monolithic 3D Integration: Materials, Devices, and Applications, Gong Xiao, NUS
Abstract: Back-end-of-line (BEOL)-compatible oxide semiconductor devices offer a compelling combination of advantages, including cost-effective large-area deposition, conformal growth on complex 3D structures, BEOL-compatible processing, decent carrier mobility, and wide bandgap characteristics. These unique properties position oxide semiconductors as powerful enablers of ultra-high-density, high-performance, and energy-efficient monolithic 3D integrated circuits. They unlock a wide spectrum of opportunities, from access transistors and active channel layers in advanced memory technologies to functional building blocks such as amplifiers for power management and voltage regulation. In this short course, I will present an overview of oxide semiconductor materials, devices, and emerging applications. The presentation will cover key material properties, critical process modules (gate stack engineering and contact technology), advanced device architectures, reliability concerns such as positive and negative bias temperature instabilities (PBTI and NBTI), and development of p-type oxide semiconductors. By combining recent advances with a candid assessment of remaining challenges, this talk aims to provide a holistic perspective on the future potential of oxide semiconductor devices and their path toward large-scale adoption.

4:05 PM
SCT8 | Enabling Advanced Optical Interconnects for AI Computing using Silicon Photonics, Ben Lee, NVIDIA
Abstract: Today's AI systems employ tightly integrated GPU clusters using low-cost high-reliability intra-cluster electrical signaling. Cable reach ultimately limits cluster size and aggravates thermal density. Efficient, low-cost, distance-insensitive off-chip communication can alleviate thermal challenges, increase cluster size, and simplify packaging. However, to address the challenge, optical interconnects must meet stringent efficiency, cost, and density targets. This short course explores where silicon photonics device performance and integration capabilities intersect future system requirements; then, it motivates an approach for efficient bandwidth scaling using highly integrated multi-wavelength photonic engines.

Demo Session & Reception

5:45 PM – 7:45 PM

Tapa 1

- **C3.1** Gleanmer: A 6 mW SoC for Real-Time 3D Gaussian Occupancy Mapping, **Zih-Sing Fu,** Massachusetts Institute of Technology
- **C4.1** A 209~245GHz CMOS Spectroscopic Transmitter with DTC-Based 4FSK Baseline Elimination and Phase-Offset Calibrated SSPLL for Chip-Scale Molecular Clocks, **Geyu Qiang,** University of Electronic Science and Technology of China
- **C15.2** A 10-MHz Integrated Voltage Regulator with Cycle-Accurate and Lightweight Predictive Control Achieving 5-mV Off-Chip Droop, **Haochang Zhi,** Southeast University
- **T5.3** Record 2Pr (>38 μC/cm2 at 0.5 V, >28 μC/cm2 at 0.4 V) of 3D MFM Capacitors Enabled by 3 nm HZO and ALD-TiN Orientation Engineering, **Yang Feng**, Shandong University
- **C24.4** A 4.32uW/Ch 256-Channel Neural Probe Integrating in-situ ADC and Stimulation for Closed-Loop DBS, **Shihui Sun,** Peking University
- **C26.2** A Cross-Band Wi-Fi/BLE Passive Tag Demonstrating Battery-Free Standard-Mask Communication with Smartphones and Battery-Free AoA Localization, **Qijing Xiao,** Zhejiang University
- **JFS5.4** A 28nm 0.7 pJ/bit Optically-Programmable, Fully-Digital SRAM-based Receiver Array with On-Chip Skew Calibration for Energy-Efficient Wireless Data Transfer, **Yifan He,** Cornell Tech

- **JFS6.2** A 1.58-W/mm^2 Extreme-Power-Density Ka-Band GaN PA with Electro-Thermal-Mechanical Co-Design in CMOS+GaN Heterogeneous Integration, **Ruitao Ma,** Hong Kong University of Science and Technology (HKUST)
- **T10.3** High Yield Sub 20 nm Ru Direct Metal Etch Enabled by Single-Exposure 0.55NA EUV and dry MOR Technology, **Martin O'Toole,** ASML
- **T12.4** A 4Gb 32-Layer 3D NOR Flash Test Chip with Innovative Design Methods Enabling Sub-100ns Low-Latency Read and Robust Retention, **Hang-Ting Lue,** Macronix International Co., Ltd.
- **T19.4** First Demonstration of High-precision (15 bits) Photonic-electronic Hybrid Compute-in-memory System Based on FeFET Pockels Photonic Memory, **Haochen Yu,** The Hong Kong University of Science and Technology (Guangzhou)

Evening Panel | AI: Grand Vision or Grand Delusion?

8:00 PM – 10:00 PM
Tapa 2
Organizers: Ibrahim Abdo, NTT, Inc.; Gary Bronner, Rambus; Tetsuya Iizuka, The University of Tokyo; Tetsu Ohtou, Tokyo Electron; Harold Pilo, Cadence; Farhana Sheikh, Intel Corp.; Jixin Yu, Sandisk; and Vita Pi-Ho Hu, National Taiwan University

Moderators: Gary Bronner, Rambus and Vita Pi-Ho Hu, National Taiwan University

Panelists: Kazunari Ishimaru, Rapidus; Hoshik Kim, SK Hynix; David Kanter, Real World Insights/ML Commons; Tom Burd, AMD; and Raja Koduri, Oxmiq Labs

The AI world stands at a critical inflection point. Ambitious scaling goals are colliding with fundamental physical and economic constraints. OpenAI's call for $1T level investments in AI infrastructure will require massive investment in new datacenters, fab capacity and energy generation. There is tension between the desired growth trajectories and manufacturing feasibility, highlighting a broader disconnect between AI aspirations and infrastructure realities. While inference represents the primary monetization opportunity for AI systems, the massive capital expenditure required for data center expansion raises questions about optimal deployment strategies. Will 100x scaling in AI datacenters deliver 100x benefit? Power is a particular concern with targets of hundreds of GW of power needed by the 2030s and each GW of power needing the equivalent of a nuclear reactor. Is today's AI ready for such investment? Is building out AI datacenters at this scale the right move or is more research needed to improve AI efficiency. Will scaling compute in support of Large Language Model (LLMs) take AI to the next level? Or are we setting ourselves up for an AI winter where AI won't be able meet its promise. Are we living through a "dot.AI" bubble, similar to the "dot-com" bubble of the internet era? Or are those concerns overblown? As the industry grapples with these competing forces, the path forward requires careful consideration of whether current scaling strategies can overcome fundamental physical limitations or whether paradigm shifts in computational approaches will be necessary to realize AI's transformative potential.

Tuesday, June 16, 2026

Opening and Plenary

8:00 AM
Tapa 1-3

Chairpersons: Benjamin Colombeau, Applied Materials and John Wuu, AMD

Welcome and Opening Remarks and Awards
Vijay Narayanan, IBM and Ron Kapusta, Analog Devices

P1.1 – 8:35 AM
Building the Engine of AI: From Foundational VLSI Technologies to System-Scale Impact (Invited); Richard Ho, OpenAI
AI is reshaping our lives. New models demonstrate increasing levels of intelligence in applications ranging from entertainment to productivity to scientific discovery. This impact is profound and accelerating, but the frontier-scale training and inference-demands this creates is placing unprecedented stress on every layer of hardware systems.

For AI to serve all of humanity, we must solve these hardware challenges at scale to overcome bottlenecks in compute, memory bandwidth, connectivity, and physical datacenter infrastructure.

Progress will not come from transistor scaling alone; it will require advances in circuits and technologies for memory integration, low-power interconnect, power delivery, thermals, and advanced packaging. These innovations must be co-optimized within a coherent system architecture. Only through holistic, system-level innovation—paired with disciplined execution—can we achieve the perf/watt and perf/TCO improvements needed to make AI's benefits broadly accessible.

P1.2 - 9:15 AM
Advanced Package for Next-Generation AI System Scaling, (Invited); Dr. L.C. Lu, TSMC
Advanced packages have emerged as the pivotal technology for next-generation AI system scaling to meet demanding performance, power efficiency, and bandwidth requirements. As compute density increases, the growing need for communication bandwidth reaches the highest priority. We will first present continuous UCIe bandwidth increase through the advancements in 2.5D package. Next, we will illustrate silicon photonics innovations, addressing communication energy efficiency and system-level interconnect challenges within the AI datacenter. As multi-layer 3D stacking for silicon, memory and package becomes the mainstream integration, thermal and power supply become the key limiters on performance. We will present the thermal optimization for 3D stacking integration to address the cooling needs, as well as the vertical power delivery network (PDN) used to reduce IR drop and improve power efficiency. Finally, we will report the key highlights of 3Dblox's crucial role in the design industry, enabling 3DIC design interoperability and heterogeneous multi-die automation, as it actively undergoes IEEE standardization.

Technical Sessions Block 1 (10:15 AM – 12:20 PM)

Session C1: High-Frequency Arrays for 6G and SATCOM

10:15 AM, Honolulu 1
Co-Chairs: Jorge Lagos, imec and Wei Deng, Tsinghua University

10:15 AM

C1.1 A 129-Gb/s CMOS W-band Polarization-Modulated Active Relay Transceiver Supporting Concurrent Dual-Polarized Dual-Stream Relay Operation, Yifu Li[1], Rui Wang[1], Yuchao Mei[1], Xinyu Sui[1], Xu Luo[1], Zheng Wang[1], Hanwen Shi[1], Qingbin Li[1], Wenlong Sun[1], Chuanhan Zhang[1], Yi Zhang[1], Jing Jin[1], Lin-Sheng Wu[1], Jian Pang[1]
[1]State Key Laboratory of Radio Frequency Heterogeneous Integration, Shanghai Jiao Tong University

This paper presents a CMOS W-band polarization-modulated active relay transceiver. The input and output of the proposed relay are reversely-switched between H pol. and V pol. by a 3-GHz polarization-switching clock. With enough TR isolation, a 16T+16R relay module in this work can support concurrent dual-stream dual-polarized relay in 64QAM. A maximum relay data rate of 129 Gb/s is also demonstrated.

10:40 AM

C1.2 A 240–270 GHz 4x4 Bi-Directional Phased-Array Transceiver with On-Chip Half-Wavelength-Spaced Array and Ultra-Low Power Consumption in 65-nm CMOS, Chun Wang[1], Olivia Angel Yong[1], Chenxin Liu[1], Wenqian Wang[1], Anyi Tian[1], Abanob Shehata[1], Takumi Kojima[1], Hans Herdian[1], Yudai Yamazaki[1], Sunghwan Park[1], Minzhe Tang[1], Dongfan Xu[1], Sena Kato[1], Yuncheng Zhang[1], Kazuaki Kunihiro[1], Hiroyuki Sakai[1], Kenichi Okada[1]
[1]Department of Electrical and Electronic Engineering, Institute of Science Tokyo

This paper presents a 300-GHz-band 2D 4×4 bi-directional phased-array transceiver with on-chip half-wavelength-spaced array and ultra-low power consumption in 65-nm CMOS. Each element comprises a phased shifter, doubler, injection-locked tripler, sub-harmonic mixer, and on-chip dipole antenna, occupying a compact core area of 0.30 mm^2. The power consumption of each element is 26 mW. The 4×4 phased-array features pitches of 0.49- and 0.50-wavelength in the E- and H-planes, respectively, with measured main-beam scanning angles of ±32°. The single chip achieves a peak EIRP of -17.1 dBm at 256 GHz.

11:05 AM

C1.3 A Ka-Band Time-Division Multi-Stream Phased-Array Transmitter with Built-in Spatial Notch Generation for 6G MIMO, Minzhe Tang[1], Ziyuan Ren[1], Yuxuan Liu[1], Yilun Chen[1], Minghao Fan[1], Dongfan Xu[1], Haiyun Gu[1], Zheng Li[1], Yi Zhang[1], Jian Pang[1], Daxu Zhang[1], Zezheng Liu[1], Dingxin Xu[1], Chenxin Liu[1], Duo Li[1], Yuncheng Zhang[1], Chun Wang[1], Yudai Yamazaki[1], Sena Kato[1], Hiroyuki Sakai[1], Kazuaki Kunihiro[1], Kenichi Okada[1]
[1]Electrical and Electronic Engineering, Institute of Science Tokyo

This work presents an area-efficient Ka-band time-division phased-array MIMO transmitter supporting two streams with only four RF elements. Nyquist-rate beam switching enables RF path sharing and MIMO streams coexistence, reducing chip area and power. A built-in spatial notch generation is embedded without extra hardware cost. A 65-nm CMOS prototype achieves 9.6 Gb/s OTA throughput via MIMO and 35 dB spatial isolation in measurement. An embedded band-pass filter with 42.5 dB/GHz roll-off achieves –24.6 dBc ACLR for OOB emission suppression.

11:30 AM

C1.4 An Ultra-Compact 4-Element 4-Beam Q-band Multibeamformer IC with IQ Pre-weighted Phase Shifter for VLEO SATCOM, Mingyu Lee[1], Hyungju Kim[1], Wonseob Lee[1], Hyeonwon Song[1], Seong-Mo Moon[2], Dongpil Chang[2], Songcheol Hong[3], Seungchan Lee[1], Jinseok Park[1]
[1]Department of Intelligent Electronics and Computer Engineering, Chonnam National University, [2]Satellite Payload Research Section, Electronics and Telecommunications Research Institute (ETRI), [3]Korea Advanced Institute of Science and Technology (KAIST)

This work proposes a compact Q-band TX/RX-configurable multibeamformer IC for very low Earth orbit(VLEO) SATCOM. The proposed IQ pre-weighted multi-beam phase shifter (PS) and small cell power dividers provide a compact chip area and low coupling errors. Additionally, the proposed analog linearizer compensates for both AM-AM and AM-PM distortions with a low insertion loss. The IC is fabricated using 65nm CMOS process, achieving low

area/element/beam of 0.48 mm^2. Across all 16 channels, measured P_{1dB} exceeds 18.2 dBm with AM-PM distortion below 2.8º. The measured RMS phase and gain errors are below 2º and 0.3 dB, respectively.

11:55 AM
C1.5 A 144Gbps D-Band Dual-Polarized MIMO High-Density Phased-Array Transceiver in 65nm CMOS for 6G UE, Yudai Yamazaki[1], Ryuji Kinugawa[1], Takumi Kojima[1], Takaya Uchino[1], Anyi Tian[1], Chenxin Liu[1], Sunghwan Park[1], Kotaro Ito[1], Sena Kato[1], Chun Wang[1], Minzhe Tang[1], Takashi Tomura[1], Yuncheng Zhang[1], Hiroyuki Sakai[1], Kazuaki Kunihiro[1], Kenichi Okada[1]
[1]Department of Electrical and Electronic Engineering, Institute of Science Tokyo

This work presents the world's first D-band dual-polarized (DP) multi-input and multi-output (MIMO) phased-array transceiver targeting 6G user equipment (UE), fabricated in a 65nm CMOS process. The proposed transceiver achieves a maximum data rate of 144 Gbps. Each IC integrates 4 vertical (V) and 4 horizontal (H) polarized TRX elements in a compact layout, resulting in power consumption per element of 124 mW in TX and 90 mW in RX. Furthermore, the transceiver is embedded within a high-power-density dual-polarized antenna-in-package (AiP) module, enabling long-distance single-stream communication up to 50 meters.

Session C2: High-Voltage-Ratio and Scalable Power Delivery

10:15 AM, Honolulu 2
Co-Chairs: Stephen Brink, Texas Instruments and Xun Liu, The Chinese University of Hong Kong, Shenzhen

10:15 AM
C2.1 A Monolithic 20W/mm^2 4.8V Input 94.8% Peak Efficiency 2-1 Switched Capacitor Voltage Regulator as First Stage Current Multiplier for Vertical Power Delivery, Minxiang Gong[1], Nicolas Butzen[1], Jingshu Yu[1], Harish Krishnamurthy[1], Sheldon Weng[1], Krishnan Ravichandran[1], Christopher Pelto[1], Ramiz Ahan[1], James Waldemer[1], James Tschanz[1]
[1]Intel Corporation

A monolithic 3.6-4.8V input 2-1 switched capacitor voltage regulator as first stage current multiplier in vertical power delivery, pushing package input to near 5V. It features stacked designs with synchronized auxiliary power trains, and clamp transistors, realizing self-generated drive voltages, efficient gate driving and fast startup, and achieving 20W/mm^2 power density and 94.8% peak efficiency.

10:40 AM
C2.2 A 2D-Extendable 48V-1V Ladder-Series-Capacitor Buck Converter for Computing Power Delivery, Zhiguo Tong[1,2], Wenjie Yang[1], Zhewen Yu[1], Shousheng Han[2], Rui P. Martins[2], Yan Lu[1]
[1]Tsinghua University, [2]University of Macau

This paper presents a 48V-to-1V two-dimensional extendable ladder-series-capacitor buck converter (LSCB) which supports both input-voltage stackable and output-current parallel architecture of power cells, to reduce switch voltage stress and extend output current simultaneously. We also implement an input-series output-parallel (ISOP) two-stage configuration to realize vertical power delivery. Fabricated in 180-nm BCD for input-series chips and 65-nm process for output-parallel chips, the cascaded ISOP system achieves a peak power density of 2132 W/in^3 while delivering a maximum output current of 25 A.

11:05 AM
C2.3 A 12V-to-1V Scalable 2N-Phase Buck Converter using 0V-Stress Balancer with Shortest I_L Balancing Path for Parallel IC Expansion in xPU Systems, Young-Jun Jeon[1], Chan-Ho Lee[1], Jong-Woo Jang[2], Sung-Wan Hong[3]
[1]Electronic Engineering, Sogang University, [2]Memory Business, Device Solutions Division, Samsung Electronics Co., Ltd., [3]School of transdisciplinary innovation, Seoul National University

This paper presents a scalable 2N-phase buck converter using a 0V-stress balancer with the shortest inductor current balancing path for parallel IC expansion in xPU systems. The proposed converter supports arbitrary phase expansion with all-phase interleaving. A prototype was tested using 1, 2, 3, and 4 chips (2, 4, 6, and 8 phases), and achieves peak efficiencies of 89.17% at 1.0 V and 90.4% at 1.2 V in 2-chip (4-phase) prototype.

11:30 AM
C2.4 A 48V-5V Coupled-Inductor Middle-Switched-Capacitor Hybrid GaN Converter Achieving 156µH Equivalent Inductance and 30mA Ripple, Zhiwen Gu[1], Hieu Pham[1], Hanh-Phuc Le[1]
[1]University of California, San Diego

This paper presents a middle-switched-capacitor (MSC) hybrid converter that employs coupling between the input and output planar inductor to achieve ideally ripple-free currents at the output, and a synchronous L-C path to also achieve the same at the input. Implemented in a 0.5-µm GaN-on-Silicon process, the 8.58-mm^2 MSC converter prototype achieves 30mA output current ripple and 160mA input current ripple, amounting to 48.8X and 9.5X ripple reduction, respectively.

11:55 AM
C2.5 A 0.5W/mm^2-Power-Density, 91.6%-Peak-Efficiency, Resonant Dual-Path Interleaved DC-DC Converter, Mauro Leoncini[1], Simone Zaffin[1], Alessandro Bertolini[2], Alessandro Dago[1], Alessandro Gasparini[2], Salvatore Levantino[1], Massimo Ghioni[1]
[1]Electronics, Information and Bioengineering, Politecnico di Milano, [2]STMicroelectronics

A DC–DC converter is presented that combines dual-path power delivery with resonant capacitor soft charging and interleaved operation to achieve high power density in sub-1V portable power management. The chip, implemented in a 0.18µm BCD technology, is co-designed together with a PCB interposer, enabling vertical stacking of passive components while leveraging the PCB parasitic inductance to realize zero-current switching. The converter operates with a 4.0-to-5.0V input voltage and delivers a current up to 7A at a voltage in the 0.6-to-1.2V range, with 91.6% peak power efficiency and 0.5W/mm^2 peak power density.

Session C3: Accelerators: Spatial and Immersive Visual Computing

10:15 AM, Honolulu 3
Co-Chairs: Yu-Hsin Chen, EnCharge AI and Wongyu Shin, Rebellions Inc.

10:15 AM
C3.1 Gleanmer: A 6 mW SoC for Real-Time 3D Gaussian Occupancy Mapping, Zih-Sing Fu[1], Peter Zhi Xuan Li[1], Sertac Karaman[1], Vivienne Sze[1]
[1]Massachusetts Institute of Technology

High-fidelity 3D occupancy mapping is essential for many edge-based applications (such as AR/VR and autonomous navigation) but is limited by power constraints. We present Gleanmer, a system on chip (SoC) with an accelerator for GMMap, a 3D occupancy map using Gaussians. Through algorithm-hardware co-optimizations for direct computation and efficient reuse of these compact Gaussians, Gleanmer reduces construction and query energy by up to 63% and 81%, respectively. Approximate computation on Gaussians reduces accelerator area by 38%. Using 16nm CMOS, Gleanmer processes 640×480 images in real time beyond 88 fps during map construction and processes over 540K coordinates per second during map query. To our knowledge, Gleanmer is the first fabricated SoC to achieve real-time 3D occupancy mapping under 6 mW for edge-based applications.

10:40 AM
C3.2 NOVA: A 5.3mJ/frame Fully Integrated Neural Video SoC with Content-Adaptive Architecture for 4K Real-time Encoding/Decoding and Processing, Youngjin Moon[1], Sangwoo Ha[1], Junha Ryu[1], Jungjun Oh[1], Jingu

Lee[1], Wooyoung Jo[1], Minsung Kim[1], Hoi-Jun Yoo[1]
[1]Korea Advanced Institute of Science and Technology (KAIST)

This paper presents NOVA, the first fully integrated neural video system-on-chip (SoC) for end-to-end decoding-processing-encoding on a single chip. A bound-guided mixed quantization (BMQ) and bound-guided unified (BU) core achieves 41.7% computation power reduction. A temporal interpolation unit (TIU) with tiered direct memory access (DMA) reduces system energy by 71.8%. An adaptive entropy coding unit (AECU) reduces coding latency by 79.2%. NOVA is fabricated in 28nm CMOS technology and achieves 31.4TOPS/W energy efficiency on 4K UVG benchmark.

11:05 AM
C3.3 SR-VLNA: A 5.0-23.9 mJ/meter Spatial Reasoning-based Vision Language Navigation Accelerator for Embodied Agents, Seryeong Kim[1], Jongjun Park[1], Sangmyoung Lee[1], Hyungnam Joo[1], Wonhoon Park[1], Seokchan Song[1], Junha Ryu[1], Gwangtae Park[1], Sangjin Kim[1], Jiwon Choi[1], Seongyon Hong[1], Wooyoung Jo[1], Hyeonrae Kim[1], Hoi-Jun Yoo[1]
[1]Korea Advanced Institute of Science and Technology (KAIST)

This paper presents SR-VLNA, an energy-efficient spatial reasoning-based vision-language navigation (VLN) accelerator. The spatial reasoning consists of 1) 3D Gaussian splatting-based semantic understanding (SU), and 2) language-based planning (LP). SR-VLNA proposes 4 features and reduces energy of SU and LP by 7.0× and 10.9×, respectively. Fabricated in 28nm FDSOI, SR-VLNA achieves SOTA SU performance at 0.004-0.021 μJ/point and is the first to demonstrate LP with 0.25-0.96 mJ/task. Finally, SR-VLNA demonstrates end-to-end VLN system with 5.0-23.9 mJ/meter.

11:30 AM
C3.4 Event-Gaze: A 28nm 441μs-Latency 12.88mW SoC for Sparse-Event-Based Dynamic Eye Tracking and Gaze Estimation in XR, Nealson Li[1], Che-Kai Liu[1], Zachary J. Ellis[1], Connor Talley[1], Young-Seok Noh[1], Kevin Patino-Sosa[1], Visvesh Sathe[1], Arijit Raychowdhury[1]
[1]Electrical and Computer Engineering, Georgia Institute of Technology

This paper introduces Event-Gaze, an eye-tracking and gaze-estimation SoC that natively processes asynchronous, sparse event-camera data. Using an algorithm-hardware co-design approach, the system performs graph construction from event streams and leverages dedicated accelerators for a Graph Neural Network (GNN) and a Transformer to enable eye tracking and gaze estimation, respectively. Fabricated in 28nm CMOS, the SoC achieves 1.12° accuracy with 441μs latency while consuming 12.88mW. Compared to baseline solutions, the proposed SoC demonstrates a 1.8× latency reduction and a 15% improvement in system power efficiency.

11:55 AM
C3.5 Uni4D: A 12nm 0.16mJ/Frame 4D Gaussian Spatial Video Rendering Processor with Unified Deformation and On-Chip Hierarchical Sorting, Cheng Nian[1], Xiaorui Mo[1], Jiaqi Zhang[1], Jiaying Peng[1], Fasih Ud Din Farrukh[1], Yifan Li[1], Wen Jia[2], Fei Chen[2], Liang Lu[3], Chun Zhang[1]
[1]Tsinghua University, [2]Research Institute of Tsinghua University in Shenzhen, [3]AMD

Uni4D is a 12-nm spatial video rendering processor for real-time 3D/4D Gaussian splatting with 0.16–0.52 mJ/frame energy. A unified deformation engine (UniMAC) combines Hex-plane sampling with MLP-based deformation, while K-plane Weights Segmentation (KWS) factorizes 3D indexing to reduce K-plane buffer area by 84%. A non-blocking key–value pair (KVP) generation engine and a hierarchical radix sorting engine jointly pipeline tile-based preprocessing and on-chip sorting, cutting Gaussian box-bounding latency by 76% and external memory access by 93.8%. Fabricated in 12-nm CMOS, Uni4D renders dynamic Gaussian scenes at up to 232 fps, achieving more than an order-of-magnitude lower energy than prior NeRF/3DGS processors.

Session T1: T1 - Technology Highlights

10:15 AM, Tapa 1-3
Co-Chairs: Yue Liang, nVidia and Shosuke Fujii, Kioxia Corporation

10:15 AM

T1.1 First demonstration of 3D Stacked FETs at Gate Pitch of 42 nm Featuring Triple Stacked Nanosheet Channels for Advanced Logic Applications, Donghoon Hwang[1], Jaeho Jeon[1], Gwangsik Kim[1], Seong Kwang Kim[1], Jiwoo Min[1], Hyunsoo Kim[1], Byungho Moon[1], Jina Kim[1], Jeongeun Her[1], Kyutae Jeong[1], Seongbin Park[1], Soobin Han[1], Hyeji Yoon[1], Jaehun Seo[1], Younsu Ha[1], Sandra Shaji[1], Aravindh Kumar[1], Tae-Hoi Park[1], Donghoon Kwon[1], Junghan Lee[1], Byounghoon Lee[1], Seungmo Ha[1], Sang-Moon Lee[1], Do-Sun Lee[1], Heonjong Shin[1], Wonchang Lee[1], Youngchai Jung[1], Harsono Simka[1], Seung Hun Lee[1], Sang Wuk Park[1], Wookhyun Kwon[1], Sangjin Hyun[1], Ilryong Kim[1], Jaihyuk Song[1]
[1]Semiconductor R&D Center, Samsung Electronics Co., Ltd.

We first present the industrial implement technology of three-dimensional stacked field effect transistors (3DSFETs) at gate pitch of 42 nm and triple-stacked nanosheet channels for n- and p-FETs, respectively. This paper provides a comprehensive overview of the key technologies required for the industrial production of 3DSFETs. Notable advancements include an epitaxial growth process that suppresses channel defect density and the incorporation of epitaxial layers with three distinct Ge contents to precisely form the middle dielectric isolation (MDI). Following the successful implementation of these essential technologies, we evaluate the electrical characteristics and discuss prospects for future logic applications.

10:40 AM

T1.2 Intel 18A-P CMOS Technology Enhancement Featuring Advanced RibbonFET (GAA) Transistors and PowerVia for High-Performance Computing, Anupama Bowonder[1], Orb Acton[1], Sayanee Adhikari[1], Ravi Aggarwal[1], Utsav Agrawal[1], Jeff Armstrong[1], Michael Asoro[1], Mehmet Aykol[1], Rohan Bambery[1], Andre Baran[1], Stuart Burgess[1], Kumhyo Byon[1], Derek Caselli[1], Maghesree Chakraborty[1], Sandrine Charue-bakker[1], Arya Chatterjee[1], Piyush Chaudhari[1], Surya Cheemalapati[1], Philip Chung[1], Dax Crum[1], Gopinath Danda[1], Ritesh K. Das[1], Nilanjan Das[1], Purnim Dhar[1], Gianna Di Francesco[1], Moshe Dolejsi[1], John Epple[1], Tawhid Ezaz[1], Sara Fathipour[1], Robin Forslund[1], Chethan Gaddam[1], Yuqian Gu[1], Bishu Guha[1], Walid M. Hafez[1], Chang Wan Han[1], Dennis Hanken[1], Daniel Harris[1], Michael Hattendorf[1], Phil Heil[1], I-chen Ho[1], David Hong[1], Mark Horsch[1], Alexander Hryn[1], William Hsu[1], Nabil S. Isa[1], Steven Jaloviar[1], Surangi Jayawardena[1], Kichang Jung[1], Jai G. Kameswaran[1], Vladislav Kamysbayev[1], Eric Karl[1], Sriharsha Karumuri[1], Steven Kirby[1], Mauro Kobrinsky[1], Santosh Krishnamurthy[1], Ojas Kulkarni[1], Avinash Kumar[1], Jurick Lahiri[1], Ankit Lakhani[1], Gyuhyon Lee[1], See-hoon Lee[1], Yean-an Liao[1], Chung-hsun Lin[1], Wenhe Lin[1], Yian Liu[1], Owen Loh[1], Gokul Malyavanatham[1], Saptarshi Mandal[1], Sanghamitra Mandal[1], Konstantinos Mavrakakis[1], Nikhil Mehta[1], Michael Mleczko[1], Andrew Moore[1], Nadjoua Moumen[1], Hema Chandra Prakash Movva[1], Ashok Murugavel[1], Saurav Nigam[1], Lucy Nolan[1], Steven R. Novak[1], Ling Pan[1], Sashidhar Panchamgam[1], Ivonne Paredes[1], Jun Kyu Park[1], Jessica Parker[1], Reken Patel[1], Pran Krishna Paul[1], Joshua Pauls[1], Leonard Pipes[1], Chetan Prasad[1], Conor Puls[1], Wei Z. Qiu[1], Rahul Ramaswamy[1], Omair Saadat[1], Joseph Saunders[1], Ameen Sayal[1], Bernhard Sell[1], Claudy Serro[1], Rushabh Shah[1], Faisal Shah[1], Manish Sharma[1], Vidhi Shingle[1], Mitchell Smith[1], Ryan Starko-bowes[1], Chen-yi Su[1], Shruti Subramanian[1], Afrin Sultana[1], Atsunori Tanaka[1], Sirnegeda Techane[1], Sunita Thulasi[1], David Towner[1], Shardul Wadekar[1], Xinning Wang[1], Yi-Ting Wang[1], Guowei Xu[1], Sulochana Dhar[1], William Grimm[1], Wenjun Li[1], Maruf Bhuiyan[1], Stephanie Chin[1], Kevin J Fischer[1]
[1]LTD 1278 Program, Intel Corporation

Intel 18A-P is a performance-enhanced, RibbonFET gate-all-around (GAA) transistor technology with backside power delivery through PowerVia. Relative to Intel 18A [1], Intel 18A-P offers over 18% lower power at iso-performance, or over 9% performance gain at iso-power. This improvement is achieved through newly added technology features, transistor performance enhancement, interconnect enhancement and design technology co-optimizations (DTCO). Added features in Intel 18A-P include additional logic VT pairs, skew corner tightening, new

low power devices in both high-density (HD) and high-performance (HP) libraries and performance-improved HP devices in both libraries. In addition, Intel 18A-P offers reduced thermal resistance for improved heat conduction.

11:05 AM

T1.3 First EUV-enabled integration route for 50 nm pitch N and PMOS transistors with 2D materials channel from a 300 mm fab , Tom Schram[1], Quentin Smets[1], Goutham Arutchelvan[2], Chia-Hsien Yao[3], Sui-An. Chou[2], Ming-Yang Li[2], Dmitry Batuk[1], Teresa Rodrigues[1], Souvik Ghosh[1], Pawan Kumar[1], Benjamin Groven[1], Ann Opdebeeck[1], Daniel Montero Alvarez[1], Yuchao Jiang[1], Tanushree Sarkar[1], Kristof Kellens[1], Evi Vrancken[1], Catarina Rega Da Silva[1], Celia Mogultay[1], Tzu-hsien Shen[1], Vina Faramazi[1], Gouri Sankar Kar[1], Cesar Javier Lockhart de la Rosa[1], Mark Vandal[2], Anthony Yen[3], Iuliana P. Radu[2], Etienne De Poortere[3]
[1]imec, [2]tsmc, [3]ASML

EUV lithography and 300mm fab processing are used to enable an innovative integration approach for transistors with 2D materials channel. The route demonstrates for the first time scaled transistors, as relevant for technology applications, contact pitch down to 50nm, active width down to 75 nm and EOT ~2nm. Quasi-CMOS integration was done by using different channel materials, MoS_2 for NMOS and WSe_2 for PMOS, side by side on the same 300mm wafer by die/ small wafer transfer. A key achievement is the demonstration of fab-processed 2D PMOS transistors nearly matching the performance of best lab devices. This solves a critical missing piece for the lab to fab transition of 2D and BEOL CMOS technologies.

11:30 AM

T1.4 A Multi-Stacked Cell Array Architecture with Wafer-to-Wafer Cu Direct Bonding for Ultra-High-Density 3D Flash Memory beyond 1,000 Word Lines, Mitsuhiko Noda[1], Susumu Hashimoto[1], Ryuta Mizumoto[1], Mamoru Watanabe[1], Genki Sawada[1], Kanta Kawasaki[1], Masahisa Sonoda[1], Eiji Kamiya[1], Tetsu Morooka[1], Keisuke Nakatsuka[1], Hiroshi Maejima[2], Shigeki Kobayashi[2], Shinji Suzuki[2], Yukihiro Sakotsubo[3], Kohei Osawa[4], Shinya Sato[3], Teiji Shibasaki[3], Masahiro Yaegashi[3], Ryoichi Honma[3], Masayoshi Tagami[1], Katsuyuki Sekine[1]
[1]Advanced Memory Development Center, Kioxia Corporation, [2]Memory Division, Kioxia Corporation, [3]Technology Development Engineering, Sandisk Corporation, [4]Device Integration Engineering, Sandisk Corporation

This paper demonstrates the world's first successful Quad-Level Cell (QLC) operation of a multi-stacked cell array CMOS directly bonded to array (MSA-CBA). This breakthrough overcomes key challenges in high stacking of 3D flash memory: cell current degradation, wafer warpage, and large block (BLK) size. These results mark a milestone toward ultra-high-density 3D flash memory with over 1,000 stacked layers, paving the way for next-generation memory technologies.

11:55 AM

T1.5 A16 Angstrom-class CMOS Technology featuring Enhanced Nanosheet Transistors with Super-Power Rail (backside direct contact power delivery) for AI and HPC Applications, Geoffrey Yeap[1], Maureen Wang[1], S.S. Lin[1], C.M. Lee[1,1], D. Hsu[1], I.W. wu[1], Y.S. Cheng[1], Y.C. Tseng[1], P.Y. Haung[1], H.Y. Chen[1], P.H. Haung[1], P.W. Wang[1], P. Hung[1], Y.K. Lin[1], C.F. Chang[1], C.Y. Kao[1], L.G. Chen[1], C.F. Tsai[1], K.F. Liao[1], P.C. JangJian[1], M.K. Chen[1], C.F. Fu[1], C.N. Lin[1], K.N. Liu[1], C.C. Lee[1], Y.H. Tang[1], C.Y. Ting[1], T.H. Chiu[1], H.C. Haung[1], K.B. Huang[1], C.L. Wu[1], C.P. Tsai[1], Y.H. Su[1], K.N. Yang[1], C.W. Lee[1], J.H. Lee[1], K.L. Chen[1], T.L. Lee[1], C.M. Liu[1], C.T. Chan[1], C.C. Yeh[1], H. Wang[1], P.Y. Lin[1], S.H. Sun[1], C.P. Lin[1], Y.C. Peng[1], H.L. Shang[1], H.C. Lin[1], J. Chang[1], X.M. Chen[1], H.T. Lin[1], K.W. Chen[1], L.C. Lu[1], Y. Ku[1], S.M. Jang[1]
[1]Taiwan Semiconductor Manufacturing Company (TSMC)

We successfully developed and qualified A16 platform technology, industry-first angstrom-class CMOS technology platform featuring Super Power Rail (SPR). TSMC's SPR is an innovative backside power delivery solution incorporating a novel backside direct contact scheme that preserves N2P gate density and NanoFlex design-technology co-optimization for maximum product design benefits. Compared with N2P (performance-enhanced N2), A16 provides a further 8%~10% faster speed at power, or 15%~20% power improvement and additional 8%~10% chip density gain. A16 is most suitable for AI/HPC products with complex signal routes and dense power delivery network. Mass production is slated for Q4'26.

Technical Sessions Block 2 (1:30 PM – 3:10 PM)

Session C4: Advanced High-Frequency Synthesizers

1:30 PM, Tapa 3
Co-Chairs: Martin Brox, Micron and Heein Yoon, Ulsan National Institute of Science and Technology (UNIST)

1:30 PM
C4.1 A 209~245GHz CMOS Spectroscopic Transmitter with DTC-Based 4FSK Baseline Elimination and Phase-Offset Calibrated SSPLL for Chip-Scale Molecular Clocks, Geyu Qiang[1], Xiao Luo[1], Xinyang Chen[1], Chang Liu[2], Cheng Wang[1]
[1]Electronic Science and Engineering, University of Electronic Science and Technology of China, [2]Chengdu Data Automation System Technologies

This paper presents a CMOS spectroscopic transmitter probing the 231.061GHz rotational spectra of OCS gas molecules for chip-scale molecular clocks (CSMC). It adopts 4FSK modulation through a DTC FSK modulator to mitigate the baseline tilting of gas cell with negligible SNR penalty. A 104.4~122.4GHz integer-*N* sub-sampling PLL (SSPLL, *N*=36) reduces the phase noise (PN) for high SNR. Phase offset calibration ensures robust locking of SSPLL under PVT variation. A 5.8~6.8GHz fractional-*N* sampling PLL (SPLL) drives the SSPLL. Fabricated on 65-nm bulk CMOS, it achieves a measured PN of −79.2dBc/Hz@2f_m=100kHz offset, and 86.9fs jitter at f_0 =231.061GHz, consuming P_{DC}=73mW. The measured 4FSK dispersion curve records a SNR of 85.7dB with only 1.9dB SNR penalty. After the CSMC locking, ADEV of 3.7×10^{-10}@τ=1s, and ADEV of 3.4×10^{-11}@τ=10^3s are measured.

1:55 PM
C4.2 A 45.3-to-53.9-GHz Circular-Dual-Core Series-Resonance-Harmonic-Extraction VCO Featuring Common-Mode Noise-Diversion Technique Achieving -112.5 dBc/Hz Phase Noise, Average 199.1/203.9 dBc/Hz FoM_A/FoM_{TA} and 0.8 dB Variation at 1-MHz Offset, Huanyu Eric Ge[1], Haikun Jia[1], Wei Deng[1], Baoyong Chi[1]
[1]School of Integrated Circuits, Tsinghua University

This paper presents a 45.3-to-53.9-GHz series-resonance harmonic-extraction VCO for high-speed wireline transceivers. It utilizes a circular dual-core topology to significantly enhance the quality factor within a compact footprint. A common-mode noise-diversion technique is proposed to suppress the $1/f^3$ noise corner and improve phase noise. The series-resonance topology enables a Q-boosted voltage swing, reaching a maximum third-harmonic output power of 4.8 dBm. Fabricated in 65-nm CMOS, the VCO achieves a phase noise of -112.5 to -111.6 dBc/Hz at 1MHz offset across the 17.3% tuning range. The design yields an average FoM_A / FoM_{TA} of 199.1 / 203.9 dBc/Hz at 1MHz offset with a variation of less than 0.8 dB, representing the state-of-the-art for oscillators operating above 30 GHz with a 2-to-3-dB lead over prior art.

2:20 PM
C4.3 A 54-to-60GHz Cascaded PLL Featuring Implicit Frequency Doubling and Harmonic-Mixing Phase Detector Achieving 53.8fs_{rms} Jitter and -281.1dB FoM_N, Yunbo Huang[1], Zunsong Yang[1,2], Li Wang[1], Kai Cheng[1], Hongyu Ren[1], Xiaoyu shan[1], Pui-In Mak[3], Yong Chen[4], Bo Li[1,2]
[1]Institute of Microelectronics, Chinese Academy of Sciences(CAS), [2]University of Chinese Academy of Sciences, [3]University of Macau, [4]Tsinghua University

This paper proposes a low-jitter and low-power mm-wave cascaded phase-locked loop (PLL). Key contributions include: 1) a low-noise class-F2 VCO with embedded capacitive 2nd harmonic extraction; 2) a 5th harmonic-mixing phase detector (HMPD) to realize a high PD gain (KPD) and low in-band noise. By combining the 2nd harmonic extraction and 5th harmonic-mixing, the fundamental PLL's output is frequency-multiplied by 10 with negligible jitter addition. The 54-60GHz PLL prototyped in 28nm CMOS achieves 53.8fsrms integrated jitter and −281.1dB division-N jitter-power figure of merit (FoMN).

2:45 PM
C4.4 A 28-GHz Quadrature LO-Phase-Shifting Digital Wave-Locked Loop (WLL) Achieving 57.2-fs$_{rms}$ Jitter, -80.6-dBc Spur$_{ref}$, and 0.59-µs Near-Integer Lock Time, Feifan Hong[1], Jiawen Chen[1], Chunxiao Liu[1], Pingda Guan[1], Robert Bogdan Staszewski[1], Teerachot Siriburanon[1]
[1]University College Dublin

We propose a 28-GHz digital wave-locked loop (WLL) achieving ultra-low jitter, low reference spurs, and rapid lock time, with integrated LO phase-shifting (LO-PS) for mm-wave MIMO phased arrays. A waveform detector, using I/Q sampling and arctangent-based phase estimation (PEU) provides full-2π monotonic detection with high resolution. Fabricated in 28-nm CMOS, the WLL achieves 57.2-fs rms jitter and −80.6-dBc reference spur at 27.375 GHz. A full 360° LO-PS range with 2.8° resolution is demonstrated, along with fast ±1.5-GHz frequency hopping and sub-0.59-µs lock time (<74 reference cycles).

Session C5: Isolated and Multi-Domain Power Conversion

1:30 PM, Honolulu 1
Co-Chairs: Qinwen Fan, TU Delft and Wanyeong Jung, Korea Advanced Institute of Science and Technology (KAIST)

1:30 PM
C5.1 4.5-W 5-V-Input DC-DC Converter with Six Isolated 15-V Outputs Achieving 175-kV/µs CMTI Without Signal Isolators, Yichen Zhang[1], Yaogan Liang[1], Michihiro Ide[1], Makoto Takamiya[1]
[1]The University of Tokyo

To reduce the volume of isolated DC-DC converters using bulky transformers and signal isolators for the gate drivers, a compact single input, six isolated output DC-DC converter using a single PCB transformer for both power and signal transfer is proposed. The proposed sensing for TX input power control prevents efficiency degradation even when subjected to large common-mode transient (CMT) without signal isolators. Fabricated in 180 nm BCD process, the proposed converter with >3 kV isolation and 314 mm^3 per output achieved 67% efficiency and CMTI of +175/−167 kV/µs at single 5 V input, six 15 V isolated outputs, and 4.5 W total power.

1:55 PM
C5.2 Class-D LC Oscillator-Based Capacitively Isolated DC-DC Converters with False-Mode Prevention Technique Achieving 91.3% and 43.3% Peak-Efficiencies, Suyang Song[1], Taekwang Jang[1]
[1]ETH Zürich

This paper presents two capacitively isolated DC-DC converters that significantly improve efficiency compared to the state of the art (SoTA). A highly integrated version (IC+ offchip LC) achieves 91.3% efficiency (23% higher than SoTA) and 9.8mA/mm^2 current density. A fully integrated version achieves 43.3% efficiency (36.3% higher than SoTA) and 61mA/mm ^{2} current density. They employ Class-D oscillators as both inverters and rectifiers to minimize switching loss. A charge-injection pair with PLL corrects false-mode oscillation caused by flying inductors and helps stabilize the converter in heavy load conditions.

2:20 PM
C5.3 A 6.8W 91.3%-Peak-Efficiency Coupled Micro-Inductor-Based Bidirectional Isolated Converter with Tri-State Magnetic Charging for Battery Active Cell Equalization, Bin Chen[1], Yike Fang[1], Tao Zhu[1], Yi Fang[1], Yuhang Zheng[1], Longcheng Pi[1], Jingni Hong[1], Tianlu Gao[1], Zheng Qiu[2], Lian Wang[2], Lenian He[1], Xiaopeng Yu[1], Xugang Ke[1]
[1]Zhejiang University, [2]Primechip Semiconductor

A coupled micro-inductor (CMI)-based bidirectional isolated converter for battery cell equalization is proposed. With tri-state magnetic charging, the converter operates in BCM, DCM and LRBM modes in a varying load current, with a f_{SW} from 4.2kHz to 1.25MHz. With active cell equalization, it supports CC/CV balance control with a

maximum 1.6A load current, achieving 99.2% SOC accuracy and 10mV droop voltage. At 12V V_{IN}, peak efficiency is 91.3% in charge state and 90.7% in discharge state respectively.

2:45 PM
C5.4 VSIMO: A DC-DC SIMO-based Current Mismatch Compensator Supporting Per-Output Bidirectional Power Flow for Stacked Voltage Domains Achieving >82% System Efficiency under up to 360-mA Load Mismatch, Yichen Xu[1], Shahreer Ahmed Al Hossain[1], Baoqi Zhu[1], Suhwan Kim[2], Ram K. Krihshnamurthy[2], Xin Zhang[3], Mingoo Seok[1]
[1]Columbia University, [2]Intel, [3]IBM T. J. Watson Research Center

This paper presents VSIMO, a single-inductor multiple-output (SIMO) compensator for a four-layer stacked-voltage-domain system. It features a novel bidirectional power-flow technique that dynamically configures each output to operate in either supply or sink mode to efficiently compensate current mismatch between domains. The 65-nm VSIMO prototype takes a 2.4-V input and regulates three outputs (0.6, 1.2, 1.8V). It achieves >82% system efficiency under up to 360-mA current mismatch, advancing the maximum current mismatch and system efficiency of stacked-voltage-domain systems.

Session C6: Intelligent Vision and Perception Sensors

1:30 PM, Honolulu 2
Co-Chairs: Bruce Rae, STMicroelectronics and Tomohiro Takahashi, Samsung Japan Corporation

1:30 PM
C6.1 A Hybrid Processing-in-Sensor Architecture with a Computational Mosaic Array and Residual-Enhanced Analog Skip Connections for Robust Color Vision, Hyoungjun Kim[1], Ethan Chen[1], Vanessa Chen[1]
[1]Electrical and Computer Engineering, Carnegie Mellon University

This paper presents a hybrid Processing-in-Sensor (PIS) architecture for always-on color vision. The proposed sensor employs a Computational Mosaic Array (CMA) that enables 4b in-pixel convolution while achieving a 55% reduction in photodiode area. Filter-less color computation is enabled by triple-well photodiodes, and the proposed time-domain median pooling achieves 98.8% data reduction. A residual-enhanced analog skip connection is incorporated to improve robustness against analog mismatch and noise. A 126 x 126 CMOS 180nm sensor array consumes 62.08 μW at 40 fps with iFoM of 97.76 pJ/pixel·frame, achieving 97.25% accuracy on face detection task.

1:55 PM
C6.2 A 50M pixels 1.22μm Pitch Pixel Image Sensor with 100dB Single Exposure Dynamic Range by Triple Conversion Gain and LOFIC, Yoshiaki Tashiro[1], Takahiro Toyoshima[1], Masataka Yamane[2], Kenichi Okumura[1], Kota Kimura[1], Maasa Murata[1], Takumi Kimura[1], Kenichiro Anjo[2], Tomohiro Oohama[1], Goki Aoki[1], Shigeto Nishijima[1], Keisuke Yoshiki[2], Satoru Umeki[2], Tetsuhiro Iwashita[2], Tomoya Ueno[2], Yorito Sakano[1], Yusuke Oike[1]
[1]Sony Semiconductor Solutions Corporation, [2]Sony Semiconductor Manufacturing Corporation

This paper presents the world's smallest 1.22 μm pixel pitch, 50-megapixel triple conversion gain sensor with a low read noise of 0.73 e^-rms, a high full well capacity of 250 ke^- and single exposure dynamic range of 100 dB. To achieve this, a new pixel architecture is introduced that incorporates a transfer-gate-type LOFIC and separates the reset and source-follower drains.

2:20 PM
C6.3 A 3.86 TOPS/W TTFS-Based Neuromorphic Image Sensor Featuring Clocked-Recharging SPAD Pixels and Sub-Pixel-Array-Level Processing, Jung-Hye Hwang[1,2], Kieop Hong[1,2], Jubin Kang[1,2], Jong-Beom Kim[1], Yunji Hong[1,2], Seong-Jin Kim[2]
[1]Ulsan National Institute of Science and Technology (UNIST), [2]Sogang University

This paper presents a neuromorphic SPAD image sensor that features in-sensor computation using a spiking neural network (SNN). Clocked-recharging (CR) SPAD pixels inherently encode light intensity into a time-to-first-spike (TTFS) format, enabling sparse, low-latency computation. An 8b exponential counter at each pixel captures spike timing, while a low-power analog counter with a CDAC implements TTFS-based weighted integrate-and-fire (IF) operation. The proposed 3×3 sub-pixel-array (SPA) convolution architecture maximizes in-sensor parallelism and achieves high throughput. The prototype sensor, fabricated in a 110-nm CMOS process, consumes 5.1 mW at 6.67 kfps, achieving a FoM of 2.6 pJ/pixel·fps·filter and an energy efficiency of 3.86 TOPS/W. Using the extracted 1st-layer feature maps (FMs) and a software-based fully connected (FC) layer, a face detection accuracy of 91.5% is achieved.

2:45 PM

C6.4 All-Digital Event-based Vision Sensor with Scene Adaptive Power-Saving Pixels and Three-Layer Neural Network for Object Detection, Houk Lee[1], Jaehyeon So[1], Chanwook Hwang[1], Woosung Chung[1], Kihoon Kim[2], Jaemin Lee[1], Jonghwan Ko[1], Jaehyuk Choi[1,3]

[1]Department of Electrical and Computer Engineering, Sungkyunkwan University, [2]Department of Semiconductor and Display Engineering, Sungkyunkwan University, [3]SolidVue

We present an all-digital synchronous event-based vision sensor (EVS) with adaptive power-saving pixels and in-sensor object-of-interest (OOI) extraction. The 200×128 single-photon avalanche diode (SPAD) array employs three scene-adaptive power saving (PS) schemes saturation-based (SAT-PS), temporal (T-PS), and spatial (S-PS) to suppress power consumption under saturation, short-window abrupt changes, and spatial redundancy. A three-layer hybrid spiking/binary neural network generates 8×8 tile-wise OOI signals to gate informative-event-focused readout. Fabricated in 90 nm, the prototype consumes 9.85 mW at 200 lux, achieving 47.6% pixel- and 21.6% core-power savings, with up to 53.77% pixel saving and 51.4% total saving in composite scenes. The measured throughput is 5.58 kfps and 714 Meps (DAQ-limited). Due to IR-drop induced margin loss, real time OOI detection was validated on an FPGA (Dice score= 0.7452).

Session C7: Accelerators: Generative Models and Specialized Decoding

1:30 PM, Honolulu 3

Co-Chairs: Lindsey Kostas, Qualcomm Inc. and Vinayak Honkote, Intel Corporation

1:30 PM

C7.1 Silicon-Oracle (Soracle): A Multi-Modal Autoregressive Model Accelerator for Context-Aware Assistance on Mobile Platform, Wenao Xie[1], Zhongqi Li[2], Jiaer Chen[2], Jerald Yoo[2], Hoi-Jun Yoo[1]

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

This paper presents Soracle, a multi-modal autoregressive (AR) model accelerator for context-aware assistance. It integrates a mixed-precision exponent-tracked LUT-counter MAC unit and a row-broadcast/column-unicast AR core for EMA-free, head-local attention execution. Weight storage and transfer overhead are further reduced by slice-shared $\log_2$-partition quantization and bandwidth-aligned, stride-adaptive streaming. Soracle also reduces redundant token computation via exponent-aligned dominant-token extraction and dominance-trackable attention process. Fabricated in 28 nm FDSOI with a die area of 20.25 mm^2, Soracle achieves state-of-the-art 83.93 mJ/frame energy consumption and 0.671 s/frame image generation speed.

1:55 PM

C7.2 SPECTRA: An Asymmetric-Precision Speculative Decoding LLM Accelerator with Product Quantization and Reconfigurable Flip-Flop Buffers in 28nm CMOS, HanGyeol Mun[1], Jian Meng[1], Chun-Ting Chen[1], Hyung Joon Byun[1], Yifan He[1], Xiaofeng Hu[1], Sungjin Park[1], Jae-sun Seo[1]

[1]Cornell Tech

We present SPECTRA, an LLM accelerator that supports speculative decoding via prefill/decode/verify phase-wise reconfigurable flip-flop buffers. With no accuracy loss, SPECTRA achieves ~4.1×/3.8×/3.7× and ~7.7×/7.9×/7.3× reductions in external memory access/energy/latency, over micro-scaling INT8 and FP16 baseline models.

2:20 PM
C7.3 A Full-MWPM Surface Code Decoder with On-the-Fly Weight Computation and Cross-Platform Adaptability Achieving 1.9×10^{-6} LER and 20.8-ns Decode Time at 4K, Haoran Lyu[1], Yulong Chen[1], Humiao Li[1], Zhengke Yang[1], Wang Liao[2], Jiamin Li[1], Longyang Lin[1]
[1]Southern University of Science and Technology, [2]Kochi University of Technology

A cryo-CMOS surface code decoder supporting full minimum-weight perfect matching (MWPM) up to distance-21 is presented. The design employs on-the-fly weight computation to reduce weight memory area by 989×, a parallel matching engine for full MWPM, and lock-pair-based speed tuning for cross-platform speed adaptability. The 40-nm test chip achieves a logical error rate of 1.9×10^{-6} with a 20.8-ns decode time at 4K, providing a 3.3× to 612× speedup over prior art while consuming 2.46nJ per decode.

2:45 PM
C7.4 A 122.2μJ/Token Reasoning LLM Accelerator with Reinforcement Fine-Tune Featuring Two-Level KV-Cache Compression and Spike-Driven Predictive Update, Mingxuan Li[1], Wenjie Ren[1], Zhuolin Li[1], Le Ye[1], Tianyu Jia[1]
[1]School of Integrated Circuits, Peking University

In this paper, we introduce an accelerator for reasoning LLM with reinforcement learning fine-tune (RLFT) via three key features: 1) two-level KV cache compression with importance-oriented KV evicting and low-rank decomposition; 2) online reward model scoring using min-hash encoding scheme; 3) spike-driven predictive gradient updating on a hybrid tensor forwarding engine using both digital transpose-free PE and compute-in-memory (CIM) macros. The above innovations enable 2.03× less KV cache, 2.27×less RM scoring latency and 3.79× less weight update. Overall, a 122.2μJ/Token energy efficiency including EMA and 747.2ms/iteration latency are obtained for Qwen3-4B reasoning and RLFT.

Session JFS1: Quantum Computing & Cryo-CMOS

1:30 PM, Tapa 2
Co-Chairs: Jaydeep Kulkarni, University of Texas at Austin and Wei-Chung Lo, Industrial Technology Research Institute (ITRI)

1:30 PM
JFS1.1 Developing a Quantum Computer with Electrically Driven Spin Qubits, Takashi Nakajima[1], Kenta Takeda[1], Akito Noiri[1], Takashi Kobayashi[2], Leon C. Camenzind[1], Tomonori Sekiguchi[3], Yusuke Kanno[3], Yusuke Wachi[3], Takayasu Norimatsu[3], Ryuta Tsuchiya[3], Hiroyuki Mizuno[3], Seigo Tarucha[1,2]
[1]Center for Emergent Matter Science, RIKEN, [2]Center for Quantum Computing, RIKEN, [3]Hitachi, Ltd.

This paper presents the implementation of spin qubits in silicon and the status of development toward scalable quantum computers. High-fidelity control of up to ten qubits is demonstrated in an array of quantum dots fabricated on a Si/SiGe quantum well wafer. Key requirements for maintaining qubit fidelities include material stacks, gate geometry, control electronics, and signal integrity. We discuss our efforts for scaling up the silicon qubit system to realize a practical quantum computer.

1:55 PM
JFS1.2 Detecting Quantum Errors with Hardware-efficient Checks, Ali Javadi-Abhari[1], Simon Martiel[2]
[1]IBM T. J. Watson Research Center, [2]IBM France Lab

Assertions are used extensively in programming to check for program bugs or runtime errors. In quantum computers, runtime errors are many orders of magnitude more likely, making it critical to detect such errors. The problem is exacerbated due to several factors: direct observation of the quantum state destroys it, and the process of detecting errors can itself introduce extra errors. Here, we introduce methods for efficiently detecting errors in quantum circuits by leveraging spacetime codes. These checks introduce little additional error and can be tailored to the underlying quantum hardware, including its qubit connectivity graph and dominant sources of noise. We show the practicality of the approach by successfully preparing maximally entangled states on 50 logical qubits. Such error detection methods form a viable bridge between near-term error mitigation and longer-term error correction in quantum computers.

2:20 PM
JFS1.3 A Cryo-CMOS Fully-Integrated Superconducting Qubit Reflectometry Readout IC with FCNN-Based Adaptive Drift Mitigation and FIR-DSM-Based Stimulus TX, Yanshu Guo[1], Heyue Li[2], Siqi Zhang[2], Jun Shi[2], Shijie Yin[2], Qichun Liu[3], Ning Deng[2], Zhihua Wang[2], Tiefu Li[2], Guoxing Wang[1], Yuanjin Zheng[4], Hanjun Jiang[2]
[1]Shanghai Jiao Tong Univeristy, [2]Tsinghua University, [3]Beijing Academy of Quantum Information Sciences, [4]Nanyang Technological University

This work presents a neural-network-augmented cryo-CMOS readout IC for superconducting quantum computing. By incorporating the FCNN-based qubit state detector, this IC enables real-time adaptation of the qubit readout threshold, resulting in readout drift monitoring and adaptive mitigation. Leveraging the FIR-DSM-based stimulus TX technique, the measured TX power is 2.6 mW, which is more than 1.9× lower than that of prior designs. A quadrature digital-IF reflected RX architecture is implemented with stimulus signal synthesis and reflected-signal processing, resulting in a fully functional readout IC with a minimum power consumption of 10.3 mW. Verified with a transmon qubit at 4 K, this IC achieves a readout fidelity of up to 95.6%.

2:45 PM
JFS1.4 Cryo-CMOS sub-1K 16nm FinFET DAC for Fast Two-Qubit Gates in Spin Qubits, Ilker Polat[1], Ramon W.J. Overwater[1], Bagas Prabowo[1], Stefano Reale[2], Davide Degli Esposti[2], Kenji Capannelli[2], Maximilian Rimbach-Russ[2], Lieven M.K. Vandersypen[2], Masoud Babaie[1], Fabio Sebastiano[1]
[1]Quantum & Computer Engineering, Delft University of Technology, [2]QuTech, Delft University of Technology

This paper presents a cryo-CMOS DAC for driving two-qubit gates in semiconductor spin qubits. Thanks to a current-integrating coarse/fine architecture implemented in a 16-nm FinFET process, the proposed design generates adiabatic waveforms for very fast (<50 ns) high-fidelity operations, while operating at sub-1K temperatures with a small footprint (0.025 mm^2) and low power (275 μW), thus demonstrating its compatibility with qubit co-integration.

Session T2: Backside Power Delivery

1:30 PM, Tapa 1
Co-Chairs: Russell McMullan, AMD and Yoshiaki Kikuchi, Sony Semiconductor Solutions

1:30 PM
T2.1 Improved Backside Contacting for CFET, Cassie Sheng[1], Steven Demuynck[1], Dmitry Batuk[1], Jishnu Ganguly[1], Andy Peng[1], Thomas Chiarella[1], Pierre Eyben[1], David Schippers[1], Roger Loo[1], Subhobroto Choudhury[1], Thomas Dursap[1], Clement Porret[1], Lea Di Donato[1], Heath Huang[1], Serena Iacovo[1], Andrea Mingardi[1], Rami Khazaka[2], Conor Patrick Cullen[2], Daniel Casey[2], Farid Sebaai[1], Pallavi Puttarame Gowda[1], Rajendra Kumar Saroj[1], Anne Vandooren[1], Camila Toledo de Carvalho Cavalcante[1], Maryam Hosseini[1], Jerome Mitard[1], Kevin Vandersmissen[1], Nicolas Jourdan[1], Naveen Reddy[1], Il Gyo Koo[1], Efrain Altamirano Sanchez[1], Alfonso Sepulveda Marquez[1], Lucas Petersen Barbosa Lima[1], Naoto Horiguchi[1], Serge Biesemans[1]
[1]imec, [2]ASM

We report on improvements to the integration flow of backside contacted monolithic CFET devices, demonstrated by enhanced survival rate and bottom pFET performance. A new BSC scheme (code name: delta), enabled by frontside-formed bottom dielectric isolation under the source/drain, reduces the variability in eSiGe:B volume and reduces the contact resistance to the device. Combined with optimized bonding, extreme wafer thinning, backside-formed BDI and BSC etch, pFET SR improves from 45% to 85% versus our previous report.

1:55 PM

T2.2 High-Performance 2nm GAA Standard Cell Architecture Exploiting Backside Power Delivery via Dynamic Area Borrowing, Hyeongyu You[1], Geonwoo Nam[1], Jaehee Cho[1], Minjae Jeong[1], Jisu Yu[1], Jungho Do[1], Hakchul Jung[1], Sanghoon Baek[1]

[1]Foundry Business, Samsung Electronics Co., Ltd.

Aggressive scaling of standard cells to a 5-track (5T) height in the 2nm GAA node creates a critical performance bottleneck by physically confining the maximum active width ($Wactive$). To overcome this limitation, we propose dynamic area borrowing (DAB), a design-technology co-optimization (DTCO) architecture enabled by a backside power delivery (BPD) and direct backside contacts (DBC). By accommodating larger active cells via row-boundary area borrowing, DAB maximizes drive strength within a uniform cell footprint. The proposed DAB achieves a 5.0%–9.0% cell-level delay reduction and a 9.8% $Fmax$ improvement in block-level evaluation on a representative ARM core without area overhead. Ultimately, DAB redefines the performance frontier of logic scaling in sub-2nm nodes, providing an effective solution for next-generation high-performance logic.

2:20 PM

T2.3 CPU Cores in GAA with Backside Power: Silicon-Validated Design Insights, Manjunath Shamanna[1], M Wills[1], J Hazra[1], M Saif[1], B Sell[1], C-H Lin[1], K Fischer[1], W Hafez[1], B Grimm[1], S Joshi[1], A Chatterjee[1], R Ramaswamy[1], E Fetzer[1], R Subramaniam[1], T Lustman[1], J Glassberg[1], S Dhar[1], X Wang[1], R Varada[1], S Saha[1], S Mudanai[1], A Sarwar[1], T Wu[1], E Karl[1], A Murugavel[1], S B Prakash[1], H Markovits[1], P Ranade[1], S Venkataraman[1], A Krishnamoorthy[1], M Doyle[1]

[1]CTaP, Intel Corporation

This paper presents key design learnings from CPU cores implemented in Gate-All-Around (GAA) transistors with Backside Power Delivery (BSPD), contrasting with FinFET technologies. Silicon data reveals improved translation of transistor improvements to core frequency, especially at low voltages (30% frequency at ~0.5V) because of lower Vt and reduced IR drop. GAA-BSPD designs exhibit altered behavior in frequency sensitivity to Vcc and temperature, while posing MCMM (multi-corner multi-mode) tool challenges. Lower Vt and IR drop expose more of the Local Layout Effects (LLE), necessitating enhanced front-end modeling. As lower Vt provides lesser benefits at high voltage in GAA-BSPD designs, we demonstrate through presilicon analysis that adding coarse-pitch metal layers serves as a low-cost, high-Vcc performance booster, improving power and frequency via reduced capacitance and wiring congestion (lower coupling = fewer Signal Integrity issues).

2:45 PM

T2.4 Backside Contact Resistivity Reduction for Nanosheet Technologies Beyond 2nm by Backside Nanosecond Laser Anneal without Impacting the Frontside Cu BEOL, Ruilong Xie[1], Kevin Brew[1], Wonkeun Chung[2], Oleg Gluschenkov[1], Jongjin Lee[2], Chen Zhang[1], Nirmaan Shanker[1], Debarghya Sarkar[1], Haojun Zhang[1], Huai Huang[1], Koichi Motoyama[1], Byounghoon Kim[2], Sangshin Jang[2], WuKang Kim[1], Dominic Picciocca[1], Yasir Sulehria[1], Wei-Shang Lo[1], Michael Gribelyuk[1], Sarabjot Singh[1], Hosadurga Shobha[1], Fabio Carta[1], Richard Johnson[1], Shay Reboh[1], Shogo Mochizuki[1], Erik Milosevic[1], Tao Li[1], Kishore Natarajan[1], Su Chen Fan[1], Tenko Yamashita[1], Kisik Choi[1], Dechao Guo[1], Kang-ill Seo[1], Huiming Bu[1]

[1]IBM Research, [2]Samsung

This paper highlights the challenges and solutions associated with forming backside source/drain (S/D) contact with low contact resistivity (RhoC). The backside nanosecond laser anneal (NLA) process window is identified by thermal simulation and experimental verification of resistance and reliability impact on the frontside Cu-based

back-end-of-line (BEOL). Optimal NLA energy dose (ED) is implemented on nanosheet transistors achieving low RhoC backside contact.

Technical Sessions Block 3 (3:25 PM – 5:30 PM)

Session C8: Compute-In-Memory

3:25 PM, Tapa 3
Co-Chairs: Yan Li, Sandisk Corp. and Masanao Yamaoka, Hitachi, Ltd.

3:25 PM
C8.1 A 2nm 234.4 TOPS/W and 511.9 TOPS/mm² Digital Computing-in-Memory Compiler with Multiple MAC Units per Weight and Multiple Data Format Support, Hidehiro Fujiwara[1], Brian Crafton[1], Haruki Mori[1], Je-Min Hung[1], Wei-Chang Zhao[1], Ashwin Sanjay LeLe[1], Bo Zhang[1], Chao-Kai Chuang[1], Hiroaki Ishii[1], Cheng-En Lee[1], Vineet Joshi[1], Xiaochen Peng[1], Xiaoyu Sun[1], Shu-Huan Hsu[1], Yao-Yi Liu[1], Takeshi Hashizume[1], Yu-Hao Hsu[1], Win-San Khwa[1], Saman Adham[1], Yu-Der Chih[1], Yen-Huei Chen[1], Hung-Jen Liao[1], Kerem Akarvardar[1], Meng-Fan Chang[1], Tsung-Yung Jonathan Chang[1]
[1]Taiwan Semiconductor Manufacturing Company (TSMC)

This paper presents a Compute-in-memory (CIM) compiler featuring multiple data format and multiple MAC units per weight, supporting flexible data format, better utilization and throughput. We implemented various configurations in a test chip with our 2nm nanosheet technology node and test chip measurement results show V_{MIN} < 0.38V, 234.4 TOPS/W at 0.5V and 511.9 TOPS/mm² at 1.2V.

3:50 PM
C8.2 A 0.0257 mm², 1.16 µW, Mask-Efficient KWS SoC with Hybrid Cap-ROM-IMC/SRAM-NMC and Transfer Learning for Flexible Multi-Keyword Recognition, Fei Tan[1,2], Zhongyu Zhao[1], Wei-Han Yu[1], Ka-Fai Un[1], Jiawei Xu[1], Yingzhe Hu[3], Mahsa Shoaran[2], Rui P. Martins[1], Pui-In Mak[1]
[1]University of Macau, [2]EPFL, [3]Spacetouch

Scaling always-on keyword spotting (KWS) system to larger vocabularies increases model size, driving higher memory and compute demands that push SoCs >100 µW, with significant off-chip loading energy and area overhead. This work presents the first KWS SoC featuring a hybrid Cap-ROM-based in-memory computing (Cap-ROM-IMC)/5T-SRAM-based near-memory computing (SRAM-NMC) architecture with a transfer-learning based model supporting 20 keywords. Fabricated in 28nm, the SoC achieves 92.1% accuracy on GSCD-v2 with the lowest reported power (1.16µW) and highest ROM memory density (2372.6KB/mm²) among prior arts.

4:15 PM
C8.3 A 28nm 44.15TFLOPS/W and 1.57TFLOPS/mm² Joint-Alignment Floating-point CIM Macro with ISO26262 ASIL-D Safety Level for Autonomous Driving, Zhen He[1], Yiqi Wang[1], Zhiheng Yue[1], Zihan Wu[1], Yang Hu[1], Fengbin Tu[2], Shouyi Yin[1]
[1]Tsinghua University, [2]Hong Kong University of Science and Technology (HKUST)

This work presents an FP-CIM for AI model acceleration in autonomous driving system with three key features: 1) A joint-alignment CIM structure performs exponent-mantissa joint-alignment to reduce 3.97x accuracy loss caused by alignment truncation by repurposing the inherent-invalid bits between the hidden bit and mantissa leading-one without extra bit-width extension. 2) A hybrid adder tree features first-level 6T and other-level 10T full adders, which precompute and reuse the common XOR subexpression shared between Sum and Cout, achieving average 1.54x area efficiency improvement. 3) A unified all-inclusive SDC decoder directly protect the final MAC result against SDC to reduce safety risks, meeting ISO26262 ASIL-D safety level with 99.50% fault coverage, while reducing 48.44% area and 60.09% power overhead over the conventional separate memory-compute protection for same safety level.

4:40 PM

C8.4 A 28nm CMOS-Integrated VCMA-MTJ Dual-Function Macro for Deterministic In-Memory Computing Synapses and Stochastic Poisson Neurons in Hybrid Vision, Yaoru Hou[1,2], Zhihua Xiao[1,2], Xuezhao Wu[1,3], Shuyu Wang[4], Zheng Zhu[3], Di Wu[3], Jiayong Wang[3], Yuesheng Li[3], Hao Cai[4], Qiming Shao[1,2]

[1]Department of Electrical and Computer Engineering, Hong Kong University of Science and Technology (HKUST), [2]AI Chip Center for Emerging Smart Systems, [3]Institute of Integrated Circuits, Shanghai University, [4]School of Integrated Circuits, Southeast University

Unlike current-driven spintronic devices, Voltage-Controlled Magnetic Anisotropy (VCMA) Magnetic Tunnel Junctions (MTJs) utilize electric-field-induced switching to overcome energy bottlenecks and enable tunable stochastic behavior. This work presents a 28 nm CMOS-integrated VCMA-MTJ dual-function macro that employs a single spintronic device to support both deterministic and stochastic neuromorphic primitives. The integrated VCMA-MTJs function as deterministic in-memory-computing (IMC) synapses for signed-weight storage and array-level MAC operations, while their voltage-tunable probabilistic switching enables the implementation of stochastic Poisson neurons. The fabricated macro is validated in a hybrid ANN/SNN hardware platform for frame–event hybrid vision detection, achieving 92.5% inference accuracy and 411 TOPS/W energy efficiency. These results demonstrate the potential of VCMA-MTJ devices as a promising solution for low-power, high-accuracy neuromorphic computing.

5:05 PM

C8.5 A 40nm 24.85-to-80.59 TFLOPS/W FP8 Computing-in-Memory Macro with Outlier-Aware Hierarchical-Alignment for Edge Transformers, Zhiwei Zhou[1], Tong Hu[1], Jia Chen[1], Jiancong Li[2], YuYang Fu[1], Yi Li[1], Xiangshui Miao[1]

[1]School of Integrated Circuit, Huazhong University of Science and Technology, [2]Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST)

This paper presents a digital Compute-in-Memory (CIM)-based macro tailored for edge Transformer acceleration. It has three features: 1) A Hierarchical Dual-Exponent Alignment scheme that separates outliers from normal values, achieving near-ideal accuracy with minimal overhead. 2) A Sign-Magnitude Input Serial Pipelined Aligner (SMSPA) reduces area and power by 41.1% and 34.1%. 3) A Shift-Concat based Sign-Inverse Adder Tree (SSAT) saves accumulation area and power by 25.9% and 18.2%. The macro achieves 24.85-80.59 TFLOPS/W in FP8. Measurements demonstrate up to 6.66 × improvements in energy efficiency, compared to state-of-the-art post-align FP-CIM schemes, while maintaining high accuracy on different tasks.

Session C9: Neural-Interface Processors & SoCs

3:25 PM, Honolulu 2

Co-Chairs: Emmanuel Quevy, Probius and Jerald Yoo, Seoul National University

3:25 PM

C9.1 A 64-Channel, Low-Power End-to-End Transformer-Based BCI SoC for Efficient Motion Intention Decoding, Fei Tan[1], Raphael Berner[1], Zhongyu Zhao[2], Yuhan Xie[1], Amirmahdi Joudi[1], Yiheng Fu[1], Icare Sakr[1,3], Thibault Collin[1,3], Vincent Rouanne[1,3], Henri Lorach[1,3,4], Ka-Fai Un[2], Pui-In Mak[2], Jocelyne Bloch[1,3,4,5], Gregoire Courtine[1,3,4], Mahsa Shoaran[1]

[1]EPFL, [2]University of Macau, [3]Neurorestore, [4]University of Lausanne, [5]Lausanne University Hospital

Real-time brain-computer interfaces (BCIs) demand fully on-chip neural decoding with tight power, area, and latency budgets. This work presents the first fully-integrated transformer-based BCI SoC that performs energy-efficient motion intention decoding from ECoG signals. The chip integrates a 64-channel analog front-end (AFE) and a distilled, light-weight transformer-based decoder featuring customized 7T-SRAM computing-in-memory (CIM) macros with hybrid sign-magnitude and two's-complement (SM-2C) computation. Fabricated in 65nm, the

decoder achieves a 67.8% accuracy on a 6-class upper-limb motion-intention dataset, consuming 47μW with 4.7mm^2, while the whole SoC occupies 10.1mm^2.

3:50 PM
C9.2 An In-Ear Sleep Modulation SoC Featuring a CNN–LSTM Accelerator with Long-Kernel Memory Reuse and Runtime Dynamic Quantization, Yaqian Xu[1], Xinyu Chen[1], Yuhan Hou[1], Jianxiong Xu[1], Hao You[1], Yu Huang[1], Bowen Liu[1], Ashley Hung[1], Andrew G. Richardson[2], Roman Genov[1], Xilin Liu[1]
[1]University of Toronto, [2]University of Pennsylvania

This work presents a 65 nm system-on-chip (SoC) for closed-loop in-ear sleep modulation. The SoC performs low-noise in-ear EEG sensing, CNN–LSTM–based sleep stage classification, and phase-specific auditory stimulation under a 1 mW power budget. A dedicated accelerator is developed with a memory reuse technique to efficiently process long temporal kernels required for low-frequency EEG feature extraction. Runtime mixed-precision dynamic quantization is incorporated, reducing model size by 3.88× with only a 0.28% loss in accuracy. A sensitivity of 98.3% is achieved for deep sleep detection, benchmarked on the public MASS dataset. Integrated with a 0.65 μV chopper-stabilized low-noise amplifier, delta-sigma ADCs achieving a 172 dB Schreier FoM, and on-chip noise generation and an audio amplifier, the chip enables effective closed-loop sleep modulation, paving ways for future ear-based healthcare applications.

4:15 PM
C9.3 TARNA: A 1.18-μW/Channel Topology-Aware Residual Recurrent CIM Processor for Zero-Shot Seizure Detection, Huanshihong Deng[1], Yiheng Fu[1], Negin Safari[1], Cong Huang[1], Fei Tan[1], Mahsa Shoaran[1]
[1]EPFL

Zero-shot seizure detection avoids patient-specific training but remains challenging under strict power constraints and patient variability. This work presents TARNA, a topology-aware recurrent network accelerator implemented as an ultra-low-power computing-in-memory (CIM) zero-shot seizure detector. TARNA combines graph convolutions with masked residual recurrent updates to eliminate nonlinear gating, while a sparsity-aware CIM macro further reduces energy. Fabricated in 65-nm CMOS, TARNA achieves 100% zero-shot event-based sensitivity and a 0.048/hr false-alarm rate on the CHB-MIT dataset, and an AUROC of 0.88 on the TUSZ dataset, operating at the lowest reported power of 22.5μW.

4:40 PM
C9.4 A 16-Channel Neural Recording IC with Spike-Driven Adaptive-Basis Compressive Sensing, Sanghoon Lee*[1], Soohyun Yun*[2,3], Seokho Yoon[1], Kyungmin Kang[1], Taejune Jeon[1], Sang Hoon Park[1], Ju Young Lee[1], Won Gi Chung[1,4], Jang-Ung Park[1,4], Ki Jun Yu[1,5,6], Joonsung Bae[2,3], Youngcheol Chae[1,3]
[1]Yonsei University, [2]Kangwon National University, [3]XO Semiconductor, [4]IBS, [5]KIST, [6]POSTECH

This paper presents a 16-channel neural recording IC that achieves massive data-rate reduction through a spike-driven adaptive-basis compressive sensing (CS). To address wireless data-rate bottlenecks in high-density recording, a two-stage compression scheme that integrates on-chip spike detection with CS encoding is proposed. The adaptive dictionary learning tracks non-stationary neural signals in real-time, maintaining high reconstruction fidelity even under significant temporal drift. Furthermore, a wide-DR analog front-end ensures robust tolerance to large transient stimulation artifacts. Measured results demonstrate 168× data-rate reduction with 96.8% classification accuracy, slashing the per-channel data-rate to 1.92kbps, while the CS block consumes only 2.2μW/ch. This enables a highly efficient and scalable interface for artifact-tolerant, long-term implantable neural recording.

Session C10: Building Blocks for Multimodal Sensors

3:25 PM, Honolulu 3
Co-Chairs: Matteo Perenzoni, Sony Semiconductor Solutions Corporation and Hirotomo Ishii, Toshiba Electronic Devices & Storage Corporation

3:25 PM

C10.1 A 75nW BJT-based Temperature Sensor with ±0.1 °C Inaccuracy from -40°C to 85°C, Jida Peng[1], Nandor G Toth[1], Kofi A. A. Makinwa[1]

[1]Microelectronics, Delft University of Technology

This paper describes a BJT-based temperature sensor that combines a dual-mode front-end (DMFE), and a low-power switched capacitor (SC) ΔΣ modulator with a 10-tap FIR-DAC. After a low-cost 1-point trim, it achieves ±0.1°C (3σ) inaccuracy from -40°C to 85°C while dissipating only 75nW. This is 3× lower than prior precision BJT-based sensors, and 2-3× more accurate than prior low-power MOS-based sensors.

3:50 PM

C10.2 A MEMS-Based Throughflow Multimodal Sensor Interface IC for Real-Time Fluid Parameter Monitoring, Arthur Campos de Oliveira[1], Jarno Groenesteijn[2], Zhong Tang[1], Hui Jiang[1], Joost Lötters[2,3], Remco Wiegerink[3], Kofi Makinwa[1]

[1]TU Delft, [2]Bronkhorst BV, [3]University of Twente

This paper describes a MEMS-based throughflow multimodal sensing system that integrates a Coriolis flow sensor with pressure and temperature sensors to enable simultaneous measurement of mass flow, fluid density, input/output pressure, temperature, and viscosity. Implemented using a single MEMS chip and a CMOS readout ASIC, it achieves state-of-the-art performance with the highest reported resolution and zero stability among MEMS Coriolis flow sensors. Compared to stainless-steel Coriolis sensors, it exhibits >22× lower temperature drift, while maintaining relative errors <±5% for most liquids and gases.

4:15 PM

C10.3 A SPAD-Based iToF Flash LiDAR Sensor with a Coarse-Fine Operation Scheme Robust Against Strong Background Light, Dongseok Cho[1,2], Hyun-Seung Choi[1], Injun Park[3], Yoondeok Na[1], Seonghun Yang[1], Minkyoung Yang[1], Yerin Youn[1], Jimin Cheon[3,4], Myung-Jae Lee[1], Youngcheol Chae[1,3]

[1]School of Electrical and Electric Engineering, Yonsei University, [2]Semiconductor R&D Center, Samsung Electronics Co., Ltd., [3]XO Semiconductor, [4]Kumoh National Institute of Technology

This paper presents a SPAD-based indirect ToF LiDAR sensor achieving a record 94 m range with 11 cm depth precision at 35 fps. To achieve high depth precision and robust operation under strong background light (BGL), we propose a coarse-fine distance measurement scheme. The coarse operation utilizes a binary search to progressively confine the target distance, while the fine operation employs a four-phase method to suppress BGL and ensure high-precision ranging. Fabricated in a 110 nm BSI CIS process, the 100×75 sensor achieves a high fill factor of 38.7% and a photon detection probability (PDP) of 42.1% at 850 nm. Experimental results demonstrate robust performance under 90 klx BGL.

4:40 PM

C10.4 A 3nm Variation-Tolerant On-Die Current Sensor for Power-Limits Management, Fikre Gebreyohannes[1], Yimai Peng[2], Sohail Asghar[1], Khaled ElGammal[1], John MacHale[1], Keith O'Donoghue[1], Keith Bowman[2]

[1]QT Technologies Ireland Limited, Cork, Ireland, Qualcomm, [2]Qualcomm Technologies, Inc., Raleigh, NC, USA, Qualcomm

A CPU core globally distributed head switch (GDHS) integrates an on-die resistor-less current sensor (I-Sensor) to measure core current (I_{IN}) for power-limits management (PLM). The I-Sensor features a replica head switch (RHS) to track the GDHS resistance across process, voltage, and temperature (PVT) variations and an analog front-end (AFE) circuit to measure the voltage drops across GDHS and RHS. With a nearly constant RHS reference current (I_{REF}) across PVT variations, the I-Sensor calculates I_{IN} from a ratio of the AFE measured GDHS and RHS voltage drops, thus cancelling out the GDHS and RHS resistance variations and AFE gain error, resulting in a fast and accurate current measurement, while minimizing post-silicon calibration. From silicon data, the I-Sensor demonstrates a

1.2MHz bandwidth and a 92.8% accuracy for ±3 standard deviations for a 100× dynamic current range (0.12A to 12A).

5:05 PM
C10.5 Unified Digital Thermal-Voltage Sensor for Thermal Management in Intel 18A/Intel 3, Shanshan Xie[1], Gregory K. Chen[1], Phil C. Knag[1], Hikmet Seha Ozturk[2], Visvesh S. Sathe[2], Carlos Tokunaga[1], James W. Tschanz[1]
[1]Intel Corporation, [2]Georgia Institute of Technology

A unified digital thermal and voltage sensor achieves inaccuracies of 1.9°C/1.3mV (Intel 3) and 3.1°C/2.1mV (Intel 18A) in a hybrid bonded 3DIC DNN processor using a high-volume manufacturing calibration flow. Aging compensation measures always-on and power-gated sensors to generate scaling factors that maintain the sensor's accuracy over its lifetime. RISC-V cores process sensor data to compute real-time thermal and voltage maps for a DNN workload, and thermal throttling response reduces temperature guardbands by maintaining a reference temperature.

Session T3: 2D Channel Materials

3:25 PM, Tapa 1
Co-Chairs: Chris Hinkle, Notre Dame and Toshifumi Irisawa, AIST

3:25 PM
T3.1 Scaling V_{DD} for P-type 2D-Material Channel Transistor: What Non-Idealities Are Still to Be Solved, Chao-Ching Cheng[1], Ang-Sheng Chou[1], Edward Chen[1], Ming-Yang Li[1], Yu-Wei Hsu[1,2], Chia-Chin Cheng[1], Seong Rae Cho[3], Haomin Liu[4], Po-Sen Mao[1], Wen-Chia Wu[1], Shuer Ku[1], Wei-Sheng Yun[1], Chih-Piao Chuu[1], Wei-Yen Woon[1], Lain-Jong Li[4], Vincent Tung[3], Chih-I Wu[2], Iuliana Radu[1], Min Cao[1]
[1]Corporate Research, Taiwan Semiconductor Manufacturing Company (TSMC), [2]Graduate School of Advanced Technology, National Taiwan University, [3]Department of Chemical System Engineering, School of Engineering, The University of Tokyo, [4]Department of Materials Science and Engineering, National University of Singapore (NUS)

This study investigates the effect of non-ideal factors in WSe_2 2D-material channel transistors, including interface traps (D_{it}), contact resistance (R_C), and band-tail states (S_{BT}), on the threshold voltage (V_T). For achieving the scaled V_{DD} = 0.75 V operation, reduction of multiple non-idealities is required beyond current state: **(i)** reducing D_{it} below 1E12 $cm^{-2}eV^{-1}$ preserves the position of V_{TCC} = - 0.1 V (constant-current V_T); **(ii)** suppressing the S_{BT} through gate dielectric engineering, with scaled *EOT* below 1 nm, is required for setting correct V_{TCC} and V_{TGM} (transconductance-defined V_T) positions; **(iii)** minimizing R_C to 100 W-mm ensures the (V_{TGM} – V_{TCC}) gap below 0.28 V. The study combines these findings and offers insights to design the desired V_T positions at scaled V_{DD} for 2D materials.

3:50 PM
T3.2 Enhancement-Mode Monolayer WSe_2 pFETs with Record I_{ON}/I_{OFF} at Contact Length of 30nm, Yuan-Chun Su[1,2], Ting-Hua Wei[1], Chenxi Lei[3], Sunhao Shi[3], Yu-Wei Hsu[1,4], Tzu-Chuan Su[1], Sin-Yue Lee[1], Hong-De Lin[1], Shao-Heng Chen[4], Hung-Li Chiang[1], Chao-Chen Lin[5], Bo-Heng Liu[6], Chien-Ying Su[6], Chi-Chung Ke[6], Shu-Jui Chang[2], Chih-I Wu[4], Chenming Hu[2], Jui-Han Fu[3], Vincent Tung[3], Tsung-En Lee[1]
[1]Dept. of Microelectronics, National Yang Ming Chiao Tung University, [2]FSTRC, National Yang Ming Chiao Tung University, [3]Dept. of Chemical System Engineering, The University of Tokyo, [4]Graduate School of Advanced Technology, National Taiwan University, [5]Taiwan Semiconductor Research Institute (TSRI), National Institutes of Applied Research, [6]National Center for Instrumentation Research, National Institutes of Applied Research

We demonstrate enhancement-mode monolayer (1L) WSe_2 pFETs, under aggressive contact-gate pitch (CGP) scaling while maintaining competitive I_{ON} and record high ON-OFF ratio of 10^9. The device performance is improved by transport in single-domain CVD-grown 1L-WSe_2, controllable UV-ozone (O_3) p-doping, and selective contact doping by TeO. At I_{OFF} = 0.3pA/μm defined at V_G = 0V, the 1L-WSe_2 pFET under channel length (L_{CH}) of 50nm exhibits a record $I_{D,Max}$ of 0.9mA/μm at V_D = −1V, a peak transconductance (G_m) of ~570 μS/μm, a low DIBL of ~50

mV/V and a contact resistance (R_C) of ~407 Ω·μm. With further aggressive contact length (L_C) scaling, enhancement-mode operation is preserved with an I_{ON}/I_{OFF} ~1.8×10^8 and a transfer length (L_T) below 100 nm. These results validate the scalability of 2D p-channel devices pursuing optimal performance, power, area (PPA) for future CMOS logic technologies.

4:15 PM

T3.3 Performance Demonstration of Monolayer MoS_2 Transistors at Cryogenic Temperatures Down to 6 K for Ultralow-Power Electronics, Zhongyunshen Zhu[1], Aijia Yao[1], Patricia Sun[2], Yixuan Jiao[1], Mohamed Mohamed[3], Ji-Hoon Park[2], Jingxian Li[2], Jing Kong[1], Jiadi Zhu[2], Tomás Palacios[1]
[1]Department of EECS, Massachusetts Institute of Technology, [2]CDimension Inc, [3]Department of MSE, MIT

We demonstrate systematic DC and noise characterization of monolayer MoS_2 transistors over a wide temperature range from 300K to 6K. Temperature-dependent trends in V_T, on and off currents, and subthreshold swing (SS) are analyzed through both individual devices and statistics. At 6K, the devices exhibit significantly improved cryogenic performance, achieving: 1) SS_{min} < 30mV/dec, 2) on/off ratio > 10^8, and 3) twofold enhancement in peak transconductance efficiency compared to 300K. A temperature-dependent compact model calibrated to the measurements is employed to simulate a scaled MoS_2 FET with ultrathin EOT, predicting ultralow-power (ULP) switching with SS = 23mV/dec and reduced energy-delay product (EDP) relative to 90nm Si and 22nm FD-SOI. In addition, MoS_2 FETs show strongly suppressed low-frequency noise at cryogenic temperatures, outperforming Si counterparts. These results highlight the promise of MoS_2 for ULP cryogenic electronics.

4:40 PM

T3.4 Atomic Layer Deposition of Epitaxial Semimetal Contacts for 2D Transistors, Huije Ryu[1], Jongwon Lee[2], Sangmin Jeon[3], Gwangsik Jeon[3], Yoonhoo Ha[1], Young Min Lee[1], Dongmin Kim[1], Hyun Mi Lee[1], Joung Eun Yoo[1], Eun-Kyu Lee[1], Eun-Hyoung Cho[1], Jaehyoung Lee[1], Yongjoon Shin[2], Chanyoung Yoo[4], Sungnam Lyu[2], Seunghun Lee[2], Sangjin Hyun[2], Yongsung Kim[1], Cheol Seong Hwang[3], Do-Sun Lee[2], Kyung-Eun Byun[1], Changhyun Kim[1]
[1]Samsung Advanced Institute of Technology, [2]Semiconductor R&D Center, Samsung Electronics Co., Ltd., [3]Department of Materials Science and Engineering, Seoul National University, [4]Department of Materials Science and Engineering, Hongik University

This work reports 8-inch wafer-scale epitaxial atomic-layer deposition (ALD) Sb contact for two-dimensional (2D) transistors using intermediate-enhanced ALD (IE-ALD) and epitaxial de-chalcogenation, achieving a record-low ALD contact resistance (R_C) of 0.6 kOhm·μm for MoS_2. Leveraging ALD conformality, we demonstrated for the first time an epitaxial contact-all-around (C-AA) structure, extending the contact interface beyond conventional planar geometries. Density functional theory (DFT) calculations show that the charge-transfer efficiency of the C-AA structure exceeds that of conventional Sb top-contacts.

5:05 PM

T3.5 Wafer-scale Two-channel Monolayer-MoS_2 Nanosheet FETs with Record-high I_{on}>2 mA/μm Enabled by Sub-1 nm CET High-κ integration, Haofei Zheng[1,2], Lingqi Li[1], Wanqing Meng[3], Yi-Hsin Tu[4], Chenyang Li[3,5], Sujuan Ding[6], Ziming Fu[3], Mingxi Chen[2], Dongzhi Chi[2], Chuanhong Jin[6], Gengchiau Liang[4], Lain-Jong Li[3], Kah-Wee Ang[1]
[1]Department of Electrical and Computer Engineering, National University of Singapore, [2]Institute of Materials Research and Engineering, Agency for Science, Technology and Research (A*STAR), Singapore, [3]Department of Materials Science and Engineering, National University of Singapore, [4]Industry Academia Innovation School, National Yang Ming Chiao Tung University, [5]Nexstrom, [6]State Key Laboratory of Silicon and Advanced Semiconductor Materials, Zhejiang University

Multi-channel Nanosheet FETs (NSFETs) are the definitive architecture for maximizing drive current (I_{on}) within a scaled device footprint. In this work, we demonstrate the first wafer-scale (2-inch) two-channel (2-ch) MoS_2 NSFETs, enabled by a CMOS-compatible, interfacial-layer-free integration of high-κ HfO_x on monolayer MoS_2. An aggressively scaled capacitance equivalent thickness (CET) of 0.71 nm is achieved and the best-performing device exhibits a record-high I_{on}>2 mA/μm and a SS of 101 mV/dec at a channel length of 50 nm.

Session T4: Oxide Semiconductors

3:25 PM, Honolulu 1
Co-Chairs: Asif Khan, Georgia Tech and Jin Cai, TSMC

3:25 PM
T4.1 Different Low-Frequency (1/f) Noise Origins in Atomically Thin ALD In_2O_3 TFTs Determined by the Quantum-Confinement Effect, Sumi Lee[1], Jian-Yu Lin[1], Chang Niu[1], Chun-An Shih[1], Linjia Long[1], Muhammad A. Alam[1], Peide D. Ye[1]
[1]Electrical and Computer Engineering, Purdue University

Low-frequency (1/*f*) noise in amorphous-oxide-semiconductor (AOS) TFTs increases as device area scales down, but its origin is obscured by misleading S_{ID}/I_D^2–V_{OV} slopes. Using ultrathin atomic-layer-deposited (ALD) In_2O_3 TFTs that span channel degeneracy and contact barriers (2 nm: degenerate, $\phi_{SBH} < 0$; 1.2 nm: nondegenerate, $\phi_{SBH} > 0$), we combine temperature-dependent S_{VG} with Dutta-Horn analysis and cleanly separate two regimes: the 2 nm devices follow tunneling-assisted carrier-number fluctuations (ΔN) ($\alpha \approx 1$, nearly flat $D(E_a)$), whereas the 1.2 nm devices exhibit $\alpha < 1$ with a redistributed $D(E_a)$, implicating band-tail-coupled ΔN. Finally, *bias*- and L_{ch}-scaling reveal a transition to contact-noise dominance at high V_{OV} only in the 1.2 nm devices.

3:50 PM
T4.2 Hydrogen Overflow as a Novel Negative V_{th} Shift Mechanism in PBTI Universal to Planar and Vertical IGZO FETs, Reika Ichihara[1], Akinori Kamiyama[1], Takao Kosaka[1], Hiroki Tokuhira[1], Nobuyoshi Saito[1], Masaya Toda[1], Kazuhiro Matsuo[1], Ken-ichi Haga[1], Takamasa Hamai[1], Shoichi Kabuyanagi[1], Takayuki Tsukamoto[1], Akihiro Kajita[1], Masumi Saitoh[1], Shosuke Fujii[1]
[1]Kioxia Corporation

We identified novel Positive-Bias-Temperature-Instability (PBTI) mechanism in InGaZnO gate stack by separating multiple PBTI factors using in-depth *C-V* analysis and atomistic calculations. In addition to well-known electron trapping and donor generation, accumulation of proton in the gate insulator triggered by hydrogen overflow from the channel was found to play a critical role in negative V_{th} shift of PBTI for the first time. This mechanism was also verified with scaled vertical channel transistors, and we demonstrated over 10 years of lifetime via process optimization based on a solid understanding of PBTI mechanism.

4:15 PM
T4.3 First Demonstration of Top-Gated Indium Oxynitride (InON)-FET with Superior Reliability and Contact Properties (I_{on} = 1.2 mA/μm, L_T = 17 nm, $\rho_c = 4\times10^{-8}$ Ω·cm²), Minjong Lee[1], Doo San Kim[1], Soham Shirodkar[1], Thi Thu Huong Chu[1], Dushyant M. Narayan[1], Dan N. Le[1], Mingyeong Jo[1,2], Jae Kyeong Jeong[2], Youngbae Ahn[3], Jayong Kim[3], Seung Wook Ryu[3], Jiyoung Kim[1]
[1]University of Texas at Dallas, [2]Hanyang University, [3]SK Hynix

This work presents InN-derived indium oxynitride (InON) as a contact-efficient, reliability-improved In-based oxide channel for back-end-of-line (BEOL)-compatible FETs. Top-gated (TG) devices deliver an on-current (I_{on}) as high as 1.2 mA/μm. With Ni contacts, the devices achieve an ultra-low specific contact resistivity (ρ_c) of 4×10^{-8} Ω·cm² (contact resistance, R_c = 0.015 Ω·cm) and a short transfer length (L_T) of 17 nm, enabling potential contact scaling with minimal contact-induced series resistance. The InON devices further exhibit improved threshold voltage (V_{th}) stability under bias stress compared with In_2O_3-based counterparts.

4:40 PM
T4.4 Breaking Mobility-V_{th}-R_{SD} Trade-off in Oxide Semiconductor FETs: Novel p-NiO Junction Engineering, Weibing Hao[1], Jiawei Xie[1], Leming Jiao[1], Jingguang Lu[1], Chen Sun[1], Yuxuan Wang[1], Kai Ni[2], Gengchiau Liang[3], Xiao Gong[1,4]

[1]National University of Singapore (NUS), [2]University of Notre Dame, [3]National Yang Ming Chiao Tung University, [4]Agency for Science, Technology and Research (A*STAR), Singapore

For the first time, junction engineering using a p-type NiO cappinglayer is implemented in ITO FETs, simultaneously enabling a significantpositive threshold voltage (Vth) for enhancement-mode operation, preservationof channel mobility (μeff), and reduction of source/drain series resistance (RSD),capabilities that are unattainable through conventional channel thickness (TCH)engineering. By modulating the carrier concentration of p-type NiO, a 2 Vpositive shift in Vth is achieved together with a field-effect mobility of 34cm^2/V·s and a 1.5× enhancement in on-state current (ION) in long-channeldevices (LCH = 10 μm). Enhancement-mode operation is also realized in short-channel devices (LCH = 150 nm), delivering a Vth of 0.6 V and a 2.9×improvement in ION compared with control devices, owing to the combinedbenefits of mobility preservation and RSD reduction. Moreover, relative todevices without NiO, the proposed devices exhibit a much lower subthresholdswing of 28 mV/dec at 77 K, along with excellent wafer-scale uniformity.

5:05 PM
T4.5 Experimental Study on Carrier Transport in Amorphous and Poly InGaOx FETs – Hall Measurement, Temperature Dependence, Velocity Saturation, Xingyu Huang[1], Anlan Chen[2], Kota Sakai[2], Saraya Takuya[2], Toshiro Hiramoto[2], Mutsunori Uenuma[3], Takanori Takahashi[4], Yukiharu Uraoka[4], Masaharu Kobayashi[2]
[1]The Univ. of Tokyo, [2]The University of Tokyo, [3]AIST, [4]Nara Institute of Science and Technology

We have systematically investigated low-field carrier transport in amorphous and poly oxide semiconductor (OS), InGaOx (IGO) FETs by employing Hall measurement for strict separation of carrier concentration (Ns) and mobility. High mobility OS FETs exhibit distinct temperature dependence at low and high carrier concentration. Poly IGO FETs show more phonon-limited scattering behavior than amorphous IGO FETs. We also investigated high-field carrier transport with short channel FETs. Velocity saturation behavior is obtained even with parasitic resistance correction, which can limit the device performance for scaling OS FETs. These findings are essential for physics-based device modeling of scaled OS FETs.

Session TFS1: 3D Memory (Flash/HBM)

3:25 PM, Tapa 2
Co-Chairs: Alessandro Calderoni, Micron and Hang-Ting Lue, Macronix International Co., Ltd.

3:25 PM
TFS1.1 3D Flash Memory Technology for the AI Era, Ryota Katsumata[1], Seiji Maeda[1], Masahiro Inohara[1], Hirofumi Inoue[1]
[1]Kioxia Corporation

Generative AI's widespread use relies on Large Language Models (LLMs) and significant data storage supported by Solid State Drives (SSDs). Key optimizations for 3D flash memory involve balancing latency, parallelism, and bandwidth. Technologies such as molybdenum word lines and tungsten bit lines with an air-gap improve latency. Packaging advancements, including vertical wire bonding, redistribution layer (RDL), and through-silicon via (TSV), enhance interface speed and power delivery, which are crucial for handling large AI model data and enabling rapid data access.

3:50 PM
TFS1.2 High Bandwidth Memory 4 (HBM4) Package Development Challenges, Jaesik Lee[1], Yeontaek Hwang[1], Jongkyu Moon[1], Juheon Yang[1], Kyoungmoo Harr[1], Minseok Jang[1], Jinwoo Choi[1]
[1]SK Hynix

We report package development challenges associated with HBM4 using Mechanical 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.

4:15 PM

TFS1.3 Demonstration of Cell Multi-Bonding (CMB) Technology for Future Vertical NAND over 1k-Layer, Jeehoon Han[1], Jaeryong Sim[1], Seogoo Kang[1], Junghoon Jun[1], Jihwan Park[1], Donghwan Kim[1], Changyeon Yu[1], Jongho Lim[1], Yusik Choi[1], Changmin Kim[1], Jisoo Yoon[1], Seonghwan Kim[1], Jaehoon Jang[1], Jaihyuk Song[1]
[1]Semiconductor R&D Center, Samsung Electronics Co., Ltd.

For the first time, 900-layer class VNAND integration was implemented by bonding two 450-layer cell wafers. The cell wafer with a large warpage could be successfully supported in the upper chuck through an appropriate warpage design during bonding process, and the alignment deterioration that occurs in bonding between cell wafers with large warpage was also solved by applying new overlay correction technology. In addition, normal cell operation characteristics were verified based on the newly introduced the bit-line (BL) and word-line (WL) structure that can drastically improve the power consumption and chip size.

4:40 PM

TFS1.4 Toward Thousands-Layer 3D NAND: Unleashing the Block Size Constraint by Enabling Lateral Sub-Block Mode, Wei Cao[1], Takayuki Inoue[1], Peng Zhang[1], Jiacen Guo[1], Xiang Yang[1], Deepanshu Dutta[1]
[1]Sandisk Corporation

Stacking more wordline layers has been the main scaling approach for 3D NAND to achieve higher bit density. However, this approach leads to growing block size, which is becoming a major challenge to maintain the bit density increase rate in the beyond-1000-layer arena. To surmount this obstacle, a novel lateral sub-block mode is demonstrated in this work based on the advanced 3D NAND platform, which allows the large-block size induced bit density loss to be recovered by as much as 80%-90%. This demonstration provides a cost-efficient pathway for future 3D NAND scaling

5:05 PM

TFS1.5 A Multi Stacked Cell Array (MSA) Structure with an Up-down Crossed Bit Line Architecture (Ud-CBL) in 3D Flash Memory, Hiroshi Maejima[1], Katsuaki Isobe[1], Toshifumi Hashimoto[1], Masaki Unno[1], Nobuaki Okada[1], Tetsuaki Utsumi[1], Toshiki Hisada[1], Takao Nakajima[1], Mitsuhiko Noda[1], Keisuke Nakatsuka[1], Masayoshi Tagami[1], Shinji Suzuki[1], Shigeki Kobayashi[1], Keiji Maruyama[1], Toshiyuki Kouchi[1], Toshiya Kotani[1], Hitoshi Shiga[1]
[1]Memory Div., Kioxia Corporation

We propose a Multi Stacked Cell Array (MSA) structure with an Up-down Crossed Bit Line architecture (Ud-CBL) to enhance the performance of 3D Flash memory. The MSA can address two key performance issues common in 3D Flash memory with increasing word line (WL) layers: reduced cell current and increased block (BLK) size. The Ud-CBL minimizes area overhead through layout innovations, halving both WL length and BLK size. For a 1Tb 3-bit per cell (TLC) 4-plane design, assuming 400 WL layers, the read time (tR) is estimated to improve by 11% compared to a conventional structure. Additionally, we present a 4KB read method leveraging the Ud-CBL that enhances power efficiency by 7%.

Wednesday, June 17, 2026

Session P2: Plenary 2

8:00 AM
Tapa 1-3

Chairpersons: Makoto Takamiya, The University of Tokyo and Masaharu Kobayashi, The University of Tokyo

Welcome and Opening Remarks and Awards
Sugako Otani, Renesas Electronics Corporation and Kazuhiko Endo, Tohoku University

P2.1 – 8:35 AM
Intelligence Accelerated: Memory Innovations to Power the AI Era, Nirmal Ramaswamy, Micron Technology

Artificial intelligence is driving exponential growth in computation, data transfer, and energy consumption. As model sizes surpass the trillion-parameter threshold and reasoning workloads increase demands on bandwidth and latency, memory has emerged as the central constraint for advanced AI systems rather than compute resources. Meeting these exponential requirements necessitates coordinated advancements in memory technologies—such as DRAM and NAND, high-bandwidth memory, and packaging solutions including hybrid bonding, fine-pitch die-to-die interconnects, and vertically stacked memory architectures. Breakthroughs in materials science, sophisticated physics-based modeling, wafer bonding, and advanced metrology solutions are essential for rapid progress. The integration of emerging non-volatile memories and CXL-enabled expansion frameworks further supports heterogeneous and composable memory systems tailored for large-scale AI. This work outlines device, packaging, energy-efficiency, and memory product innovations required to address the evolving needs of AI workloads over the next decade.

P2.2 – 9:15 AM
Meeting AI Demand Through Equipment Innovation and AI-Driven Manufacturing: Progress and Challenges, Yoshinobu Mitano, Executive Vice President & General Manager, Tokyo Electron Ltd

Our industry has grown together with the evolution of semiconductor manufacturing equipment. Driven by the rise of AI and its prospects, the role of semiconductors continues to expand, and equipment has become even more critical. At the same time, it is increasingly important to apply AI to equipment technologies and fab operations themselves. In this talk, we will present our efforts on both fronts.
We will introduce how new equipment and process technologies contribute to the advancement of AI systems toward more high performance with less power consumption, and how we are using AI in equipment and fab technology.
We will outline current progress, challenges, and next steps. By addressing these two sides together, we aim to show how the semiconductor industry can meet the needs of AI today and in the future.

Technical Sessions Block 4 (10:15 AM – 12:20 PM)

Session C11: Wireless Power Transfer & Energy Harvesting Interfaces

10:15 AM, Honolulu 1
Co-Chairs: Jason T. Stauth, Dartmouth College and Sung-Wan Hong, Seoul National University

10:15 AM

C11.1 A 48 V, 25 W, 6.78 MHz Self-Starting Class-D WPT Transmitter with Self-Biasing and Autonomous Clocking Achieving 91% Peak DC-DC Efficiency, Saurabh Kale[1], Bernhard Wicht[1]

[1]Institute of Microelectronic Systems, Leibniz University Hannover

This work presents a self-sustained Class-D wireless power transmitter operating without auxiliary supplies or level shifting. High-side bias is generated using dual-edge capacitor charge harvesting, reducing the required harvesting capacitance, while switch-node voltage sensing is used for autonomous clock recovery eliminating cross-domain gate control. A dedicated startup mechanism initializes operation from rest, enabling entire operation solely with a DC bus voltage and a single clock input. Implemented with a monolithic control stage driving a discrete GaN half-bridge, the system delivers 25 W at 6.78 MHz from 48 V with 91% DC-DC efficiency over 40 mm transfer distance.

10:40 AM

C11.2 A 6.78MHz Fractional-Resonance Full-Wave Current-Mode Wireless Receiver with Real-Time SAR-based Optimum Tracking and Power-Aware Division Control, Jong-Hun Kim[1], Seung-Ju Lee[1], Seogyong Jeong[2], Mun-Jung Cho[1], Yeon-Woo Jeong[1], Min-Sik Kim[1], Myeong-Ho Kim[1], Dong-Chan Lee[1], Min-Gyu Jeong[1], Geon Kim[1], In-Hye Hur[1], Eun-Jae Ko[1], Se-Un Shin[1]

[1]Pohang University of Science and Technology, [2]Samsung Electronics Co., Ltd.

This work presents a 6.78 MHz fractional-resonance current-mode wireless receiver with successive-approximation (SAR) based optimum tracking and power-aware division control for implantable devices. A new full-wave topology and dithering control method enable fractional resonance to enhance energy extraction, while the controller autonomously adapts to link variations. Fabricated in 180 nm CMOS, it achieves 81.2% efficiency at 18.9 mW and the widest received power range among state-of-the-art designs.

11:05 AM

C11.3 A 6.78MHz Fully Integrated Dual-Output Symmetric 1/16-Cycle Regulating Rectifier Achieving 92.57% Efficiency and 3.7 V Voltage Difference, Xiaguang Li[1], Chao Yang[1], Shiheng Yang[2], Sijun Du[3], Xiaoyang Zeng[1], Zhiyuan Chen[1]

[1]Fudan University, [2]University of Electronic Science and Technology of China, [3]Delft University of Technology

This paper presents a 6.78-MHz fully integrated dual-output symmetric 1/16-cycle regulating rectifier that employs adaptive duty-ratio control and symmetry-based phase scheduling to enable fine-grained intra-cycle regulation with fast transient response and low ripple. In addition, dual-mode regulation, pulse merging, and gate-charge recycling effectively suppress ripple and balance transient performance and overall efficiency. The fabricated chip delivers up to 110 mW with fully integrated output capacitors, achieves a peak receiver efficiency of 92.57%, a maximum output voltage difference of 3.7 V, a 72.5% end-to-end efficiency, and the highest FoM with >21× improvement compared to prior work.

11:30 AM

C11.4 A CMOS Energy Harvesting System Integrating SIDO/DISO Continuously-Scalable- Conversion-Ratio SC Converter and Energy-Recycling BSI Photovoltaics, Minsung Kim[1], Daehong Kim[1], Minjae Kim[1], Joongyu Kim[1], Byeongwoo Yoo[1], Sung-Yun Park[1]

[1]Pusan National University

We present a near-infrared energy harvesting system that integrate a continuously-scalable-conversion-ratio (CSCR) switched-capacitor (SC) converter, backside-illuminated (BSI) CMOS photovoltaics (PVs), a maximum power point tracking (MPPT), and all controls in a single chip. The PVs recycle incident BSI light to enhance photo-collection efficiency (η_{PC}) and enable 3-D stacking of energy-reservoir capacitors by optical grating and reflector. The CSCR SC converter alternates between single-input-dual-output and dual-input-single-output with low ripple according to MPP and improves power conversion efficiency (η_{PCE}) via adaptive virtual level selection and flying

capacitor-split schemes. The system achieves 21% η_{PC} at low illumination intensity of 90 μW/mm^2 and 91.1% peak η_{PCE}.

11:55 AM
C11.5 A Universal Piezoelectric Energy-Harvesting Interface with Machine Learning-Based Excitation Identification and Cross-Topology MPPT for Diverse Input Vibrations, Guangshu Zhao[1], Haoyang Sang[1], Chenxi Wang[1], Hao Wu[1], Litao Zhang[1], Ningchao Lin[1], Kaixue Ma[2], Rui Paulo Da Silva Martins[1,3], Man-Kay Law[1]
[1]University of Macau, [2]Tianjin University, [3]Instituto Superior Tecnico/University of Lisboa

This work presents the first universal piezoelectric energy harvesting (PEH) interface that achieves machine learning (ML)-based coarse/fine excitation type identification and ML-assisted cross-topology maximum power point tracking (MPPT), capable of selecting the best-matched topology under distinct input excitation types. Moreover, by decomposing prior PEH topologies into a functional control framework, the proposed interface can realize arbitrary PEH topologies. The fabricated interface achieves near-normalized peak output across diverse excitations and a versatility FoM of 7.39, corresponding to a 1.3-2.4× compared with prior arts.

Session C12: True Random Number Generators

10:15 AM, Honolulu 2
Co-Chairs: Chris Kim, University of Minnesota and Noriyuki Miura, The University of Osaka

10:15 AM
C12.1 An 80.198um^2 Compact TRNG with Bias Mitigation, Proven PVT Tolerance in 4nm, Yunhyeok Choi[1], Yong Ki Lee[1], Bohdan Karpinskyy[1], Jieun Park[1], Hyunwoo Ko[1], Yongsoo Kim[1], Taewook Park[1], Jisu Kang[1], Sungha Lee[1], Gapkyung Kim[1], Jonghoon Shin[1], Hyo-Gyuem Rhew[1]
[1]Samsung Electronics Co., Ltd.

This paper presents a ring-oscillator (RO) based true random number generator (TRNG) that improves a 97.54% bias reduction, achieving a high entropy per bit of 0.986 (NIST SP 800-90B IID-track) without post-processing at 4 Mbps. Enabled by a simple yet novel bias-mitigation method, the proposed TRNG occupies only 80.198um^2

10:40 AM
C12.2 A 0.33 pJ/b Differential CMOS Thyristor-based TRNG with PVT Tolerance and LSB Min-Entropy Preservation in 28nm Process, Jaerok Kim[1], Donguk Seo[1], Yoonmyung Lee[1]
[1]Electrical and Computer Engineering, Sungkyunkwan University

This paper presents a differential CMOS thyristor-based TRNG (DCT-TRNG) achieving high energy efficiency and PVT robustness with three key techniques: 1) bias self-regulation using source followers and charge pumps to regulate noise accumulation time, 2) a glitch-filtering latch that prevents state-dependent sampling to preserve LSB entropy, and 3) differential short-window activation to minimize active power. Fabricated in 28nm CMOS, the prototype achieves 0.33 pJ/b energy efficiency and 13.3 Mb/s throughput at 0.8 V. It maintains consistent entropy across -40 to 125°C and 0.8-1.1 V, passing all NIST tests without post-processing.

11:05 AM
C12.3 A 5GS/s Staggered Working Point Ensemble TRNG with a Stochastic Model Achieving 4064.5 Mbps Entropy-Throughput and 997.6 fJ/entropy-bit, shaonan Wu[1], Ziheng Ma[1], Bohan Yang[1], Yuanxiao Wang[1], Zhiqi Yang[1], Weinan Chen[1], Hanning Wang[1], Min Zhu[2], Aoyang Zhang[1], Leibo Liu[1]
[1]School of Integrated Circuits, Tsinghua University, [2]Wuxi Micro Innovation Integrated Circuit Design Company

This paper presents a robust metastability-based TRNG featuring a Staggered Working Point Ensemble (SWE) architecture. To address the fundamental trade-off between tracking latency and power consumption caused by dynamic metastable point drift, the SWE scheme replaces real-time feedback loops with a spatial

ensemble strategy, ensuring continuous entropy coverage. A physical-process-based stochastic model validates the design's provable min-entropy. Fabricated in 28 nm CMOS, the prototype achieves a peak throughput of 4.19 Gb/s and a best-in-class energy efficiency of 324 fJ/bit. The design demonstrates resilience against aging and power injection attacks, passing both NIST SP800-90B and AIS-31 test suites.

11:30 AM
C12.4 A 0.511 pJ/bit Data-Independent and Post-Processing-Free 6T-SRAM-Based TRNG Exploiting Differential Bit-Line Leakage, Jeongho Kim[1], Donghyun Park[1], Hyeri Kang[2], Minhyeok Jeong[1], Donguk Seo[1], Kyungjun Lee[1], Yoonmyung Lee[1]
[1]Electrical and Computer Engineering, Sungkyunkwan University, [2]LG Electronics

This paper presents a 6T SRAM-based true random number generator (TRNG) that achieves the lowest reported energy among memory-based TRNGs while enabling data-independent operation without disturbing stored data. Fabricated in 28 nm CMOS, the TRNG exploits differential bit-line leakage to eliminate post-processing and achieves 0.511 pJ/bit at 0.7 V while passing NIST tests. To the best of our knowledge, it is the first memory-based TRNG demonstrating fully data-independent operation, with only 2.85% energy variation across data patterns. The proposed architecture is fully compatible with conventional 6T SRAM arrays without requiring any modification to the memory cell.

Session JFS2: Sensor Devices & Circuits

10:15 AM, Honolulu 3
Co-Chairs: Rashmi Jha, Univ of Cincinnati and Aaron Voon-Yew Thean, National University of Singapore

10:15 AM
JFS2.1 Flexible Monolithic 3D Integration of RRAM-based CIM, CNT CMOS Circuits and IGZO FET-based Sensor and Display for Augmented Reality Contact Lenses, Weijin wang[1], Yibei Zhang[1], Jianshi Tang[1], Qiqi Zhang[2], Ziruo Wang[2], Lei Gao[3], Haitao Xu[2,3], Yanbo Su[1], Jiaming Li[1], Ruofei Hu[1], Zhidong Tang[1], Kexin Wang[1], Yijia Fan[1], Jian Yuan[1], Peng Yao[1], Dong Wu[1], Ning Deng[1], Bin Gao[1], He Qian[1], Huaqiang Wu[1]
[1]School of Integrated Circuits, Beijing Advanced Innovation Center for Integrated Circuits, BNRist, Tsinghua University, [2]Institute of Carbon-based Thin Film Electronics, Peking University, [3]Beijing Institute of Carbon-based Integrated Circuits

Augmented reality (AR) contact lenses emerge as a promising immersive platform offering seamless, eye-integrated augmented experiences. In this work, for the first time, we present M3D-FAR, a prototype Monolithic 3D integration chip featuring Flexible Augmented Reality contact lenses with three interconnected functional layers on the same flexible substrate: 1st layer of 8×8 Ta_2O_5-based resistive random-access memory (RRAM) array for digital computing-in-memory (CIM), 2nd layer of carbon nanotube (CNT) CMOS circuits for logic and data interface, and 3rd layer of $InGaZnO_x$ field-effect transistors (IGZO-FET) for 8×24 photosensor array and display driver circuits. The structural integrity and proper function of the M3D-FAR chip was validated by structural analysis and electrical measurements. Furthermore, system-level benchmark in a typical image-based interactive task shows that the M3D-FAR architecture could achieve 10.45× speed-up compared to its 2D counterpart and consume 5.39× lower energy than GPU.

10:40 AM
JFS2.2 Depletion-Mode III-V/Si SISCAP Mach-Zehnder Modulator Breaking the Efficiency-Bandwidth Trade-Off for Co-Packaged Optics, Donggil Kang[1], Shin Hyung Lee[1], Jae-Hoon Han[2], Jongmin Kim[3], Seok-Geun Ahn[4], Minhyuk Jung[4], Hyun-Chul Jung[4], SangHyeon Kim[1]
[1]School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), [2]Korea Institute of Science and Technology, [3]Korea Advanced Nano Fab Center, [4]Advanced PKG Lab., Samsung Electronics Co., Ltd.

Design-technology co-optimization in co-packaged optics has become important; designing optical modulators would be a good example because their key figure of merits, such as high bandwidth, efficiency, and thermal robustness, will be impacted by the layout, device structure, etc. This work demonstrates a III–V/Si SISCAP Mach–Zehnder modulator utilizing depletion-mode operation, Franz–Keldysh–assisted modulation, and a traveling-wave electrode design, achieving a modulation efficiency of 0.146 V·cm and an electro-optic bandwidth of 26.3 GHz, corresponding to a record optical modulation amplitude (OMA) line.

11:05 AM

JFS2.3 Back-Illuminated CMOS-SPAD With Full-Surround Anode Structure Achieving Sub-50 ps Timing Jitter for High-Precision Physical-AI Sensors, Joo-Hyun Kim[1,2], Doyoon Eom[1,2], Hyun-Seung Choi[1,2], Eunsung Park[1], Woo-Young Choi[1], Myung-Jae Lee[1,3]

[1]Department of Electrical and Electronic Engineering, Yonsei University, [2]Post-Silicon Semiconductor Institute, Korea Institute of Science and Technology (KIST), [3]TruPixel

We present a Full-Surround anode SPAD (FS-SPAD) fabricated in a back-illuminated 90 nm CMOS image sensor (CIS) foundry technology with high performance timing jitter for short/mid- range dTOF senser. The FS-SPAD improves carrier transport by co-optimizing the p-epitaxial (p-epi) layer and the anode, which is ease-modification with other non-isolated SPADs. The SPAD achieves a breakdown voltage (V_B) of 15.8 V and demonstrates a timing jitter of 49 ps (FWHM), a photon detection probability (PDP) of 24.2%, and a normalized dark count rate (nDCR) of 12.57 cps/μm^2.

11:30 AM

JFS2.4 A Highly Scalable and Noise-Robust Direct Digital Readout (DDRO) for E-Skin Temperature Sensor Arrays Achieving 12.2 pJ/Sensor and 710 ns/Row, Kosuke Kayano[1], Sang-Gyu Koh[2], Shiyuan Sun[1], Satoshi Hamasuna[1], Takaaki Miyasako[2], Tadasu Hosokura[2], Takeaki Yajima[1]

[1]Kyushu University, [2]Murata Manufacturing Co., Ltd.

Direct Digital Readout (DDRO) is proposed as a scalable, noise-robust readout architecture for high-density, multi-point e-skin targeting human-skin-level spatial resolution. In contrast to conventional analog-sensor readout with ADCs, where wiring-induced disturbances and parasitic resistance/capacitance corrupt analog signals on the crossbar interconnects and rapidly collapse signal margin as arrays scale, DDRO digitizes each phase-transition sensor output at the sheet interface and transports fully digital signals on the interconnects. This structural digitization suppresses array-scaling noise sensitivity and enables high scalability without tightening analog accuracy requirements. The concept is also aligned with human thermal perception, which exhibits a nonlinear response close to binary discrimination. A prototype system combining a flexible 1024-sensor sheet and a DDRO chip demonstrates 710 ns/row latency and 12.2 pJ/sensor energy, enabling full-frame readout of all 1024 sensors within 23 us while consuming 12 nJ.

11:55 AM

JFS2.5 A 3D-Conformal Flexible X-ray Detector with 2.06pA$_{rms}$ Ultra-Low Noise Readout Circuit Enabling Low-Dose Computed Tomography Imaging, Zhongqiu Shen[1], Kaiwen Zhou[1], Shi Chen[1], Shen Ye[1], Zhiliang Hong[1], Yingguo Yang[1,2], Jiawei Xu[1]

[1]Fudan University, [2]Shanghai Synchrotron Radiation Facility (SSRF), Shanghai, China

This paper presents the first demonstration of a flexible X-ray imaging system for low-dose computed tomography (CT). The system features a 3D-conformal flexible X-ray detector that integrates ultra-high-mobility ITZO thin-film transistors (TFTs) with perovskite-based X-ray sensing layers, achieving a sensitivity of 296μC/(Gy·cm^2). The detector is co-designed with an ultra-low-noise, scalable current readout consisting of 6-channel ping-pong C-TIAs and a shared CT-IADC with an optimized reset scheme, delivering 2.06pA$_{rms}$ input-referred noise over a 5kHz bandwidth. Experiments with the TFT-based flexible detector and a real X-ray source validate the system's capability to reconstruct both gradient and patterned features, demonstrating its practical viability for low-dose CT imaging.

Session T5: Technology Highlights 2

10:15 AM, Tapa 1-3
Co-Chairs: Anabela Veloso, imec and Masahiko Kanda, Toshiba Electronic Devices & Storage Corporation

10:15 AM
T5.1 Vertically Stacked DRAM Technology for Scaling Evolution, Seung Uk Han[1], Jin-Woo Han[1], Gyuhwan Oh[1], Sang-Il Han[1], Sungmin Park[1], Juhyun Kim[1], Taekyong Kim[1], Taejin Bae[1], Taegyu Kang[1], Joongchan Shin[1], Seongbin Hong[1], Bowon Yoo[1], Hongrak Choi[1], Yujin Kim[1], Sang Hyun Sung[1], Kyoseon Choi[1], ManBok Kim[1], Jin Kwan Lee[1], Janghee Lee[1], Chang Hwa Jung[1], Jungmin Oh[1], Myoung Ho Jung[1], Hojin Lee[1], Yungjun Kim[1], Soongun Lee[1], Kyuman Hwang[1], Moon-Hyun Cha[1], Taeyoon An[1], DK Lee[2], BC Kim[2], Dongwan Kim[2], Kwangho Ahn[2], Kwangmin Park[1], Sang Wuk Park[1], Seung Hun Lee[1], Sangmin Hwang[2], Siwoo Lee[2], Sangjin Hyun[1], Se Geun Park[1], Jaihyuk Song[1]
[1]R&D Center, Samsung Electronics Co., Ltd., [2]DRAM Pathfinding, Samsung Electronics Co., Ltd.

Vertical Stacked DRAM (VS-DRAM) has emerged as a promising candidate to extend DRAM generations beyond the 10nm node, leveraging cell stacking similar to V-NAND. Demonstrating reliable device operation at higher tiers is crucial to ensuring its competitiveness. For the first time, this paper presents experimental results of a 16-tier VS-DRAM structure. The study integrates Gate-All-Around (GAA) type cell transistors (CTR) and horizontal storage capacitors (CAP) and compares these results with conventional DRAM. Additionally, a core/peri (C/P) device is fabricated on a separate wafer and its backside is bonded to the cell wafer, enabling the implementation of Peri-on-Cell architecture. The findings highlight the potential of VS-DRAM and pave the way for future DRAM innovations.

10:40 AM
T5.2 Demonstration of CFET Inverters on Si (110) with 2X2 RibbonFETs at 45nm Gate Pitch with PowerVia and Direct Backside Contacts, Jami Wiedemer[1], David Bennett[1], Evan Clinton[1], Cheng-ying Huang[1], Rohit Galatage[1], Munzarin Qayyum[1], Valur Gudmundsson[1], Hojoon Ryu[1], Ramanan Chebiam[1], Nicole Thomas[1], Pratyush Buragohain[1], Patrick Morrow[1], Ashish Agrawal[1], Thoe Michaelos[1], Rambert Nahm[1], Natalie Briggs[1], David Kohen[1], Andrey Vyatskikh[1], Scott Goodwin[1], Susmita Ghose[1], Niels Zussblatt[1], Yuwen Huang[1], Dincer Unluer[1], Madeleine Beasley[1], Miles Miles[1], Li Huey Tan[1], Matthew Elkins[1], Seda Cekli[1], Richard Hermann[1], Leah Shoer[1], Makram Qader[1], Umang Desai[1], Trenton Edwards[1], Priyaa Prasad[1], Jeff Armstrong[1], Mithun Ghosh[1], Yean-An Liao[1], Vadym Kapinus[1], Dhairya Dixit[1], Phuong Tran[1], Kai Cheong[1], Adnan Fatehi[1], Adedapo Oni[1], Noel Franco[1], Brian Krist[1], Quan Shi[1], Shishir Agrawal[1], Anand Krishnamoorthy[1], Tai-hsuan Wu[1], Sabareesh Sridhar[1], Sonia Leon[1], Matthew Metz[1], Jessica Torres[1], Hoon Sung[1], Gilbert Dewey[1], Richard Schenker[1], Mauro Kobrinsky[1], Marko Radosavljevic[1]
[1]Intel Corporation

Complimentary Field-Effect Transistor (CFET) technology is considered a key path to continuing scaling in the semiconductor industry beyond RibbonFET. In this work, we demonstrate fully integrated monolithic CFET inverters at contacted poly pitch (CPP) of 45nm. Stacked RibbonFET transistors on Si (110) substrate with 2 NMOS and 2 PMOS ribbons (2X2) are integrated with backside power via, direct backside contacts (BSCON), and an epi-epi via (EEV) intraconnect. A vertical nanoribbon depopulation process is demonstrated to eliminate the effects of parasitic transistors without requiring a split-gate process. Additionally, we demonstrate hybrid PMOS Si (110) on NMOS Si (100) stack with middle dielectric isolation (MDI) less than 10nm to enable power, performance, and area (PPA) improvement. Cell PPA analysis for 2X2 ribbon stack shows significant frequency benefit compared to 1x1.

11:05 AM
T5.3 Record $2P_r$ (>38 μC/cm² at 0.5 V, >28 μC/cm² at 0.4 V) of 3D MFM Capacitors Enabled by 3 nm HZO and ALD-TiN Orientation Engineering, Yang Feng[1,2], Xiaolin Wang[2], Jiezhi Chen[1], Xiao Gong[2]
[1]Shandong University, [2]National University of Singapore (NUS)

For the first time, we propose and demonstrate orientation engineering of ALD-TiN electrodes as an effective approach to enhance the remnant polarization ($2P_r$) in ultrathin 3 nm HZO layers. We promote ferroelectricity in

HZO without introducing additional process steps or integration complexity. Compatible with 3D structures, this approach is successfully extended to the first experimentally demonstrated 3 nm HZO-based 3D ferroelectric capacitors. The devices deliver a $2P_r$ >38 μC/cm^2 at V_{op} of 0.5 V, setting a new benchmark for 3D MFM capacitors. Importantly, a record-high $2P_r/V_{op}$ ratio of 76 is achieved, which is the highest reported to date for HZO layers with a physical thickness ≤3 nm. Even at a reduced V_{op} of 0.4 V, the devices continue to meet the IRDS $2P_r$ requirement (≥26.5 μC/cm^2).

11:30 AM
T5.4 High-Temperature Resilient SiGe Nanosheet PFET RMG Towards Multi-Tiered Sequential Integration, Nirmaan Shanker[1], Shogo Mochizuki[1], Takashi Ando[1], Shay Reboh[1], Thanh Nguyen[1], Debarghya Sarkar[1], Eunsoo Cho[1], Wai-Kin Li[1], Huimei Zhou[1], Richard Southwick[1], Juntao Li[1], Anisha Pawar[1], Ruturaj Pujari[1], Paul Jamison[1], Su Chen Fan[1], Brown Peethala[1], Hemanth Jagannathan[1], Ravikumar Ramachandran[1], Tenko Yamashita[1], Dechao Guo[1], Huiming Bu[1]
[1]IBM Research

We demonstrate, for the first time, SiGe nanosheet (NS) replacement-metal-gate (RMG) PFETs with high temperature stability exceeding 900°C, validated through sequential integration of a top Si NFET over a bottom SiGe PFET. Interfacial layer (IL) regrowth is suppressed via IL optimization and Al-doped work function metal (WFM). We extend the oxygen-vacancy–based Fermi-level pinning theory by showing that both n- and p-type WFMs converge to the same equilibrium oxygen vacancy concentration after the high-temperature budget required for sequential integration. Consequently, V_T is tuned by the SiGe–Si valence band offset rather than the WFM and the WFM can be optimized to mitigate IL regrowth. Together with our prior demonstration of high-temperature-stable NFETs, this work lays the foundation for multi-tiered sequential integration.

11:55 AM
T5.5 A 2.1-μm Pixel-Pitch CMOS Image Sensor with 65% MTF/35% QE IR Global Shutter and RGB Rolling Shutter Sequential Operation for In-cabin Applications, Mizuha Hiroki[1], Tatsuya Takeuchi[1], Yuta Nakamoto[1], Yhang Ricardo Sipauba Carvalho da Silva[1], Yuki Yoshimura[1], Hiroki Hagiwara[1], Ryotaro Takata[1], Kaihei Hotta[1], Rui Nishikawa[1], Yuji Nishimura[2], Shingo Yamaguchi[2], Ayumi Takayama[2], Ryosuke Nakamura[1], Takahiro Toyoshima[1], Yorito Sakano[1], Yusuke Oike[1]
[1]Sony Semiconductor Solutions Corporation, [2]Sony Semiconductor Manufacturing Corporation

In this work, we propose a 2.1-μm RGB-IR CMOS image sensor with sequential operation combining rolling shutter (RS) and global-shutter (GS) modes for in-cabin applications. As an RGB sensor, it achieves a dynamic range (DR) of 112 dB at T_j=85 °C, enabling high-quality visible imaging. For the IR sensor, we achieve state-of-the-art characteristics with 65% modulation transfer function (MTF) and 35% infrared quantum efficiency (IR-QE), facilitating gaze detection under IR illumination.

Technical Sessions Block 5 (1:30 PM – 3:10 PM)

Session C13: In-Body Bioelectronic Systems

1:30 PM, Tapa 3
Co-Chairs: Phillip Nadeau, Analog Devices and Chihiro Okada, Sony Semiconductor Solutions Corp.

1:30 PM
C13.1 A Split-Inference Intracortical Interface IC for Battery-Free mm-Scale Magnetoelectrically Powered Brain Implants, Mustafa Kanchwala[1], Jianxiong Xu[1], Mohammad Abdolrazzaghi[1], Wonjune Kim[2], Gerard O'Leary[1], Yu Huang[1], Junyu Ma[1], Jose Sales Filho[1], Sudip Nag[1], Hanfeng Cai[1], Qiaosong Deng[1], Weian Deng[1], Anush Mutyala[1], Theeban Kumaresan[1], Shidu Ren[1], Chae Lim[1], Mandana Movahed[1], Homeira Moradi[1], George Eleftheriades[1], Taufik A. Valiante[1], Jacob T. Robinson[2], Roman Genov[1]
[1]University of Toronto, [2]Rice University

This paper presents a brain-interfacing IC that enables depth-scalable edge inference within the tight form factor and energy budget of a minimally invasive, battery-free implant. A 181-dB FoM noise-shaping SAR ADC with an LNA is shared across 20 of 32 recording electrodes. On-chip inference combines boosted decision-stump spike classification with first-layer spiking- and deep-neural-network (SNN/DNN) inference, reducing the output data rate by up to 16,732×. An event-based body-coupled communication (BCC) TX streams intermediate activations to a wearable device for remaining-layer execution. The communication and computation energy costs are 1.23 nW and 53 nW per input, respectively, enabling wireless powering by a mm-scale magnetoelectric film.

1:55 PM

C13.2 A Wireless Pressure Sensor Implant with MRI-Encoded Data Uplink, Biqi Rebekah Zhao[1], Alexander Chou[1], Changuk Lee[1], Aviral Pandey[1], Jade Pinkenburg[1], Liz Murray[1], Karthik Gopalan[1], Chunlei Liu[1], Michael Lustig[1], Rikky Muller[1]
[1]Electrical Engineering and Computer Sciences, University of California, Berkeley

This work presents a wireless implantable pressure sensor using active MRI-based uplink for simultaneous sensing and localization. The system integrates a MEMS pressure transducer with a 2x1mm^2 28nm IC powered by a 1MHz inductive link. Data uplink is performed by actively modulating MR image voxels using an on-chip micro-coil. A 12-bit 1st-order CTΣΔ ADC digitizes pressure data at 375μW. In-scanner tests demonstrate 15% modulation depth at 56μW.

2:20 PM

C13.3 GMINI: Gut Modulating Neural Interface Chip for Closed-loop Non-invasive Therapies, Adam Gierlach[1], Hyemin Stella Lee[1], Deniz Umut Yildirim[1,2], Giovanni Traverso[3], Anantha P. Chandrakasan[1]
[1]Electrical Engineering & Computer Science, Massachusetts Institute of Technology, [2]MediaTek Inc., [3]Mechanical Engineering, Massachusetts Institute of Technology

Current ingestible devices are limited by high power usage and slow sensor readouts, limiting awareness and functionality. Our Gut Modulating Neural Interface chip known as GMINI leverages gastric neural signals to provide faster, more effective therapies such as electrical stimulation of the gut in one package – with an 8 channel neural recorder, adaptively duty cycling slow wave tracker, behavior detection ML accelerator, and programmable current stimulator. With real-time adaptive thresholding, we achieve a 48% improvement in tracking the slow wave period, enabling adaptive duty cycling and more effective electrical stimulation therapy synchronized with the gut. Our gastric signal-optimized CNN accelerator consumes the lowest energy (0.266 μJ/class) among comparable workloads to efficiently determine patient behaviors. This enables ingestible devices to stay active longer and trigger intelligently depending on the patient's needs.

2:45 PM

C13.4 A 50-mm^2 Crystal-Free Bidirectional WPDT System with Carrier-Reuse Clocking for Implantable Neural Interfaces, Cong Ding[1], Mingxiang Gao[2,3], Shaokang Zhao[1], Niels Kuster[2,3], Armin Tajalli[1,4], Mahsa Shoaran[1]
[1]EPFL, [2]ETH Zürich, [3]IT'IS Foundation, Zurich, Switzerland, [4]University of Utah

Neural implants require compact and energy-efficient wireless power and data transfer (WPDT) while avoiding reliance on costly on-board clock sources. This work presents a crystal-free, full-duplex WPDT link that reuses a 500-MHz power carrier as the sole system clock for synchronized downlink and uplink telemetry. A clock-data recovery (CDR)-less coherent ASK downlink achieves 5 Mb/s, while maintaining uninterrupted power transfer through a 6-mm implant loop. The recovered 500-MHz carrier also wirelessly locks a fast-settling UWB PLL, enabling a duty-cycled uplink supporting 5–500Mb/s. Burst-mode operation enables data-rate-scalable uplink power, ranging from 0.23 mW at 5 Mb/s to 4.93 mW at 500 Mb/s, and extends the uplink range to 40 cm using an on-chip antenna. The complete system, including the loop antenna and chip, occupies 50 mm^2.

Session C14: Adaptive and Agile Radio Architectures

1:30 PM, Honolulu 1
Co-Chairs: Vanessa Chen, Carnegie Mellon University and Shuichi Nagai, Panasonic Industry Co., Ltd.

1:30 PM
C14.1 A Reconfigurable 24-28GHz Time-Division Full-Duplex Transceiver with 59dB Self-Interference Rejection over 400MHz, Dongfan Xu[1], Haiyun Gu[1], Minzhe Tang[1], Yuxuan Liu[1], Ziyuan Ren[1], Yilun Chen[1], Minghao Fan[1], Duo Li[1], Zheng Li[1], Yi Zhang[1], Daxu Zhang[1], Zezheng Liu[1], Chun Wang[1], Sena Kato[1], Yudai Yamazaki[1], Hiroyuki Sakai[1], Yuncheng Zhang[1], Kazuaki Kunihiro[1], Kenichi Okada[1]
[1]Institute of Science Tokyo

This work presents a reconfigurable 24–28GHz time-division (TD) full-duplex (FD) transceiver in 65nm CMOS. It supports at least 400MHz bandwidth with up to 59dB self-interference (SI) rejection, achieved by isolating TX/RX in time domain without an extra SI cancellation (SIC) path. The proposed architecture overcomes SIC bandwidth limits and allows switching to half-duplex (HD). The OTA measurements verify 400MHz 64QAM OFDM FD and HD operations, offering a low-complexity mm-Wave FD solution.

1:55 PM
C14.2 A 1.5-to-20.5 GHz Intrinsic-Spurious-Rejection Real-IF Dual-Translation Transceiver Chipset Achieving 10-Gb/s 1024-QAM and 5-µs Hopping, Dongze Li[1], Wei Deng[1], Yichi Sun[1,2,3], Xiangyu Nie[1], Fuzhi Xue[1], Haikun Jia[1], Liangjiang Zhou[2,3], Rui Wu[2,3], Baoyong Chi[1]
[1]Tsinghua University, [2]Aerospace Information Research Institute, Chinese Academy of Sciences, [3]University of Chinese Academy of Sciences

This work presents a 1.5-to-20.5 GHz intrinsic-spurious-rejection real-IF dual-translation transceiver (TRX) chipset for agile frequency-hopping (FH) communications (FHC). By processing signals in the real-IF domain, both the transmitter (TX) and receiver (RX) adopt a dual-conversion architecture that fundamentally eliminates I/Q imbalance, achieving inherent sideband and image suppression for superior spectral purity. To enable rapid frequency switching, a fast-hopping PLL is co-designed to generate LO signals for the chipset. Measurement results demonstrate a 10 Gbps data rate with 1024-QAM modulation and a hopping time of 5 µs. To the authors' knowledge, this is the first complete CMOS TRX designed for fast FHC.

2:20 PM
C14.3 A 57–71-GHz CMOS Phased-Array Transceiver with Aperture-Tuning Antenna Achieving 47.9–62.2% EIRP Efficiency Improvement, Minghao Fan[1], Yilun Chen[1], Zheng Li[1], Ziyuan Ren[1], Minzhe Tang[1], Junqing Liu[1], Yuxuan Liu[1], Dongfan Xu[1], Zezheng Liu[1], Yudai Yamazaki[1], Sena Kato[1], Kazuaki Kunihiro[1], Hiroyuki Sakai[1], Yuncheng Zhang[1], Kenichi Okada[1]
[1]Department of Electrical and Electronic Engineering, Institute of Science Tokyo

This work presents a 60-GHz aperture-tuning antenna and transceiver for wide frequency coverage in the 5G NR FR2-2 n263 band. Distributed tuners reconfigure the radiation to enable mode-dependent operation, achieving frequency-selective EIRP enhancement under narrow instantaneous bandwidth operation. Fabricated in 65-nm CMOS, a dual-polarized transceiver with switchless antenna-port mode-separated TX/RX operation is implemented. The proposed design achieves measured TX EIRP improvements of 62.2% (2.1 dB) at 57 GHz and 47.9% (1.7 dB) at 71 GHz without additional power consumption. Using a customized OFDM waveform with 14-GHz frequency coverage, a maximum data rate of 56 Gb/s is demonstrated with 16-QAM modulation.

2:45 PM
C14.4 An 8-Element 200-MHz Bandwidth Oversampling Digital Slepian Beamformer, Zhengqi Xu[1,2], Zhiyuan Zhao[1], Coleman DeLude[3], Justin Romberg[3], Michael Flynn[1]
[1]University of Michigan, Ann Arbor, [2]Marvell Semiconductor, [3]Georgia Institute of Technology

We introduce an 8-element digital Slepian beamforming receiver. The Slepian architecture reduces the number of digital fractional delay filters (FDFs) by at least 4 times. An array of bandpass time-interleaving noise-shaping SAR ADCs provides a 61 dB array SNR, and enables 800 MS/s multiply and accumulate (MAC) and FDF. Oversampling improves FDF efficiency. Fabricated in 28 nm CMOS, the beamformer supports a 200 MHz baseband bandwidth and consumes only 10.4 mW per beam for 8 beams with an EVM_{rms} of -42 dB.

Session C15: Power Integrity: Regulators, References, and Monitors

1:30 PM, Honolulu 2
Co-Chairs: Samira Zaliasl, Ferric Inc. and Kazuki Fukuoka, Renesas Electronics Corporation

1:30 PM
C15.1 A Robust, Digital-Friendly, Differential and Distributed Supply Monitor Architecture for Automotive SoC in 28nm, Stefano Bonomi[1], Frank Preamassing[1], Sie Boo Chiang[1], Werner Grollitsch[1], Roberto Nonis[1]
[1]Infineon Technologies

The proposed supply monitor digital architecture is applied in Automotive applications on Sense and Adapt and on State of Health systems. It operates within the core supply domain using core regular devices without additional sensors or analog references. It requires a one-time, one-point calibration at wafer test. It ensures a robust, digital-friendly and easy-to-integrate implementation that satisfy the ASIL safety level and temperature range -40°C to 160°C of Automotive applications. Fabricated in n28 the active area is 0.01 mm^2, power consumption 540uW. It tracks variations of 217mV in 1us, FoM is 2.45 pJ/conv-step., an inaccuracy of 10mV, after stress and with supply aggressors.

1:55 PM
C15.2 A 10-MHz Integrated Voltage Regulator with Cycle-Accurate and Lightweight Predictive Control Achieving 5-mV Off-Chip Droop, Haochang Zhi[1], Shaojie Xu[1], Kaize Zhou[1], Zhangrui Qian[1], Shengdao Ren[2], Tianshu Liu[2], Jintao Li[3], Yun Li[3], Weiwei Shan[1], Wanyuan Qu[2]
[1]Southeast University, [2]Zhejiang University, [3]University of Electronic Science and Technology of China

This paper presents a 10-MHz integrated voltage regulator featuring cycle-accurate, lightweight predictive control. A bandwidth-matched weighted training scheme selects 78 sparse transition detectors (TDs) that capture only power-delivery-network-shaped (PDN) frequency components. These TDs are used to predict the cycle-accurate core current I_{Core}; an adaptive PDN emulator with in-situ calibration then maps I_{Core} to the regulator-relevant load-current envelope I_{Load} for proactive regulation. Across 18 workloads, the predictor achieves 3.3% normalized root-mean-square error (NRMSE) on I_{Core}, which translates to 1.5% NRMSE on I_{Load}. Fabricated in 28-nm CMOS, the DC-DC prototype supports 0.81-A load steps with 66-ns edges, suppresses on-die second droop from 72-mV to less than 6-mV, and reaches 90.1% peak efficiency.

2:20 PM
C15.3 A Hybrid LDO with Region-Detection and Power-Supply-Rejection-Based Control Achieving High PSR and Fast Transient Response, Yu-Ting Huang[1], Xiao-Quan Wu[1], Chi-Yu Chen[1], Po-Jui Chiu[1], Chien-Wei Cho[1], Ke-Horng Chen[1], Seng-Yu Peng[1], Tsung-Heng Tsai[1], Ying-Hsi Lin[2], Shian-Ru Lin[2], Tsung-Yen Tsai[2]
[1]EE, National Yang Ming Chiao Tung University, [2]Realtek Semi.

The proposed hybrid LDO integrates region detection (RD) and gate-edge detection (GED) techniques to optimize the ALDO–DLDO current ratio to break the PSR trade-off. At 10kHz, the proposed HLDO achieves -49dB and -40dB PSR when V_{DO} is 200mV and 50mV, respectively, maintaining high PSR up to 10MHz range. For transient response, GED enables fast settling time; under a 1mA–200mA load step, the undershoot is 90mV with an 8.7ns settling time, and the overshoot is 92mV with a 10.2ns settling time.

2:45 PM
C15.4 A 0.75V, 1.43%-Inaccuracy 4T MOS-Based Voltage Reference in 251μm^2 Area, Szu-Lin Liu[1], Bei-Shing Lien[1], Wei-Lin Lai[1], Ching-Lin Jen[1], Yi-Chen Lu[1], Yung-Chow Peng[1], Kenny Cheng-Hsiang Hsieh[1]
[1]Taiwan Semiconductor Manufacturing Company (TSMC)

This paper presents a 4-transistor (4T) MOS-based voltage reference in a 3nm FinFET node. By using single threshold voltage (V_{TH}) design and identifying the temperature-slope correlation between the required CTAT voltage and the nMOS-V_{TH}, this voltage reference achieves a 3σ-inaccuracy of 1.43% and a temperature coefficient (TC) of 42.7 ppm/°C from a 0.75V-supply. The whole circuit consumes 12μW power and only occupies area of 251 μm^2.

Session C16: Low-Power Oscillators and Digital Techniques

1:30 PM, Honolulu 3
Co-Chairs: Gautam Gangasani, Apple and Kenichi Okada, Institute of Science Tokyo

1:30 PM
C16.1 A 5.7–6.8GHz Capacitively-Coupled Series-Assisted Parallel-Resonance VCO Achieving -148.8 dBc/Hz PN, 193.5 FoM, and 202.0 FoMA at 10MHz Offset, Long He[1], Jiang Gong[1], Taekwang Jang[1]
[1]ETH Zürich

This paper presents a capacitively-coupled series-assisted parallel-resonance (CSA-PR) VCO that achieves ultra-low phase noise with high power and area efficiency. By capacitively coupling a parallel LC tank with low-impedance auxiliary elements, the design provides passive voltage gain and waveform shaping to reduce both thermal and flicker noise. The tunable tank impedance enables flexible trade-offs between power and noise. Fabricated in 22-nm FDSOI CMOS, the single-core VCO achieves -148.8dBc/Hz phase noise and 193.5dBc/Hz FoM at 10MHz offset while consuming 11.8mW in a 0.14mm^2 area.

1:55 PM
C16.2 A Compact Magnetically Coupled Dual-Mode Class-F-1 VCO Covering 5.8-9.97 GHz with 196.2 dBc/Hz FoM and 70-320 kHz 1/f^3 Corner without Harmonic Tuning, Yue Wu[1], Yatao Peng[1], Jiawei Li[1], Fengen Yuan[1], Wenhua Gong[1], Jun Yin[1], Rui P. Martins[1], Pui-In Mak[1]
[1]State Key Lab. of Analog and Mixed-Signal VLSI, IME/ECE of FST, University of Macau

A wideband dual-mode inverse class-F VCO with intrinsic harmonic self-alignment is presented. A magnetically coupled quad-mode resonator enables tuning-free harmonic shaping using a single switching group, without explicit common-mode resonators. Implemented in 65-nm CMOS, the VCO achieves a 52.4% tuning range, FoM of 196.2 dBc/Hz, and 1/f^3 PN corners of 70~320 kHz, demonstrating state-of-the-art performance among mode-switching oscillators.

2:20 PM
C16.3 A Versatile On-Chip Jitter Measurement Circuit with Dynamic Quantizer-Resolution Adaptation for Arbitrary Clocks across 1–8GHz and 20–130pspp Jitter, Junseok Lee*[1], Chanwoong Hwang*[1], Jeongbeom Seo[1], Jaehyouk Choi[1]
[1]Seoul National University

This work presents a versatile on-chip jitter measurement (OCJM) circuit. By employing a dynamic quantizer-resolution adaptation (DQRA) technique that optimizes the quantizer resolution according to the jitter magnitude, the proposed OCJM accurately constructs jitter histograms for arbitrary clocks with wide frequency and jitter variations. Implemented in 28nm CMOS, the prototype uses only 15 quantizers to measure clock jitter across frequencies from 1 to 8 GHz and magnitudes from 20 to 130ps$_{pp}$.

2:45 PM

C16.4 A 20T Low-Power Single-Phase Full-Static Contention-Free Flip-Flop with Inter-Latch Transistor Sharing, Minkyu Ko[1], Bomin Joo[2], Byeonggon Kang[3], Unbok Wi[1], Jieun Kim[1], Dalhee Lee[2], Changbeom Kim[2], Hayoung Kim[2], Sanghoon Baek[2], Bai-Sun Kong[1]

[1]Sungkyunkwan University, [2]Samsung Electronics Co., Ltd., [3]University of California, San Diego

This paper proposes an area-efficient single-phase contention-free flip-flop (FF) with inter-latch transistor sharing. The flip-flop reduces the area by reconfiguring the primary and secondary latches of C^2MOS FF to share transistors. It has lower DQ latency due to a shorter setup time, resulting from fewer logic stages in the input setup path. It can also provide robust operation with a static contention-free structure. The proposed FF is fabricated in a 28-nm process. The measurement results show that the proposed FF achieves reductions of up to 23.3% and 17.7% in layout area and DQ latency, respectively. In terms of power-delay-area product (PDAP), the proposed FF achieves up to 35.2% improvement among low-power FFs. It can also ensure reliable operation at supply voltages as low as 0.3 V, thanks to its fully static contention-free operation.

Session T6: BEOL Processes and Devices

1:30 PM, Tapa 1

Co-Chairs: Ming Cai, Google and Tetsu Ohtou, Tokyo Electron Limited

1:30 PM

T6.1 Topological Semimetal MoP as Future Conductors in Back-End-of-Line: Thin Film Deposition, Resistivity Scaling, Processability, Integration, and Reliability, Keun Wook Shin[1], Hyangsook Lee[1], Changhyun Kim[1], Dae-Jin Yang[1], Young Min Lee[1], Hyeong-Seok Jang[1], Yujin Han[1], Youngchul Leem[1], Gi-Young Jo[1], Jeong Yub Lee[1], Kyung-Eun Byun[1], Sang Won Kim[1]

[1]Samsung Advanced Institute of Technology

To overcome the RC delay bottleneck in interconnect applications, topological semimetals (TSMs) have emerged as promising candidates for post-Cu interconnects. This study investigates MoP as a representative TSM, with a focus on its crystallographic structure, electrical properties, processability, integration, and reliability. This work provides the first experimental demonstration that MoP thin films exhibit reduced resistivity with decreasing thickness, a behavior that is preserved in narrow line structures. These lines sustain a J_{max} exceeding 100 MA/cm^2 and an electromigration (EM) lifetime of more than 20 h, highlighting MoP's strong potential for advanced interconnect technologies.

1:55 PM

T6.2 Resistivity scaling of topological CoSi wires to sub-10 nm for post-Cu interconnects, Guy Cohen[1], Franco Stellari[1], Peter Kerns[1], Leonidas Ocola[1], John Bruley[1], Lynne Gignac[1], Vesna Stanic[1], Christian Lavoie[1], Nathan Marchack[1], Joshua T. Smith[1], Vijay Narayanan[1], Ching-Tzu Chen[1]

[1]IBM T. J. Watson Research Center, IBM Research

We experimentally demonstrate the nanoscale resistivity-scaling behavior of topological conductors using cobalt monosilicide (CoSi) nanowires (NWs) fabricated on 200 mm wafers with CMOS-compatible processes. Both templated near-epitaxial and polycrystalline CoSi NWs are evaluated, with cross-sectional areas down to 35 nm^2 (~6 nm diameter). Unlike Cu, which shows a steep resistivity increase below ~20 nm linewidth, templated near-epitaxial CoSi NWs show an exceptional ~80% resistivity reduction when the NW cross-sectional area scales below ~100 nm^2 as compared to mm-scale wires. Furthermore, resistivity of polycrystalline CoSi NWs with random grain orientations also approaches the theoretical CoSi bulk single-crystal resistivity limit when the NW cross-sectional area scales below ~200 nm^2. These results provide wafer-scale validation that topological conductors can enable interconnect scaling beyond Cu.

2:20 PM

T6.3 BEOL-Compatible Oxide Semiconductor Twin-Gate FeFET Achieving a Record-High Memory Window of 25.4 V and a Near-Ideal Memory Window-to-Operation Voltage Ratio of 1.88, Xinming Li[1], Ning Ji[1], Leming Jiao[1], Yi Tong[1], Gan Liu[1], Chaoming Wu[1], Duy Hieu Trinh[1], Zhilun Zhang[1], Jiawei Sun[1], Kai Ni[2], Gengchiau Liang[3], Chen Sun[1], Dong Zhang[1], Xiao Gong[1]
[1]National University of Singapore, [2]University of Notre Dame, [3]National Yang Ming Chiao Tung University

We report, for the first time, an innovative twin-gate (TG) architecture combined with a meticulously designed interlayer demonstrated in BEOL-compatible oxide-semiconductor FeFETs (OS-FeFETs) with channel length (L_{CH}) as small as 30 nm. This design introduces multiple degrees of freedom in optimizing voltage distribution across the gate stack, leading to record-breaking performance. Specifically, a record-high memory window (MW) of 25.4 V among all reported FeFETs is achieved at an operation voltage (V_{op}) of 13.5 V, corresponding to a record-high MW/V_{op} ratio of 1.88 that approaches the theoretical limit. Moreover, our devices maintain MW exceeding 15 V after 10^7 cycles, effectively alleviating the long-standing trade-off between large MW and endurance in FeFETs.

2:45 PM

T6.4 Performance Improvement from Subtractive Ruthenium Interconnects with Airgap, Akshit Peer[1], Ananya Dutta[1], Saima Siddiqui[1], Jonathan Bunyan[2], Weijie Feng[2], YuWen Huang[2], Christopher Jezewski[1], Robert Jordan[1], Abinasha Kalita[2], Joshua Kevek[1], Emmannuel Khora[1], Brian Krist[1], Nikhil Mehta[2], Matthew Metz[1], Robert Seidel[1], En-Min Shih[2], Supanee Sukrittanon[2], Mauro Kobrinsky[1]
[1]Intel Foundry Technology Research, Intel Corporation, [2]Intel Foundry Technology Development, Intel Corporation

Subtractive ruthenium (sRu) interconnects have shown great promise for future interconnect pitch scaling, especially for pitch ≤ 22 nm. In this publication, we first demonstrate R&D results from a fully functional sRu and airgap (AG) interconnect architecture integrated with RibbonFET devices on an advanced R&D test vehicle. Implementation of sRu interconnects with airgaps resulted in approximately 2% ring oscillator frequency improvement at matched leakage, as compared to a damascene Cu reference, primarily driven by a significant line-to-line capacitance reduction. Airgap is shown to enable up to a 35% reduction of line-to-line capacitance. We also identified opportunities to improve via resistance by up to 50% for both upper and lower vias.

Session T7: Gate Stack Processes

1:30 PM, Tapa 2
Co-Chairs: Vamsi Paruchuri, ASM and Yoshiaki Kikuchi, Sony Semiconductor Solutions

1:30 PM

T7.1 Ultra-scaled high-k gate stacks (CET 0.9 nm) enabled by a low-thermal-budget (500 °C) oxygen-passivated interfacial layer (O-PAS IL), Yukinori Morita[1,2], Takamasa Kawanago[1,2], Takefumi Kamioka[1,2], Yuichiro Mitani[2,3], Toshihide Nabatame[2,4], Takashi Onaya[2,4], Naoki Fukata[2,4], Jevasuwan Wipakorn[2,4], Kazuhito Tukagoshi[2,4], Takuya Hoshii[2,5], Kasidit Toprasertpong[2,6], Atsushi Tamura[2,6], Koji Kita[2,6], Naoya Okada[1,2], Kenzo Manabe[1,2], Wataru Mizubayashi[1,2], Hiroyuki Ota[1,2], Takashi Matsukawa[1,2], Shinji Migita[1,2]
[1]SFRC, AIST, [2]LSTC, [3]Tokyo City University, [4]NIMS, [5]Science Tokyo, [6]The University of Tokyo

We propose an oxygen-passivated interfacial layer (*O-PAS IL*) technology as a novel scheme for forming thin oxides in environments where oxidation and etching coexist, and demonstrate an ultra-thin IL even in low thermal budget (500 °C) processes. To introduce an IL of < 0.3 nm, which is thinner than the native Si oxide films, we optimized the entire process of ultra-thin IL formation and subsequent gate stack formation to avoid unexpected IL increases. Using this novel scheme, a capacitance equivalent thickness of approximately 0.9 nm with highly reliable high-k gate stacks is achieved.

1:55 PM

T7.2 Dipole material-independent carrier mobility in dipole-first nMOS gate stack and introduction of multi-V_t patterning-tolerant dipole-middle integration for NS/CFET, Hiroaki Arimura[1], Leo Lukose[1], Jimmy Stiers[1], Stephan Brus[1], Jacopo Franco[1], Ju-Geng Lai[1], Mohamed Ben Chroud[1], Camila Cavalcante[1], Thomas Chiarella[1], Inge Vaesen[1], Thierry Conard[1], Matias Bejide[1], Jerome Mitard[1], Min-Soo Kim[1], Steven Demuynck[1], Lucas Petersen Barbosa Lima[1], Serge Biesemans[1], Naoto Horiguchi[1]
[1]imec

We report on an evaluation of dipole-first nMOS gate stack, and a new dipole multi-V_t gate stack scheme, called dipole-middle. On planar RMG nFET using dipole-first gate stack, we find a strong, shifter material- and stack-independent correlation among D_{it}, carrier mobility and V_{tlin}, indicating that mobility cannot be maintained unless T_{inv} is increased. Moreover, the shifter patterning on SiO_2 IL introduces damage in the dipole-first scheme. Thus, a novel dipole-middle scheme has been introduced, which protects IL during the patterning, while allowing to reduce drive-in thermal budget. EWF/V_t reduction of MOSCAP/single nanosheet nFET are demonstrated with LaO_x dipole-middle stack using low temperature lock-in anneal. This could provide a large benefit in dual dipole integration for multi-V_t CFET devices.

2:20 PM

T7.3 Mitigating intrinsic fixed-charge-induced V_{fb} fluctuation using a "neutral TiO_2 dipole" layer, Takefumi Kamioka[1,2], Toshihide Nabatame[2,3], Hiroyuki Matsukawa[2,4], Takamasa Kawanago[1,2], Yukinori Morita[1,2], Kasidit Toprasertpong[2,4], Yuichiro Mitani[2,5], Takashi Onaya[2,3], Naoki Fukata[2,3], Wipakorn Jevasuwan[2,3], Kazuhito Tsukagoshi[2,3], Takuya Hoshii[2,6], Atsushi Tamura[2,4], Koji Kita[2,4], Naoya Okada[1,2], Kenzou Manabe[1,2], Wataru Mizubayashi[1,2], Hiroyuki Ota[1,2], Takashi Matsukawa[1,2], Shinji Migita[1,2]
[1]AIST, [2]LSTC, [3]NIMS, [4]The University of Tokyo, [5]Tokyo City University, [6]Institute of Science Tokyo

Conventional dipole engineering with La_2O_3 and Al_2O_3 enables Vth modulation but suffers from increased fixed charges and limited controllability of small V_{fb} shifts, particularly when sub-monolayer dipole layers form non-uniform distributions. This study first demonstrates that these intrinsic fixed charges in metal-oxide/SiO_2 stacks lead to V_{fb} variability—not previously recognized as a major source —, which leads to hinder precise V_{th} tuning. In this work, we introduce a "neutral dipole" (ND) concept based on TiO_2, which produces no measurable dipole-induced V_{fb} shift on SiO_2. When combined with La_2O_3 or Al_2O_3, the resulting mixed dipole layers (LTO/ATO) provide small, well-controlled V_{fb} modulation with minimal CET penalty. The ND approach also suppresses device-to-device V_{fb} fluctuation. These findings show that ND engineering effectively addresses intrinsic fixed-charge issues and enables more precise V_{th} control for future transistor technologies.

2:45 PM

T7.4 Effective Dipole Strength Modulation using Imprinted Gadolinium Dipole (IGD) Layer to Achieve Fine V_t Tunability for Low Thermal Budget CFET Devices, Sangkuk Han[1], Wonjae Choi[2], Jaewon Chung[2], Haesoo Jang[2], Kyungwook Park[1], Sangmyun Lim[2], Soyoung Park[2], Sangwoo Jeong[2], Hanbyul Kim[3], Yongjoo Park[3], Jinho Ahn[1], Changhwan Choi[1]
[1]Division of Materials Science & Engineering, Hanyang University, [2]Department of Semiconductor Engineering, Hanyang University, [3]SK Trichem Co., Ltd.

We investigate dipole-first gate stacks using rare-earth oxides to enable multi-Vt for CFET devices. LaO_x and GdO_x DF stacks are systematically compared in terms of dipole strength, interface quality, trap charge density, and gate leakage. LaO_x exhibits strong dipole behavior, resulting in a V_t reduction of ~155 mV/cycle. GdO_x shows a moderate dipole effect with a negative Vt shift of ~80 mV/cycle. However, the 1 cycled-ALD of GdO_x induced an excessively strong dipole, leading to an abrupt V_t modulation. Based on this behavior, an imprinted Gd dipole gate stack is designed to further modulate the effective dipole strength of a subsequent LaO_x DF stack. The proposed IGD scheme achieves a fine Vt tunability of ~40 mV/cycle with negligible EOT penalty. In addition, the IGD stack

reduces the effective trap density by 92% and suppresses gate leakage current by more than one order of magnitude compared to the LaO_x DF stack.

Technical Sessions Block 6 (3:25 PM – 5:30 PM)

Session C17: High Speed ADCs

3:25 PM, Tapa 3
Co-Chairs: Trevor Caldwell, Alphawave Semi and Yun-Shiang Shu, MediaTek Inc.

3:25 PM
C17.1 A 12-bit 4-GS/s ADC with PLL-based Sub-ranging RO-TDC, Yimeng Wang[1,2], Xinyang Di[1,2], Jingxiang Wang[1,2], Hongjiang Chen[1,2], Ran Zhang[1,2], Sai-Weng Sin[1], R.P. Martins[1], Mingqiang Guo[1]
[1]State-Key Laboratory of Analog and Mixed-Signal VLSI and IME/ECE-FST, University of Macau, [2]UM Hetao IC Research Institute

This paper proposes a PVT-robust sub-ranging ring oscillator-based time-to-digital converter (SSRO-TDC) with phase-locked loop (PLL), integrated into a 12b 4-GS/s time-interleaved pipelined ADC as a backend. With frequency multiplication by PLL, the proposed background-running sub-ranging RO (SRRO) is shared among TI channels, enhancing time resolution and PVT robustness and improving power efficiency. The proposed offset calibration algorithm suppresses the time offset in SRRO. Measurement result shows an SNDR of 58.5 dB at Nyquist frequency with FoM_W of 15.94 fJ/conv.-step.

3:50 PM
C17.2 A 3.2GS/s 2nd-order Noise-Shaping-Pipelined ADC with 400MHz-BW 69dB-SNDR Featuring Gain/Quantizer-Embedded Parallel Voltage/Time-Domain Loop Filter, Yuhang Peng[1,2], Shizhan Mo[1,2], Liang Qi[3], Sai-Weng Sin[1], Rui P. Martins[1], Mingqiang Guo[1]
[1]University of Macau, [2]UM Hetao IC Research Institute, [3]Shanghai Jiao Tong University

This paper proposes a 3.2GS/s 2nd-order error-feedback (EF) Noise-Shaping-Pipelined ADC. Benefiting from the fast and parallel operation of the proposed gain/quantizer-embedded parallel voltage/time-domain loop filter, this work breaks the speed bottleneck of NS-ADC with high efficiency. The operating frequency of the loop filter is increased to more than 3.2 GS/s. Fabricated in 28nm CMOS, this ADC achieves 69dB SNDR over a 400Mhz bandwidth (BW) and only consumes 25.8mW, leading to a 170.9dB FoMs and a 14fJ/conv.-step FoMw. The prototype ADC achieves the highest BW among all discrete-time oversampling ADCs and the best FoM_S/FoM_W among all oversampling ADCs with BW > 250 MHz.

4:15 PM
C17.3 An 8bit 4.8GS/s 10.9fJ/conv.-step TI-SAR ADC with 4bit/3cycle Ping-Pong Architecture and Data-Driven Zero-Overhead Background Calibration , Changjoo Kim[1], Sooho Park[1], Minkyun Shim[1], Woojin Lee[1], Hyunjin Kim[1], Donghwi Seo[1], Jaeuk Lee[1], Chulwoo Kim[1]
[1]Korea University

This paper presents a 4-channel 4.8GS/s 8bit time-interleaved SAR ADC in 14nm FinFET that overcomes the throughput limitations of conventional ping-pong architectures. While traditional ping-pong schemes are limited to 1bit/cycle, the proposed 4bit/3cycle architecture utilizes a background-calibrated time reference using forbidden-code detection to accelerate conversion speed. Uniquely, the design features a data-driven zero-overhead background offset calibration scheme that exploits idle cycles identified during conversion, eliminating the throughput penalty of dedicated calibration phases. The ADC achieves FoM_W of 10.9fJ/conversion-step, demonstrating the best energy efficiency among state-of-the-art ADCs operating above 3GS/s.

4:40 PM

C17.4 A 14b 500MS/s Pipelined-SAR ADC with Loop-Unrolled Correlated Level Shifting and Rapid Sequential Sampling, Yong Huo[1], Yuefeng Cao[1], Minglei Zhang[1], Yan Zhu[1], Rui P. Martins[1], Chi-Hang Chan[1]
[1]State-Key Laboratory of Analog and Mixed-Signal VLSI, IME, University of Macau
This paper presents a 14b 500MS/s Pipelined-SAR ADC, which features the loop-unrolled correlated level shifting (LU-CLS) and rapid sequential sampling (RSS) to enable a swift, accurate and PVT-robust residue amplification at wideband inputs. The ADC achieves 70.12dB SNDR at the Nyquist input and consumes 7.6mW, leading to 175.3dB FoMs.

5:05 PM

C17.5 A 19.4mW 2.5GS/s Pipelined-SAR ADC Achieving 60.8dB SNDR Employing a Dual-Zero-Compensation Ring Amplifier, Jingxiang Wang[1,2], Kaize Zhang[1,2], Xiuyuan Zhao[1,2], Hongjiang Chen[1,2], Haoyu Li[1,2], Sai-Weng Sin[1], Rui P. Martins[1], Mingqiang Guo[1]
[1]University of Macau, [2]UM Hetao IC Research Institute

In this paper, a double-zero compensated ring amplifier is proposed, which introduces two left half-plane (LHF) real zeros to improve the phase margin. As a result, the proposed architecture achieves high open-loop gain and stable operation. The amplifier is integrated into a two-stage pipelined SAR ADC that achieves a signal-to-noise ratio (SNDR) of 60.8 dB at a 2.5 GS/s sample rate and Nqyuist rate inputs. This corresponds to a factor of merit (FoMs) of 168.9 dB and a factor of merit (FoMw) of 8.7 fJ/conversion step.

Session C18: Display Interfaces and Data Conversion Circuits

3:15 PM, Honolulu 2
Co-Chairs: Markus Dielacher, Infineon Technologies and Sai-Weng Sin, University of Macau

3:15 PM

C18.1 A 3200×2880, 120Hz OLED Display Driver IC with Miller-Charge Cancellation (MCC) Achieving a 13.3V/μs Slew Rate under Worst-Case Full-Channel Load, Suyun Chae[1], Taeksu Kwon[2], Yousung Park[1], Insuk Kim[2], Yeongmin Kim[2], Siwoo Kim[2], Hyun-Sik Kim[1]
[1]Korea Advanced Institute of Science and Technology (KAIST), [2]Samsung Electronics Co., Ltd.

This paper presents a 2880-channel 10b driver IC for 3200×2880 120Hz OLED displays. Miller-charge cancellation boosts the slew rate (SR) by 3.3× without static power, while a fast-slew input method relieves dynamic loading. Fabricated in 28nm CMOS, the 10μm-pitch chip attains a worst-case SR of 13.3V/μs (σ_{SR} = 7.5%) under full-channel transitions, improving color fidelity and uniformity on a product-level 6.9-inch OLED panel.

3:40 PM

C18.2 A 95.0 dB Dynamic Range Zero-Bias PV Light-to-Digital Converter for Seawater Monitoring with Single Point Calibration, Sanghyuck Moon[1,2], Ashfakh Huluvallay[1], Hyunwon Chung[1], Myeongsu Ko[1], Seokhyeon Jeong[1], Jungho Lee[1], Mohammad Khreishah[1], Jeongtaek Chang[1], Hung Do[1], Caitlyn Sutherland[3], Jason Kapit[3], David Nicholson[3], William Reinhardt[4], Dennis Sylvester[1], Mark Miskin[4], David Blaauw[1]
[1]Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, [2]System LSI Division, Samsung Electronics Co., Ltd., [3]Woods Hole Oceanographic Institution, [4]Electrical and Systems Engineering, University of Pennsylvania

We present a single chip, self-zero-biased light sensor for sea-water analysis in a wireless underwater vehicle. The sensor achieves linear irradiance-to-frequency conversion, single point calibration, and light-proportional power consumption. It has a measured dynamic range of 56,500× with temperature-robust operation in 180 nm CMOS.

4:05 PM

C18.3 A 28nm 2GS/s Remote-Sensing On-Chip Oscilloscope Using Charge-Redistributing Voltage Transfer and Time-Domain Stochastic ADC with 1.6mV Resolution, Minyoung Kang[1], Dongkwon Lee[1], Sangsu Jeong[1],

Kwanghyun Shin[1], Dongsuk Jeon[1]
[1]Seoul National University

This paper proposes a high-resolution, multi-point on-chip oscilloscope utilizing time-interleaved distributed sensors to minimize self-induced droop and a shared central processing unit with a stochastic time-domain analog-to-digital converter, combined with fast and power-efficient repeater-less charge-redistributing transmission. Fabricated in 28nm, the design achieves a high resolution of 1.6mV at 2GS/s over 0.59-1.1V, outperforming prior designs with minimal local sensor power consumption to avoid self-induced VDD droops.

4:30 PM
C18.4 An 11b Source-Driver IC with Resistor-Resistor-Embedded (RRE) DAC and Multi-Step Offset Calibration Achieving 1723 μm²/Channel and 1.6 mV DVO for 6285-PPI OLED-on-Silicon Displays, Chong-Hwa Bae[1], Hyeon-Ho Park[1], Wiman Yoo[1], Junghwan Oh[1], Dong-Hyun Kim[1], Gang-Bae Park[1], Mungyu Kim[2], Chankeun Kwon[2], Dongwan Ha[2], Jong-Seok Kim[1]
[1]Department of Electronic Engineering, Hanyang University, [2]Samsung Display

This paper presents an 11b (4b+4b+3b) source-driver IC (SD-IC) for OLED-on-Silicon (OLEDoS) displays that achieves a small channel area by using a three-stage resistor-resistor-embedded (RRE) DAC, deep-N-well (DNW)-based low-voltage DAC (LVDAC), and DEMUX structure. The proposed data-dependent current source (D^2CS), spare amplifier, R-share logic, and multi-step offset calibration (MOCU) suppress deviation of voltage output (DVO) under massive channel operation. In addition, a proposed high-pass-filter-assisted dynamic current source (HPDCS) enables 120 Hz frame-rate operation. The proposed D^2CS and global amplifier array with replica (GAAR) enable nonlinear gamma curve implementation. The proposed 11b SD-IC was fabricated in a 65 nm CMOS process. Despite supporting 11b resolution, the proposed SD-IC achieves the smallest channel area (1,723 μm^2 per channel), the lowest DVO (1.6 mV), and the fastest settling time (0.69 μs) under the 1000-channel load condition.

4:55 PM
C18.5 A 10-Bit Source-Driver IC with Mismatch-Resilient Cross-Point DAC Achieving 1223μm²/Channel and 4.3mV DVO for Mobile OLED Displays, Yousung Park[1], Hyeong-Joon Kim[1], Suyun Chae[1], Jeeyeon Eom[1], Hyun-Sik Kim[1]
[1]Korea Advanced Institute of Science and Technology (KAIST)

This paper presents a 10-bit source-driver IC for OLED displays. The cross-point (XP) DAC enables a compact layout footprint through symmetric centroid-based mismatch resilience. Also, the improved correlated double integration (I-CDI) not only eliminates DAC errors but also auto-zeros amplifier offset with minimal hardware overhead. The 272-channel SD-IC prototype, fabricated in 180nm CMOS, achieves a state-of-the-art area (107.6×11.3 μm^2/channel) and a 4.3mV DVO with a 1-H scan time of 1.5μs.

Session C19: Bio-Sensing and Molecular Interfaces

3:25 PM, Honolulu 3
Co-Chairs: Carolina Mora Lopez, imec and Yohei Nakamura, Hitachi, Ltd.

3:25 PM
C19.1 300-Channel Photometry Neural Probe with In-Pixel Adaptive-Noise ΔΣ Modulators Achieving 16.8 fA/√Hz Noise for High-Sensitivity Fluorescence Recording, Mengyu Li[1], Xiaolin Yang[1], Alejandro Lopez Rodriguez[2], Bastien Duckert[1], Pieter Neutens[1], Valentina Restrepo Jaramillo[3], Phillippe Coppejans[1], Carolina Mora Lopez[1]
[1]imec, [2]EPFL, [3]Universidad de Antioquia

This work presents a 300-channel photometry neural probe with in-pixel readout for scalable, high-resolution fluorescence sensing. Each channel integrates an on-chip photodiode and a ΔΣ-based adaptive-noise readout with serialized data transfer. By exploiting a duty-cycled unipolar current DAC (IDAC) with a passive integrator, the readout achieves an adaptive noise floor with a compact implementation. The input-referred noise decreases at lower signal levels, enabling an 88.7dB dynamic range which is 12dB higher than the peak SNR. The system achieves an ultra-low noise floor of 16.8fA/√Hz while consuming 70nW per channel. Each channel occupies 0.0015mm^2 with a 25µm pitch, supporting high-spatial-resolution neural recording along the probe shank.

3:50 PM

C19.2 A Scalable, Unified Current-Mode ΔΣ Redox/pH Readout with Shared Potentiostat and Flicker-Noise Cancellation for High-Density Electrochemical Sensing, Joan Aymerich[1], Gerald Topalli[2], Jose Cisneros Fernandez[1], Javier Cuenca Michans[3], Valentina Restrepo Jaramillo[4], Chutham Sawigun[1], Carolina Mora Lopez[1]
[1]imec, [2]Rice University, [3]UAB, [4]Universidad de Antioquia

This work presents an area-efficient electrochemical-sensing platform based on current-mode delta-sigma modulators (I-DSMs) enabling scalable in-pixel neurochemical sensing. It supports amperometric redox and ISFET-based ion/pH sensing using 1st-order, 1-bit I-DSMs that directly digitize current in sub-0.01-mm^2 per-channel footprints. The architecture employs a shared multi-channel potentiostat, an auto-zeroed current DAC suppressing 1/f noise and mismatch, and a switchless back-to-back (B2B) diode mitigating ISFET offset and drift. A prototype in 130-nm CMOS validates the architectures, achieving 100-dB dynamic range (DR) with 12.6-pA$_{rms}$ noise for amperometry and 0.016-pH resolution at nW-level power for the ISFET.

4:15 PM

C19.3 An 83.5dB DR, 74pA$_{RMS}$ Neurotransmitter Sensor Interface with Drift Correction, Wangbo Chen[1], Jose R Lopez-Ruiz[1], Patrick McCommons[1], Noah Leonardo[1], Jill Becker[1], Sung-Yun Park[1,2], Michael P Flynn[1], Euisik Yoon[1]
[1]University of Michigan, Ann Arbor, [2]Pusan National University

State-of-the-art neurotransmitter sensors use fast-scan cyclic-voltammetry (FSCV) to achieve high sensitivity and selectivity. The main challenge in FSCV is the need for high dynamic range (DR) and a low noise floor. We address this challenge by introducing a dynamic template-tracking sensor interface which takes advantage of the quasi-repetitive characteristics in non-faradaic (NF) currents. The prototype has a DR of 83.5dB, a 73.8pA$_{RMS}$ noise floor, and is resilient to sensor drift. This work reports the highest DR and maintains simplicity of a one-working-electrode/channel setup.

4:40 PM

C19.4 A 16-channel 97pArms 1MHz Bandwidth IC for Solid State Nanopore Sequencing, Qiuyang Lin[1], Wim Sijbers[1], Yixiong Hu[1], Eric Beamish[1], Wouter Botermans[1], Carolina Mora Lopez[1], Nick Van Helleputte[1]
[1]imec

A fully integrated 16-channel nanopore interface IC fabricated in 0.13µm CMOS is presented. The design achieves 20MΩ transimpedance, 1MHz BW, and 97pArms noise at 1.7mW per channel, and demonstrates reliable parallel detection of labeled DNA translocation events using solid-state nanopores.

5:05 PM

C19.5 Filter-free Time-gated Fluorescence CMOS Biochip with Fast Thermal Cycling for PoC DNA Testing, Nicholas Vitale[1,2], Davood Zare[1], Mark McDermott[1], Arjang Hassibi[1], Tom Lee[2]
[1]Siomyx Inc., [2]Electrical Engineering, Stanford University

A 16×24 time-gated fluorescence CMOS biosensor array is presented for nucleic acid analysis. Time-interleaved measurement of excitation and fluorescence emission enables each pixel to achieve 9.9 fA input-referred fluorescence photocurrent sensitivity, 129 dB detection dynamic range, and >133 dB suppression of combined excitation and autofluorescence background, eliminating the need for an optical emission filter. Integrated on-chip

heating with passive cooling supports thermocycling with heating and cooling rates exceeding 18 °C/s, as well as precision melt-curve analysis for immobilized DNA probes on pixel.

Session T8: DRAM

3:25 PM, Tapa 2
Co-Chairs: Gary Bronner, Rambus and Vita Pi-Ho Hu, National Taiwan University

3:25 PM
T8.1 3D Orthogonal Die Stacking Technology for DRAM-on-GPU Integration Using Contactless Die-to-Die Interface, Yuki Mitarai[1], Masaya Kawano[1], Hung-Chih Huang[1], Takafumi Fukushima[2], Tadahiro Kuroda[1,3], Mototsugu Hamada[1], Atsutake Kosuge[1,3]
[1]University of Tokyo, [2]Tohoku University, [3]RIKEN

Massive orthogonal stacking assembly of IC (MOSAIC) technology was demonstrated to overcome thermal limitations in 3D memory-on-GPU configurations. By stacking memory dies orthogonally to the GPU, heat is dissipated effectively, enabling twice the memory capacity. A contactless die-to-die interface, achieving up to 4 Gbps/ch in a prototype, mitigates assembly challenges, enabling TSV-free 3D integration.

3:50 PM
T8.2 Highly stackable 3D DRAM of Dual-gate IGZO 2T0C with Record 3 bits/cell and 400s Data Retention, Fuxi Liao[1], Zhengyong Zhu[2], Guanhua Yang[3], Chao Zhao[2], Menggan Liu[4], Xuanming Zhang[4], Kaifei Chen[4], Wendong Lu[4,5], Zihan Li[4], Naide Mao[4], Bok-Moon Kang[2], Jinghong Shi[2], Meichen Jin[2], Chang Liu[2], Jing Liang[2], Gui-Lei Wang[2], Congyan Lu[4], Jinshan Yue[4], Lingfei Wang[4], Jiawei Wang[4], Di Geng[4], Nianduan Lu[4], Chao Zhao[2], Arokia Nathan[6], Ling Li[7], Ming Liu[4]
[1]State Key Lab of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics, Chinese Academy of Sciences(CAS), [2]Beijing Superstring Academy of Memory Technology, [3]Key Lab of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics, Chinese Academy of Sciences(CAS), [4]Institute of Microelectronics, Chinese Academy of Sciences(CAS), [5]Anhui University, [6]Cambridge University, [7]Key Lab of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics, Chinese Academy of Sciences(CAS)

For the first time, we have experimentally demonstrated a 4-tiers 3D dual-gate IGZO 2T0C DRAM for ultra high-density memory. Benefiting from the DG control, the read transistor in both 4-tier 3D DG 2T0C cell exhibit high Ion of 12 μA/μm and low SS of 103mV/dec. High stability of 80 mV PBTI and -30 mV NBTI is achieved under 85℃ tests for 1 hour. The 3D DG 2T0C performance basic read/out operation with ultra-long data retention of 400s and a record 3-bit multi-bit storage.

4:15 PM
T8.3 A High-Performance Heterogeneous 2T0C DRAM with PSLC-Si/IGZO Complementary Gain Cell for Monolithic 3D Leveraging Capacitive Coupling, Jongyoun Park[1], Heejae Jeong[2], Jiwoo Won[2], Junhyeok Lee[1], Chunghyun Ahn[1], Sangsu Lee[1], YoungWoo Cho[1], Geuntae Park[2], Sangho Yu[2], Hyuk-Jun Kwon[2], Hyun-Yong Yu[1]
[1]Department of Electrical Engineering, Korea University, [2]Department of Electrical Engineering and Computer Science, Daegu Gyeongbuk Institute of Science and Technology

This study experimentally demonstrates a high-performance monolithic 3D (M3D) integrated 2-transistor-0-capacitor (2T0C) DRAM cell utilizing a complementary gain cell (CGC) with n-IGZO and p-channel PSLC-Si. Using patterned seedless laser crystallization (PSLC), we achieved a world-record grain size of 32.3 μm and a hole mobility of 265 cm^2/Vs for the read-transistor (R-Tr). The CGC architecture leverages capacitive coupling (C.C.) to boost the sensing margin (SM) to ~10^6, satisfying the 1k-array criterion (SM > 10^4) and outperforming conventional 2-nIGZO 2T0C. Simulation confirms an estimated read time of 11.8 ns. These findings establish the M3D CGC 2T0C DRAM as a candidate for high-density, and high-performance 3D memory solutions that overcome the limitations of conventional architecture.

4:40 PM
T8.4 Vertical-Die (V-die) 3.5D Integration for Cool Ultrahigh-Bandwidth Memory Systems, Heesoo Yang[1], Hyeongjun Kim[2], Yurim Choi[1], Seungmin Lee[1], Dongyun Kam[1], Haksoon Jung[1], Yongwoo Lee[1], Seongju Kim[3], Jimin Kwon[1,2,4]
[1]Department of Electrical Engineering, Ulsan National Institute of Science and Technology (UNIST), [2]School of Semiconductor Materials and Devices Engineering, Ulsan National Institute of Science and Technology (UNIST), [3]Department of Creative Convergence Engineering, Hanbat National University, [4]School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST)

This work explores a 3.5D V-die memory architecture as an alternative integration paradigm, where conventionally fabricated DRAM dies are rotated upright to exploit three-dimensional space at the system level. We assess the feasibility of the V-die concept by benchmarking its achievable bandwidth, capacity scalability, and thermal behavior against state-of-the-art HBM. By standing dies upright to expand exposed surface area and enabling direct liquid cooling between adjacent dies, V-die supports substantially higher thermal design power while unlocking scalable I/O density and capacity beyond conventional HBM.

5:05 PM
T8.5 Electrical Characteristics of the $4F^2$ Vertical Gate (VG) DRAM integrated with Bit-Line Shielding (BLS) and Back Gate (BG) Transistor, Joodong Park, Seung Wan Chu, Junho Cheon, Jinsun Cho, Eunhyup Doh, Jungmin Han, Seung Bum Hong, Choong-ki Kim, Daeik Kim, Jin Ar Kim, Sunghyun Kim, Yongtaik Kim, Kyongsoo Kum, Sein Kwon, Dong Jae Lee, Dong Ryeol Lee, Junghak Lee, Eunshil Park, Jongbum Park, Dong Hee Shin, Eunji Shin, Jino Song, Minchul Sung, Wansik Yoo, JeongTae Hwang, Seungbum Kim, Kyoungchul Jang, Youngmann Cho, Seonyong Cha
R&D Division, SK Hynix

This paper reports the electrical characteristics of $4F^2$ vertical gate (VG) DRAM such as data retention and sensing capability. The proposed VG DRAM incorporates key technologies including bit-line shielding (BLS) to suppress coupling noise and shared back gate (BG) structure to enhance threshold voltage (Vth) control and passing gate effect. Die-thinning technology is also introduced to support reliable circuit operation in wafer-bonded structure. Measured results demonstrate the feasibility of $4F^2$ VG DRAM architecture providing a viable development path toward competitive future DRAM products.

Session T9: Ferroelectric Materials for Memory

3:25 PM, Honolulu 1
Co-Chairs: Olivier Weber, ST Microelectronics and Shosuke Fujii, Kioxia Corporation

3:25 PM
T9.1 Origin and Modeling of the Access-Rate-Dependent Degradation and Dual-Interface Optimization in Ultra-Thin Nanolaminate HZO, Chen-Yi Cho[1,2], Yu-Ling Shih[1], Ming-Yeh Lao[1], Yu-An Chen[1], Yu-De Lin[2], Po-Chun Yeh[2], Shyh-Shyuan Sheu[2], Tuo-Hung Hou[1]
[1]National Yang Ming Chiao Tung University, [2]Industrial Technology Research Institute (ITRI)

Access-rate (AR)-dependent degradation under a non-continuous cycling (NCC) with hold times between operations is a new showstopper for the practical use of $Hf_{1-x}Zr_xO_2$ (HZO)-based FeRAM. We provide new insights into this critical reliability issue from the following perspectives: (1) reveal strong thickness- and hold-polarity-dependent AR degradation originating from interface property and asymmetry down to the ultrathin thickness of 4 nm, (2) develop a comprehensive physics-based AR degradation model that bridges continuous cycling (CC)-to-NCC behavior based on the interplay between charge trapping and depolarization-field (E_{dep})-driven redistribution, and (3) demonstrate an optimized symmetric ZHZ tri-layer nanolaminate based on new physics understanding and largely mitigate AR-dependent degradation.

3:50 PM

T9.2 Improved disturb-free memory window, retention and endurance in low-voltage 3D-FeFETs by controlling strain and trap properties in interfacial layers, Kunifumi Suzuki[1], Kiwamu Sakuma[1], Reika Ichihara[1], Viktoria Schlykow[1], Hiroki Tokuhira[1], Rieko Tanaka[1], Natsuki Iguchi[1], Takuma Jike[1], Kazuhiro Matsuo[1], Daisuke Hagishima[1], Makoto Fujiwara[1], Masumi Saitoh[1]
[1]Frontier Technology R&D Institute, Kioxia Corporation

A large memory window (MW), with suppressed read disturbance, good retention, and high endurance, was demonstrated in 3D-FeFETs, enabling both QLC and low-latency applications. Through in-depth charge analysis, we clarified that read disturbance, which is one of the major issues in both QLC and low-latency operations, occurred when discharging is followed by depolarization. Significant suppression of depolarization was achieved by engineering the charge de-trapping rate in both the gate-side and channel-side interfacial layers. We also found that thickness control of the gate interfacial layer effectively induced strain in the ferroelectric layer, enhancing remanent polarization. These optimizations enabled 3D-FeFETs to achieve a large MW of 7.8V after 10^5 P/E cycles and a disturb-free MW exceeding 6V after 10^7 disturbances without retention loss, which is sufficient for low-voltage and low-latency NAND.

4:15 PM

T9.3 First Demonstration of Highly-Scaled Trench-Structured Ferroelectric Capacitive Memory with Record-High Capacitive Memory Window and Integrated ITO Sensing FET, Jingguang Lu[1,2], Zuopu Zhou[3], Leming Jiao[1], Zijie Zheng[1], Jiawei Xie[1], Weibing Hao[1], Umesh Chand[4], Akhil Ranjan[4], Yu-Chieh Chien[4], Khee Yong Lim[2], Gengchiau Liang[5], Xiao Gong[1,4]
[1]ECE, National University of Singapore (NUS), [2]Technology Development, Globalfoundries, Singapore, [3]School of Electronic Science and Technology, Xiamen University, China, [4]IME, Institute of Microelectronics, Agency for Science, Technology and Research (A*STAR), Singapore, [5]Industry Academia Innovation School, National Yang Ming Chiao Tung University

A highly-scaled inversion-type FCM is demonstrated with trench structure on silicon and scaled to minimum 0.02 μm^2. The MFS structure combined with high-aspect-ratio trench enables a >10× capacitance ratio and a record-high capacitance difference of ~19 $\mu F/cm^2$. Integrating an ITO sensing FET with FCM, for the first time, we experimentally demonstrate the sensing of individual highly scaled FCM, validating the scalability of this emerging memory

4:40 PM

T9.4 Engineering 3D HZO ferroelectric capacitors to scale down 22nm embedded FeRAM, Carine Jahan[1], Simon Martin[1], Romuald Blanc[1], Gustav Persson[1], Fabien Grimaud[1], Mélanie Louro[1], Liam Hosier[1], Nicolo Giovannelli[1], Théo Monniez[1], Grégory Boniface[1], Olivier Glorieux[1], Christelle Boixaderas[1], Antonio Roman[1], Mehdi Oujanba[1], Messaoud Bedjaoui[1], Julien Mercier[1], Stéphane Minoret[1], Catherine Euvrard[1], Julian Sturm[1], Cécile Candegabe[1], Sébastien Kerdilès[1], Corinne Comboroure[1], Estelle Guyez[1], Stéphane Pocas[1], Denys Ly[1], Sébastien Martinie[1], Yann Beilliard[1], Gabriel Pares[1], Fabrice Bernard-Granger[1], Jean Coignus[1], Nicolas Bernier[1], Olivier Billoint[1], François Andrieu[1], Laurent Grenouillet[1]
[1]CEA-Leti

This paper demonstrates 22nm FeRAM arrays featuring for the first time at this node 3D HZO ferroelectric capacitors (FeCap), BEOL-integrated with 2 different schemes. Array functionality with gaussian bit distributions is reported down to 0.047μm^2 1T-1C bitcell at 1.3V with bitcells featuring a logic transistor and a 0.0078 μm^2 footprint 3D FeCap with an aspect ratio (AR) of about 4:1. To improve further both MW and density, 17:1 AR 7nm HZO 3D FeCaps with 0.0028μm^2 footprint are demonstrated with measured remanent polarization 2.Pr of 1100$\mu C/cm^2$ (normalized by footprint) at 1.5V and wake-up free behavior in agreement with a ~80% orthorhombic phase fraction as measured by precession electron diffraction (PED). These results pave the way to ultra-dense embedded FeRAM at nodes below 22nm.

5:05 PM

T9.5 Record-Thin (~2 nm) Wakeup-Free HZO Ferroelectric MFM via Atomic Layer Etching (ALE): Grain Size Enlargement and Ferroelectric Phase Engineering , Xiaopeng Li[1,2], Chaoming Wu[1], Duy Hieu Trinh[1], Xiaoyu Dou[2], Jixuan Wu[2], Wei Shi[1], Xinming Li[1], Ning Ji[1], Zhilun Zhang[1], Dong Zhang[1], Jiezhi Chen[2], Xiao Gong[1,3]
[1]National University of Singapore (NUS), [2]Shandong University, [3]Agency for Science, Technology and Research (A*STAR), Singapore

In this work, we report the world's thinnest 2 nm $Hf_xZr_{1-x}O_2$ (HZO) metal-ferroelectric-metal (MFM) capacitor, achieving 2Pr of ~20 μC/cm^2 at 0.65 V, wakeup-free operation, ultra-low leakage, and a record-high E_{BD} of ~11.3 MV/cm. This breakthrough is enabled by a series of process innovations, highlighted as follows: (1) A novel Atomic Layer Etching (ALE)-based grain size enlargement and ferroelectric (FE) phase engineering approach is proposed and demonstrated, overcoming the physically-intrinsic grain size limitation in ultra-thin (sub-2.5 nm) HZO films. (2) A novel Fluorine Treatment (F treat.) is introduced to repair oxygen vacancies (Vo) for further performance enhancement. This work pushes the frontier of FE memory scaling and significantly enhances the competitiveness of FE memory as the mainstream next-generation ultra-low power memories.

Session TFS2: Advanced 3D Logic

3:25 PM, Tapa 1
Co-Chairs: Hsinyu (Sidney) Tsai, IBM and Kazuyuki Tomida, Rapidus Corp.

3:25 PM

TFS2.1 CFET Demonstration for Future Logic and SRAM Technology, Szuya Sandy Liao[1], Lu Yang[1], Wei-Xiang You[1], Ting-Yun Wu[1], Tsung-Kai Chiu[1], Jui-Chien Huang[1], Kelvin Elphick[1], Ku-Feng Yang[1], Chu-Hsuan Sha[1], Guan-Ren Wang[1], J.H. Jhanf[1], Y.R. Joy Cheng[1], Ming-Feng Shieh[1], Chia-Min Chang[1], Huang-Ming Chen[1], Pei-Ren Jeng[1], Shou-Yi Chang[1], M.H. Matt Yeh[1], Weng Chang[1], Simon Jang[1], Min Cao[1]
[1]TSMC

The Complementary Field-Effect Transistor (CFET) architecture offers a pathway for extending Moore's Law beyond state-of-the-art nanosheet CMOS scaling. This paper reviews our prior CFET device demonstrations on 300mm silicon, progressing from nFET-on-pFET stacked transistors to 101-stage ring oscillators and the world's most compact 6T SRAM bit-cells. In this work, we further present two key milestones: first, doubling number of stacked nanosheets to boost ring oscillator frequency by +15%; and second, enabling multi-Vt integration capability through dipole patterning approaches, achieving six distinct Vt values (3-Vtn and 3-Vtp) with a 160mV/120mV span range for N/P FETs on the same wafer.

3:50 PM

TFS2.2 Backside Power: Enabling Energy Efficient Performance on Advanced Node Designs, Eric Karl[1], Shishir P Agrawal[1], Som Bangalore Prakash[1], Seung-june Choi[1], Kevin J Fischer[1], Lei Jiang[1], Daniel Pantuso[1], Ramesh K Subramaniam[2], Xiaofei Wang[1], Xinning Wang[1], Tai-Hsuan Wu[1]
[1]Intel Foundry, Intel Corporation, [2]Silicon & Platform Engineering, Intel Corporation

Backside power delivery (BS-PDN) decouples signal routing from power delivery, reducing interconnect scaling cost/complexity. Key benefits include 11% routed area reduction and 10x dynamic droop reduction, enabling 5–6% frequency uplift or >15% dynamic power reduction vs. comparable frontside interconnect technology.

4:15 PM

TFS2.3 A Product-Oriented 3rd-Generation 2nm GAA Platform Technology with Enhanced Speed and Power Efficiency across Nanosheets, Seulki Park[1], Dongchan Jeong[1], Sanghoon Lee[1], Seungkwon Kim[1], Minseong Lee[1], Byungha Choi[1], Sada-aki Masuoka[1], Chang-Woo Sohn[1], Sung Won Kim[1], Younghun Jung[1], Taeyoung Kim[1], Ju-Youn Kim[1], Kichul Park[1], Hyunjo Kim[1], Jaehun Jeong[1], Hee Sung Kang[1], Yuri Yasuda-Masuoka[1], Ja-Hum Ku[1]
[1]Foundry Business, Samsung Electronics Co., Ltd.

The state-of-art 3rd generation 2nm GAA platform has been developed to offer further Power-Performance enhancement with better DTCO options. To boost intrinsic performance in a wide range of nano-sheet, novel Si Surface roughness control and enhanced S/D strain technique have been introduced to obtain +10~12% performance. For further product level gain, significantly improved MOL/BEOL resistance -30~40% and capacitance -15~17% are also offered simultaneously. SF2P GAA provides the best PPA matrix with DTCO solution.

4:40 PM
TFS2.4 Backside Split-VDD Write-Assist Driven SRAM Macro Scaling in Nanosheet-Era, Ankit Singh[1], Dawit Abdi[1], Pieter Weckx[1], Aishwarya Singh[1], Arvind Sharma[1], Dwaipayan Biswas[1], Fernando Garcia-Redondo[2]
[1]imec, [2]imec UK

SRAM's scaling at advanced CMOS nodes requires robust write-assist schemes to ensure reliable write operations. This reliance becomes even more critical due to aggressive voltage scaling at advanced CMOS nodes for Vmin applications. This work leverages the backside routing offered in A14 nanosheet technology to enable backside split-VDD implementation without impacting bitcell footprint scaling. This technique offers enhanced writability without requiring self-timed circuits or large capacitors used in conventional negative-bitline write-assist. The proposed approach supports up to 512 rows per bitline and delivers a 40% improvement in write energy along with macro level area savings of 12% area per IO column, compared to these conventional assist schemes. The proposed design enables robust SRAM operation at 0.7 V and delivers 100 mV improvement in bitcell write trip point and 200 fJ write energy per bit for 512 rows per bitline with 86% array efficiency and 44.9 Mb/mm2 density for 512x256 subarray.

5:05 PM
TFS2.5 Area and Performance of Staggered-Channel Nanostack SRAM Bitcells, Chen Zhang[1], Carl Radens[1], Utkarsh Bajpai[1], Shay Reboh[1], Shahrukh Khan[1], Matt Deming[1], Nikhil Jain[1], Nick Lanzillo[1], Indira Seshadri[1], Debarghya Sarkar[1], Yann Mignot[1], Nirmaan Shanker[1], Stuart Sieg[1], Imran Younus[1], Shirin Akter[1], Feng Liu[1], Uzma Rana[1], Susan Fan[1], Ravi Ramachandran[1], Tenko Yamashita[1], Dechao Guo[1], Huiming Bu[1]
[1]IBM Research

A staggered-channel SRAM bitcell design in the Nanostack sequential stacking architecture is presented. We show the role of top-bottom gate merge contact, a unique requirement in sequential stacking, in determining cell height and how the staggered design where top and bottom channels are not aligned helps scaling. We also show for the 1st time in the world the gate merge structure developed on silicon. >40% cell height reduction versus the state-of-the-art non-stacked cells can be achieved within the current patterning capability. Finally, simulation shows Nanostack bitcell has an improved wordline (WL) performance and comparable bitline (BL) performance compared to non-stacked bitcells.

Session Banquet: Hawaiian Luau and Banquet

7:00 PM, Grand Lawn

Included with full conference registration but ticket is required for entry.

Thursday, June 18, 2026

Technical Sessions Block 7 (8:00 AM – 9:40 AM)

Session C20: Next-Generation Optical Transceivers

8:00 AM, Tapa 2
Co-Chairs: Harold Pilo, Cadence and C. Patrick Yue, Hong Kong University of Science and Technology & Stanford University

8:00 AM
C20.1 An 800 Gbps/Fiber Silicon Photonic Microring-Based DWDM Transceiver in an Open-Cavity Package, Cooper S Levy[1], Jahnavi Sharma[1], Zhe Xuan[1], Duanni Huang[1], Junyi Gao[1], Songtao Liu[1], Xinru Wu[1], Xiaoxi Wang[1], Susnata Mondal[1], Sashank Krishnamurthy[1], Dan Lake[1], James E Jaussi[1]
[1]Intel Corporation

Silicon photonic (SiPh) micro-ring resonator (MRR) based optical interconnects integrated with XPU/switch packages can enable very high data rates per-fiber through dense wavelength division multiplexing (DWDM). We demonstrate an O-band DWDM link with simultaneous 16-λ transmission at 50 Gbps/λ, for an aggregate data-rate of 800 Gbps/fiber, with BER<1e-9. An Intel 22 nm CMOS electronic integrated circuit (EIC) is 3D assembled with an Intel Fab11X photonic IC (PIC) and placed in an open-cavity substrate. The PIC integrates a 16-λ laser, λ-interleavers, MRRs for modulation and λ-demultiplexing, and optical amplification for link margin. The 22nm EIC features several high-speed techniques to push per-λ data rate.

8:25 AM
C20.2 A 32Gb/s Optical Receiver utilizing a Differential TIA with -17.3dBm Sensitivity in a 3D-stacked Silicon Photonics Platform, Georgios Kalogerakis[1], Sanquan Song[1], Li Xu[1], Nandish Mehta[1], Angad Rekhi[1], Nikola Nedovic[1], Benjamin G. Lee[1], Brian Zimmer[1], Steve Tell[1], Yoshinori Nishi[1], Xi Chen[1], Ward Lopes[1], Trey Greer[1], C. Thomas Gray[1]
[1]Nvidia

This work presents a 32Gb/s optical receiver that utilizes a differential transimpedance amplifier (TIA) from a single supply in a 3D-stacked silicon photonics (SiPh) platform. The receiver sensitivity (in terms of Optical Modulation Amplitude or OMA) at the photodiode (PD) is -17.3dBm and -18.9dBm at 32Gb/s and 28Gb/s, respectively. The energy efficiency at 32Gb/s is 0.484pJ/b. The receiver consists of a 7nm FinFET CMOS electronic IC (EIC) stacked on top of a 65nm silicon photonics IC (PIC) via Cu-Cu hybrid bonding.

8:50 AM
C20.3 A 0.69-pJ/b 4x80-Gb/s MRM-Based Coherent Optical Transmitter with Time-Multiplexed Thermal Tuning, Pengyu Zeng[1], Marziyeh Rezaei[1], Daniel Sturm[1], Asha Rashmi Nayak[1], Scott Li[1], Sajjad Moazeni[1]
[1]University of Washington

Co-packaged optics (CPO) is a critical technology for scaling bandwidth density in AI compute clusters. To maximize spectral efficiency and minimize the required number of laser lines, we demonstrate a 4-channel coherent optical transmitter using micro-ring modulators (MRMs) capable of 4x80 Gb/s QAM-4 modulation. Implemented in GlobalFoundries (GF) 45nm monolithic silicon photonics, the SoC achieves a 320 Gb/s aggregate data rate on a single wavelength. The design incorporates an on-chip PLL, high-speed clock distribution, and a full time-multiplexed thermal tuning digital control loop. Achieving an energy efficiency of 0.69-pJ/b, our approach can be either deployed as a lower power and latency alternative to PAM-4 or scaled up to QAM-16 for higher data rates in future AI fabrics.

9:15 AM

C20.4 A 0.416-pJ/b, 32-Gb/s 3D-Stacked Optical NRZ Receiver with -18.5-dBm OMA Sensitivity Using Self-Timed Decision Feedback Equalization, Li Xu[1], Sanquan Song[1], Nikola Nedovic[1], Georgios Kalogerakis[1], Nandish Mehta, Angad Rekhi, Brian Zimmer, Stephen G. Tell, Yoshinori Nishi, Xi Chen, Ward Lopes, Benjamin G. Lee, Thomas H. Greer III, John W. Poulton, C. Thomas Gray

NVIDIA

This work presents a 32Gb/s optical non-return-to-zero (NRZ) receiver with 3D-stacked 7nm CMOS and 65nm photonic dies. It uses a low-bandwidth front-end followed by self-timed decision feedback equalization (STDFE). OMA sensitivity for BER<10^{-12} and PRBS31 is -18.5dBm. The receiver consumes 0.416pJ/b at 32Gb/s and occupies 3360μm^2.

Session C21: SoCs for Reasoning and Robotics

8:00 AM, Tapa 3

Co-Chairs: Ben Keller, NVIDIA and Yukihiro Sasagawa, Socionext Inc.

8:00 AM

C21.1 TinyNPU: A 3nm 0.06-134.36 μJ/token DCIM-Based Ultra Low Power NPU for Always-On Reasoning on Wearables, Yi-Yen Hsieh*[1], En-Jui Chang*[1], Chieh-Fang Teng*[1], Po-Hua Huang[1], Hsien-Kai Kuo[1], Hsing-Chang Chou[1], Kai-Siang Yang[1], Shiyi Ou[1], Chiao-Chun Huang[1], Cheng-Xin Xue[1], Wei-Ting Hung[1], Zhe Wan[1], Hung-Wei Lin[1], Gajanan Jedhe[1], Yi-Chen Chu[1], Tun-Yi Chang[1], Chia-Hsien Chou[1], Kun-Geng Lee[1], Perry H. Wang[1], Chetan Deshpande[1], Jenwei Liang[1], Chien-Hung Lin[1], Tsung-Yao Lin[1], Yucheun Kevin Jou[1]

[1]MediaTek Inc.

TinyNPU is the first commercial 3-nm DCIM NPU for ultra-low-power always-on reasoning (AoR) in wearables. By co-optimizing the DCIM architecture, an L1-efficient composite flow, and software-hardware co-optimized scheduling, it delivers 1.47 TOPS with 512 8-bit MACs and 256 KB on-chip memory. Measurement results show up to 10 days of battery life on smart glasses and 31.8× less energy on Transformer models vs. state-of-the-art works, enabling efficient, responsive AoR for next-generation wearables.

8:25 AM

C21.2 μAgent: A 7nm 404.3mJ/Action Edge SoC for Agentic AI Reasoning with Adaptive Block-wise Mixed-Format 4-bit Quantization and Decoupled Access-Execute, Wonsuk Jang[1], Christian Kubicka[1], Allen Pan[1], Pullabhatla Smriti[1], Win-San Khwa[2], Meng-Fan Chang[2], Priyanka Raina[1], Thierry Tambe[1]

[1]Electrical Engineering, Stanford University, [2]Taiwan Semiconductor Manufacturing Company (TSMC)

μAgent is a 4 mm^2 TSMC 7 nm SoC enabling large language model (LLM)-based agentic AI reasoning on edge devices, with only 64MB of off-chip HyperRAM. Its LLM accelerator, μThinker, reduces external memory access via block-wise mixed-format 4-bit quantization with adaptive block/element size. μThinker efficiently couples to the CPU via an ARM Coprocessor interface and decoupled access-execute streams, minimizing communication stalls. Operating up to 1GHz, μAgent achieves 6.8TOPS/W while outperforming standard format (INT4/FP4) quantization on navigation task.

8:50 AM

C21.3 Sirius: A Dual-Chiplet System for Multimodal Embodied AI with Heterogeneous RVV Cores, Dense and Sparse Accelerators, Vikram Jain[1], Hasan Nazim Genc[1], Coleman Hooper[1], Minh Nguyen[1], Rahul Kumar[1], Rohan Kumar[1], Di Wang[1], Kevin Anderson[1], Yufeng Chi[1], Shengjun Kris Dong[1], Dima Nikiforov[1], Bob Zhou[1], Borivoje Nikolic[1], Yakun Sophia Shao[1]

[1]Electrical Engineering and Computer Science, University of California, Berkeley

Sirius is a dual-chiplet system targeting multimodal embodied AI on resource-constrained platforms. It integrates heterogeneous RISC-V Vector (RVV) cores with dense and sparse accelerators to efficiently execute

diverse workloads, including transformers, vision backbones, and robotics pipelines. Sirius addresses deployment challenges through three innovations: (1) support for diverse non-uniform quantization schemes via programmable LUTs and outlier-aware dense and sparse (D+S) decomposition, (2) a logically shared, physically distributed scratchpad memory system enabling kernel fusion and heterogeneous streaming, and (3) chiplet-based disaggregation using a custom interconnect. Fabricated in Intel16, Sirius delivers up to 2.24 TFLOPS/W for dense GeMM, 7.11 TFLOPS/W for SpMV, and achieves end-to-end 16 tokens/sec at 23 mJ/token on Qwen2.5-0.5B SLM, and 0.97 TFLOPS/W for robotics perception.

9:15 AM
C21.4 A Scalable Vision-Language-Action Edge Processor with Token Filtering and Hybrid Processing-Near-Memory for 6.7-ms Perception-to-Action Latency in Robot Control, Jin Wang[1], Zhengke Yang[1], Yulong Chen[1], Yufu Zhang[1], Haolong Li[1], Ruijie Peng[1], Zhen Kong[1], Liang Ran[2], Xinhe Feng[2], Zhichao Lyu[3], Liang Zhao[3], Jiamin Li[1], Yida Li[1], Feichi Zhou[1], Longyang Lin[1]
[1]Southern University of Science and Technology, [2]Beijing Pixelcore Technology, [3]Hefei Reliance Memory

This work presents a scalable Vision-Language-Action (VLA) edge processor for real-time robot control. A SimHash-based temporal token filtering eliminates redundant visual-token computation across action steps, while a hybrid RRAM-SRAM Processing-Near-Memory architecture with compression and dynamic orchestration achieves zero external memory access. A 16-chip module in 55 nm achieves 6.7-ms perception-to-action latency (150 Hz) with 12.74 mJ energy, delivering a 1.87–88.45× latency reduction over prior art and demonstrating end-to-end VLA inference for the first time.

Session C22: Fast and High-PSR On-Chip LDO Regulation

8:00 AM, Honolulu 2
Co-Chairs: Gael Pillonnet, CEA-Leti and Po-Hung Chen, National Yang Ming Chiao Tung University

8:00 AM
C22.1 A 1-to-1050-mA DLDO with Rising-Edge Computational Control and Load-Dependent Feedback Achieving 46-mV/ns DVS Rate and 0.15-ps FoM, Yichen Xu[1], Rentao Wan[1], Mao Li[1], Zhaoqing Wang[1], Suhwan Kim[2], Ram K. Krihshnamurthy[2], Xin Zhang[3], Mingoo Seok[1]
[1]Columbia University, [2]Intel, [3]IBM T. J. Watson Research Center

This brief presents a 1 mA–1050 mA computational DLDO for modern SoCs that achieves fast load transient, high efficiency, and fast DVS. A rising-edge computation enables accurate load prediction with only 1 nF Cout, while a load-dependent feedback maintains low ripple across a wide load range. Reusing the computation hardware for DVS improves the DVS rate by 4.6×. Fabricated in 28-nm CMOS, the DLDO achieves 0.15 ps FoM, 46.2 mV/ns DVS, <17 mV ripple, and 99.98% peak current efficiency.

8:25 AM
C22.2 A Scalable Distributed Digital LDO Achieving Fast and Accurate Current Balancing, Sang-Yun Nam[1], Won-Gyu Kim[1], Jaeseung Lee[2], Jun-Hyeok Yang[2], Sung-Wan Hong[3]
[1]Sogang University, [2]Samsung Electronics Co., Ltd., [3]Seoul National University

This paper presents a scalable distributed digital LDO (SD-DLDO) achieving fast and accurate current balancing for advanced SoCs. The proposed design utilizes a group-based balancing technique, reducing calibration steps to log4N, which is much faster than conventional approaches. In addition, a compact analog current sensor enables high-accuracy sensing. The prototype achieves a convergence time of 134.6 µs with a 5.2% error rate, proving its practicality for large-scale.

8:50 AM
C22.3 A Fully Integrated, GBW-Enhanced Tri-Loop, Maximum Ripple Rejection Tracking Analog LDO Achieving Better than -25.5dB PSR Up to 1GHz and 13ns Settling Time, Yunbeom Hwang[1], Yongkeon Kwon[1], Minkyu Song[1], Seongjun Shin[1], Junseo An[2], Jun-Eun Park[1]
[1]Department of Electrical and Computer Engineering, Sungkyunkwan University, [2]Department of Semiconductor and Display Engineering, Sungkyunkwan University

This paper presents a fully integrated, gain-bandwidth (GBW)-enhanced tri-loop low-dropout regulator (LDO) achieving wideband power-supply rejection (PSR) and fast transient response. The proposed LDO combines a gain-enhanced slow loop (SL), a current-mirror-based fast loop (CM-FL), and a regulated shunt-feedback-inverter-based ultrafast loop (RSFI-UFL) to simultaneously enhance PSR and transient performance. An on-chip maximum ripple rejection tracking (MRRT) controller adaptively aligns a notch-like PSR-boosting frequency to suppress the dominant switching-regulator ripple. Fabricated in a 40-nm CMOS process with an on-chip output capacitor of 40 pF, the prototype achieves PSR better than −25.5 dB up to 1 GHz, −52.8 dB at 10 MHz, and a 13-ns droop settling time with 50-mV undershoot for a 0.15-to-10-mA load step.

9:15 AM
C22.4 A Floating-Inverter Integrator LDO Achieving Near-Zero Line Regulation via Parasitic-Synthesized Supply-Tracking Offset, Jaehoon Lee[1,2], Yeji Min[2], Jae-Yoon Sim[2]
[1]Samsung Electronics Co., Ltd., [2]Pohang University of Science and Technology

This paper presents an analog LDO achieving near-zero line regulation (< 50 μV/V). The proposed architecture harnesses the intrinsic parasitic effect—traditionally considered a source of error—within a Floating-Inverter Integrator (FII). By operating the FII under an asymmetric loop configuration, this parasitic-induced charge is converted into a deterministic supply-tracking offset. This offset effectively synthesizes the reference modulation required to neutralize supply disturbances, bypassing the fundamental limitations of finite loop gain with negligible hardware overhead. Implemented in 40 nm CMOS, the LDO demonstrates a 50 μV/V line regulation and >70 dB PSR, confirming that the parasitic-converted offset enables robust regulation comparable to infinite-gain systems.

Session JFS3: DTCO

8:00 AM, Honolulu 1
Co-Chairs: Harsono Simka, Samsung and Rock-Hyun Baek, Pohang University of Science & Technology

8:00 AM
JFS3.1 5T+1 CFET SRAM with Dual-Bitline-per-Cell-Height Enabling Differential Readout and 12.5% Area Reduction, Dawit Burusie Abdi[1], Ankit Singh[1], Pieter Weckx[1], Arvind Sharma[1], Juergen Boemmels[1], Sheng Yang[1], Lynn Verschueren[1], Geert Hellings[1]
[1]imec

This work presents the 5T+1 CFET SRAM, a 5T-footprint bitcell exploiting vertical n/p stacking to integrate a transmission-gate access device without area penalty. A dual-bitline-per-cell-height architecture is proposed to enable differential sensing using the bitline of an unselected cell as a reference. The bitcell integrates a split-V_{DD} in-bitcell write-assist to improve writability without impacting bitcell footprint. DTCO evaluation shows 12.5% smaller bitcell than 6T CFET, 7–15% faster write, and 21% standby leakage reduction, while maintaining comparable RSNM and >150 mV five-percentile HSNM for unselected cells under split-V_{DD} operation.

8:25 AM
JFS3.2 Scaling High Density Designs Towards Angstrom Nodes: BEOL Routability, Pin Accessibility Challenges and Impact of Gear Ratio, Ji-Yung Lin[1], Yu-Hua Yang[2], Halil Kükner[1], Shen Yang[1], Lynn Verschueren[1], Jürgen Bömmels[1], Francesco Dell'Atti[1], Odysseas Zografos[1], Anita Farokhnejad[1], Geert Hellings[1]
[1]imec, [2]Novatek Microelectronics Corporation

A comprehensive study on the design scaling in angstrom nodes is performed on a high-density processor core. The scaling trends foresee the degradation of BEOL routability and pin accessibility over technology nodes. In results, designs in A7 node reach -6% utilization compared to N2 node, enlarging area by 8% from the expected level. Nevertheless, increasing M1 density with higher gear ratio (GR) is proposed to recover routability. GR 2:1 is shown to significantly increase utilization in A7 designs, reducing 14%/8% area vs. GR 1:1/3:2 while keeping performance & power impacts within 2%. Thus, GR 2:1 is required to achieve expected area scaling for high-density designs in angstrom nodes.

8:50 AM
JFS3.3 An Innovative DTCO for 2 nm GAA Technology with Hyper-cells in an Extremely High-Density Library for AI/HPC and Mobile SoC applications, Hakchul Jung[1], Jinyoung Lim[1], Ingyum Kim[1], Jisu Yu[1], Daeyeon Kim[1], SoonHo Choi[1], Bomin Joo[1], Jaehee Cho[1], Jiyun Han[1], Junil Choi[1], Jungho Do[1], Sangdo Park[1], Sanghoon Baek[1], Jongshin Shin[1]
[1]Foundry Business, Samsung Electronics Co., Ltd.

In this paper, Hyper-cells are introduced as key technology for high-performance and high-efficiency logic design in 2nm GAA (Gate-All-Around) technology. A Hyper-cell is enabled with wide nanosheets to maximize drive current and area efficiency in an extremely high-density architecture. With state-of-the-art DTCO (Design-Technology Co-Optimization), a block containing Hyper-cells achieves a significant speed improvement of up to 21 %, emerging as a key technology that enhances competitiveness in AI, HPC, and mobile-SoC applications.

9:15 AM
JFS3.4 ESD and Circuit Performance DTCO in 2nm CMOS Nanosheet Technology, Tao-Yi Hung[1], Li-Wei Chu[1], Wun-Jie Lin[1], Jam-Wem Lee[1], Kuo-Ji Chen[1]
[1]Taiwan Semiconductor Manufacturing Company (TSMC)

In this paper, low cap. electro-static discharge (ESD) design-technology co-optimization (DTCO) for high-speed input/output (I/O) circuit are proposed in a 2nm CMOS nanosheet technology. Several optimization methodologies are introduced and demonstrate to 9A charged-device model (CDM) robustness with only 100fF. Moreover, the new design can also suppress the V_{clamp} to <3V which will benefit interface circuit design flexibility

Session T10: Process Technology for CMOS Scaling

8:00 AM, Tapa 1
Co-Chairs: Naomi Yoshida, Applied Materials and Tetsu Ohtou, Tokyo Electron Limited

8:00 AM
T10.1 First Demonstration of Metal Interconnect At-Resolution Stitching with 0.55NA EUV Lithography for A10 Node and Beyond, Yannick Hermans[1], Yu Fang[1], Bart De Wachter[1], Vincent Wiaux[1], Tatiana Kovalevich[1], Stephane Lariviere[1], Andrea Mingardi[1], Ivan Ciofi[1], Victor Blanco[1], Sandip Halder[1]
[1]Angstrom Patterning Department (APD), imec

We report the first comprehensive assessment of at-resolution stitching for BEOL interconnects using 0.55NA EUV lithography, a key enabler for continued transistor scaling. With 0.55NA anamorphic optics limiting the maximum field size, at-resolution stitching is being considered to extend the field size for large chip applications. Based on experimental characterizations complemented by simulations, we demonstrate similar electrical performance and reliability for at-stitch and in-field interconnects down to metal pitch 24 nm, supporting future BEOL integration with 0.55 NA EUV lithography.

8:25 AM

T10.2 First Demonstration of a Functional Gate-All-Around Nanosheet CMOS Circuit Using Single Exposure High NA EUV BEOL Interconnects, Christopher Penny[1], Julien Frougier[1], Belle Antonovich[1], Christopher Carr[1], Victor Chan[1], Zheng Chen[1], Scott Halle[1], Teresa Howe[1], Atharv Jog[1], Jeong Kim[1], Nicholas Latham[1], Joe Lee[1], Shravan Matham[1], Abhinav Mamidala[1], Jaime Morillo[1], Matthew Wehrman[1], Indira Seshadri[1], Jed Rankin[1], Luciana Meli[1]

[1]IBM Research

We demonstrate BEOL patterning simplification using Single-Exposure (SE) High Numerical Aperture (NA) EUV with Metal-Oxide Resist (MOR), replacing Low NA EUV Self-Aligned Litho-Etch Litho-Etch (SALELE) multi-patterning while achieving superior Health of Line electrical yield. Gate-All-Around (GAA) nanosheet Ring Oscillator (RO) CMOS circuit functionality is demonstrated using SE High NA EUV for BEOL interconnect. Compared with the control SALELE wafers, the High NA EUV wafers were able to demonstrate comparable yield and AC performance in RO circuits.

8:50 AM

T10.3 High Yield Sub 20 nm Ru Direct Metal Etch Enabled by Single-Exposure 0.55NA EUV and dry MOR Technology, Alejandro Berdonces Layunta[1], Martin O'Toole[2,2], Etienne De Poortere[2], Tsann-Bim Chiou[2], Cyrus Tabery[2], Stefan Decoster[1], Victor M. Blanco Carballo[1], Bart De Wachter[1], Shubhankar Das[1], Elisabeth Camerotto[3], Zhengtao Chen[3], David Hellin[3], Anuja De Silva[3], Gosia Jurczak[3]

[1]imec, [2]TDC, ASML, [3]LAM Research

To enable continued BEOL scaling below 20 nm, Ru Direct Metal Etch (DME) integration has been proposed as an alternative to Cu damascene due to its lower resistivity. However, at these extremely aggressive pitches, the demands on lithography and patterning yield become increasingly critical. In this work, we demonstrate a simplified single-exposure 0.55NA EUV dark-field patterning approach using metal-oxide dry photo resist (MOR) and dry development for Ru DME. Compared to multi-patterning schemes, the proposed flow eliminates pitch walk and reduces hard-mask complexity. Electrical yield is evaluated using large-area serpentine and fork structures for sensitive detection of line breaks and bridges. Optimized conditions result in dies having no defects associated with line breaks or bridges for 20 nm pitch structures with line lengths exceeding 1m, while preliminary data indicates feasibility at 18 nm pitch. These results establish a viable path toward high-volume manufacturing of sub-20 nm Ru DME interconnects.

9:15 AM

T10.4 Direct Synthesis of Multi-stacked Channels and Doped Source/Drains for TMD CFETs, Shinichi Tanabe[1], Naoya Okada[2], Toshifumi Irisawa[2], Hao Cheng[1], Hitoshi Miura[1], Hisashi Warashina[1], Yu-min Huang[1], Keiichi Nagasaka[1], Atsuki Fukazawa[1], Hiroki Maehara[1]

[1]Tokyo Electron Ltd. (TEL), [2]National Institute of AIST

A simple LSI-compatible technique for synthesizing multi-stacked MoS_2 channels (up to 10 stacks) directly on Si substrates is demonstrated by utilizing post annealing of MoO_x. Thickness-controlled high-quality bilayer MoS_2 films with mobility reaching 12 cm^2/Vs on a wafer scale is achieved, which is the highest mobility reported for MoS_2 directly synthesized on SiO_2 by post annealing of MoO_x. In addition, multi-stacked WSe_2 channels are also demonstrated through annealing of WO_x in a selenium environment. This approach also enables formation of doped thick-layered MoS_2 source/drains (S/Ds) by incorporating metals into the MoO_x prior to the post annealing. The capability to fabricate both MoS_2 and WSe_2 multi-stacked channels along with gate-all-around (GAA) compatible S/Ds offers promising prospects for the integration of transition-metal dichalcogenide (TMD) complementary field-effect transistors (CFETs).

Session T11: FeRAM Array and Module

8:00 AM, Honolulu 3

Co-Chairs: Karl Hofmann, Infineon and Hang-Ting Lue, Macronix International Co., Ltd.

8:00 AM

T11.1 3D Vertical FeFET Array with Record Endurance (>10^{12}), Fast Writing (±2V, 20 ns), Disturb Immunity, and Kb-scale Verification for High Density 1T RAM, Yuejia Zhou[1], Yuancheng Yang[2], Liang Chen[1], Mingxiang Yu[1], Wenpu Luo[1], Zhiyuan Ning[1], Weiqin Huang[1], Xiaojian Xu[1], Jing Guo[2], Runteng Zhu[1], Hanyong Shao[1], Zhiliang Xia[2], Zongliang Huo[2], Kechao Tang[1,3], Ru Huang[1,3]

[1]School of Integrated Circuits, Peking University, [2]Yangtze Memory Technologies Holdings Co., Ltd., [3]Beijing Advanced Innovation Center for Integrated Circuits

We report a comprehensively optimized 3D vertical ferroelectric FET (FeFET) memory, with highlights below: 1) Record-level metrics of the 3D cells, including endurance >10^{12}, fast write at ±2 V/20 ns, and minimal read-after-write delay < 100 ns; 2) First demonstration of 3D array reliability, showing disturb immunity with aggressively scaled 40-nm vertical pitch, and error-free Kb-scale readout; 3) Systematic process optimization with in-depth nano-analysis, revealing mechanisms related to FE grain and channel oxygen engineering. The proposed 3D FeFETs enable high density 1T RAM, offering a low-cost and high-speed solution for the expanding data in AI era.

8:25 AM

T11.2 First Demonstration of Bit-Cost Scalable and BEOL-Compatible Monolithic 4-Tier FeRAM Operating at 125°C , Eknath Sarkar[1], Jay V Sonawane[1], Hyeonwoo Park[1], Emmanuel Quezada[1], Kiseok Lee[1], Dyutimoy Chakraborty[1], Jaewon Shin[1], Faaiq G Waqar[1], Chengyang Zhang[1], Hyun Jae Lee[1], Mengkun Tian[1], Susan Trolier-McKinstry[2], Vijaykrishnan Narayanan[3], Asif I Khan[1], Shimeng Yu[1], Suman Datta[1]

[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, [2]Department of Materials Science and Engineering, Pennsylvania State University, [3]Department of Computer Science and Engineering, Pennsylvania State University

We report the first experimental demonstration of a four-tier, bit-cost scalable (BiCS), back-end-of-line (BEOL) compatible monolithic 3D 1T1C Ferroelectric Random Access Memory (FeRAM), integrating 5nm thin HZO ferroelectric capacitors with dual-gate Indium Gallium Oxide (IGO) access transistors. All tiers show uniform switching characteristics with $2P_r \approx 60–70\ \mu C/cm^2$ and $2V_C \approx 1.6–2.1$ V. Refresh-free operation at 125°C makes the technology ideal for monolithic stacking directly atop high-performance, power-intensive compute dies. We demonstrate successful read with write-back operations, 10-year data retention at 125°C. Benchmarking projections indicate that this BiCS-based approach enables aggressive density scaling, potentially reaching 4.36 Gb/mm^2 for a 100-tier stack.

8:50 AM

T11.3 High-Performance Normally-off IGZO FeFETs with Thermally Stable Ru S/D for Power-Efficient Memory Arrays, Seung Dam Hyun[1], Jee-Eun Yang[1], Ha-Jun Sung[1], Hojung Kim[1], Chang Seung Lee[1], Kyooho Jung[2], Jung-Kyun Kim[1], Youngkwan Cha[1], Ki Deok Bae[2], Garam Park[1], Kwang-Hee Lee[1], Moonil Jung[1], Daehyuk Son[1], Younjin Jang[1], Na-Rae Han[1], Hyerim Hong[1], Seung-Geol Nam[2], Duk-Hyun Choe[2], Sangwook Kim[1], Jinseong Heo[1]

[1]Future Tech Platform, Samsung Advanced Institute of Technology, [2]Advanced Device Platform, Samsung Advanced Institute of Technology

We propose a Ru source/drain (S/D) engineering strategy to achieve normally-off IGZO-based ferroelectric FETs (FeFETs) with a positive threshold voltage (Vth). Owing to the high oxide formation energy of Ru, oxygen scavenging from the IGZO channel is suppressed, enabling stable Vth control under high-temperature annealing. Ru S/D FeFETs demonstrate excellent endurance up to 10^9 program/erase (PGM/ERS) cycles with maintaining positive Vth. The resulting normally-off FeFETs enable energy-efficient read operation in 3D arrays, reducing read energy by 60% compared to conventional S/D devices.

9:15 AM

T11.4 Interfacial engineering of scaled HZO in 3D-trench BEOL-compatible FeCAPs for high endurance (≥10^{13} cy.) and $2P_R$ (>40 $\mu C/cm^2$) at low (≥1.3 V) operating bias, Jasper Bizindavyi[1], Mihaela Ioana Popovici[1], Eyup Can Demir[1], Waleed Maqsood[1], Shruthi Subramanian[1], Gwon Kim[1,2], Alexandru Pavel[1], Nicolò Ronchi[1], Attilio

Belmonte[1], Jan Van Houdt[1,3]
[1]imec, [2]Sungkyunkwan University, [3]KU Leuven

Scaling the operating bias (V_{op}) of ferroelectric capacitors (FeCAPs) without degrading the remnant polarization $2P_R$ is a critical challenge, as it is increasingly difficult to crystallize and stabilize the orthorhombic (o) phase when scaling hafnium zirconium oxide (HZO). In this work, we demonstrate 300mm 3D-trench BEOL-compatible FeCAPs with excellent $2P_R$ (>40 μC/cm^2) and endurance (≥1e13 cy.) at low V_{OP} (1.3 V) in scaled (5 nm) HZO by seed interfacial engineering. We show that this is achieved through an effective reduction of the crystallite nucleation barriers during HZO deposition on the seed layer, enabling efficient crystallization and transition of non-FE phases into the FE o-phase.

Technical Sessions Block 8 (9:55 AM – 12:00 PM)

Session C23: ASICs for Emerging Workloads

9:55 AM, Tapa 3
Co-Chairs: Thierry Tambe, Stanford University and Chia-Hsiang Yang, National Taiwan University

9:55 AM
C23.1 A 90.6 pJ / pixel 1280x800 65 fps cm-Scale Fully Integrated Video Encoder System Utilizing a 2x2 Microlens Array, Heejin Yang[1], Pierre Abillama[1], Xiangdong Wei[1], Mohammed Ashfakh Ali Huluvallay[1], Qirui Zhang[1], Junhao Chen[1], Jayeon Yi[1], Chieh-Shen Chen[1], Andrea Bejarano-Carbo[1], Michael Alexander Stankovich[1], Martin Lentz[2], Yves Bellouard[2], Patrick Stockton[3], James Friel[3], Andrew Trickey-Glassman[3], David Blaauw[1], Dennis Sylvester[1], Hun-Seok Kim[1]
[1]University of Michigan, [2]EPFL, [3]Areté

This work presents a compact (31.5×10.7×4 mm^3) video encoder system with a commercial imager, flash, a custom 2×2 microlens array, and an image signal processing (ISP) chip using novel 2×2 intra-frame compression. Leveraging the subframe redundancies inherent to the microlens array, the system achieves > 59× compression with ≥ 32 dB PSNR. With this, the ISP chip fabricated in 28nm CMOS attains 90.6 pJ/pixel energy efficiency at 1280×800, 65fps video stream.

10:20 AM
C23.2 Specter: 376 fps 2.67 nJ/pixel 4DGS Processor for Spatiotemporal Reconstruction with Contribution-Ordered Rendering and Deformation-Aware Modeling, Jae-Young Kim[1], Teokkyu Suh[1], Junsoo Lim[1], Minjun Park[1], Hyunmin Choi[1], Jaeha Min[1], Donghyuk Kim[1], Joo-Young Kim[1]
[1]Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST)

We present Specter, a 4D Gaussian Splatting (4DGS) processor that enables real-time spatiotemporal reconstruction (STR) with efficient end-to-end on-chip rendering and modeling. Specter addresses key inefficiencies-redundant external memory access (EMA) from over-extended Gaussian bounds during tile allocation, wasted computation from coarse tile-based rendering, and high dynamic scene modeling overhead-with three solutions: (i) opacity- and shape-aware bound tightening to reduce redundant Gaussian-to-tile allocation and EMA, (ii) contribution-ordered rendering with hierarchical skipping to minimize non-contributing pixel computations, and (iii) hash-based deformation detection to skip modeling of non-deformed tiles. Fabricated in 28nm CMOS, Specter operates at 50-560 MHz with 0.75-1.1 V supply and occupies 12.96 mm^2 die area. Specter achieves up to 376 fps at 2.67 nJ/pixel for rendering and 112.3 iterations/s for modeling, respectively, while supporting 1352×1014 resolution (2.14× more pixels than prior works).

10:45 AM
C23.3 HCNP: A 70.2 TOPS/W Hybrid CIM-NPU Processor with In-Streaming Processing for Energy Efficient CNN/Transformer Acceleration in HMD, Seungbin Kim[1], Hoichang Jeong[2], Dongwook Kim[1], Jiwon Kim[1], Kyuho Jason Lee[2]
[1]Ulsan National Institute of Science and Technology (UNIST), [2]Yonsei University

This paper presents HCNP, an energy-efficient hybrid CIM-NPU processor designed to accelerate diverse DNN topologies under the strict power constraints of head-mounted display platforms. Unifying CIM and NPU architectures through *in-streaming processing* enables layer fusion, reducing off-chip access by 79.7%. The NPU core co-designed with *runtime KV cache eviction* reduces the memory footprint by 79.8% and eliminates 32.1% of computation. Optimizing *flexible-route network-on-chip* and *single-column-multiple-output* architectures facilitates a hierarchical inter-/intra-core dataflow strategy, maximizing CIM utilization to 81.8%. Fabricated in a 28nm CMOS process, HCNP achieves 70.2 TOPS/W, reducing energy consumption by 32.3% and 93.6% compared to state-of-the-art CIM for mobile-targeted CNNs and state-of-the-art NPU for large language models, respectively.

11:10 AM
C23.4 A 248.5 GFLOPS/mm^2, 1.67 TFLOPS/W Streaming Processor with Hardware-Level Scheduling for Advanced Spectrum Sensing, Hyunwon Chung[1], Parin Senta[1], Jason Yu[1], Yukun Fang[1], Pierre Abillama[1], Kuan-Yu Chen[1], Ronald Dreslinski[2], David Blaauw[1], Hun-Seok Kim[1]
[1]Electrical and Computer Engineering (ECE), University of Michigan, Ann Arbor, [2]Computer Science and Engineering (CSE), University of Michigan, Ann Arbor

This work presents a fast-programmable streaming processor for spectrum sensing in dynamic RF environments. To respond to rapidly changing RF channels, the 2D array of 77 processing elements is tightly integrated with a light-weight kernel launcher that achieves 10 ns array reconfiguration without host-CPU intervention. The processor is able to analyze multiple heterogeneous signals/channels in highly dynamic conditions and achieves a compute density of 248.5 GFLOPS/mm^2 and an energy efficiency of 1.67 TFLOPS/W.

Session C24: Circuits for Bio-Signal Acquisition and Neural Interfaces

9:55 AM, Honolulu 2
Co-Chairs: Inhee Lee, University of Pittsburgh and Hyung-Min Lee, Korea University

9:55 AM
C24.1 A Sub-μW VCO-Based Direct Digitization Front-End with an Auto-Ranging Technique Achieving 1.8 NEF and 182.3 FoM$_{DR}$ for Energy-Efficient Neural Recording, Hong Liao[1], Xiaoxu Yang[1], Minghao Chen[1], Yuqian Li[1], Lei Qiu[1], Patrick Mercier[2], Miao Meng[1]
[1]Tongji University, [2]University of California, San Diego

This work presents a sub-μW VCO-based direct-digitization neural recording front-end with an auto-ranging technique, featuring ultra-low power consumption, compact chip area, wide dynamic range (DR), and low input-referred noise (IRN). This front-end achieves high-fidelity neural signal acquisition by employing a low-noise VCO-based quantization core to form a high-resolution quantization window while large-amplitude, linearity-insensitive artifacts can be recorded by seamlessly and dynamically shifting the quantization window via the proposed auto-ranging quantizer. Fabricated in a 65 nm CMOS process with 0.035 mm^2 active area. The proposed IC achieves 82.2 dB DR and 42 nV/$\sqrt{}$Hz IRN across a 10 kHz signal bandwidth (BW), while consuming 0.97 μW from a 0.8 V supply, resulting in a 1.8 NEF and 182.3 FoM$_{DR}$.

10:20 AM
C24.2 A 256-Channel, 32-Interconnect, 1.25-NEF, In vitro and In vivo Neural Interface with Auto-Calibrated Input Impedance Boosting and On-Chip Spike Detection for Flexible Capacitive Microelectrode Array, Yiyang Chen[1], Suhyeon Nam[2], Sun-woong Byun[2], Pietro Velcich[1], Shinil Cho[2], Chanwoo Lee[2], Ju Young Lee[2], Sean

Weaver[3], Enrico Brugnolotto[3], Eunkyu Hwang[2], Dongseok seo[2], Eunee Lee[2], Ki Jun Yu[2], Woojun Choi[2], Taekwang Jang[1]
[1]ETH Zürich, [2]Yonsei University, [3]Catalan Institute of Nanoscience and Nanotechnology

This work presents a capacitive neural interface for 256-channel simultaneous recording with 32 interconnects. Fabricated in a 22-nm CMOS, the prototype is evaluated using two flexible capacitive microelectrode arrays (MEAs) for *in vitro* and *in vivo* applications. To the authors' best knowledge, this is the first to demonstrate *in vivo* or *in vitro* measurement of the series-parallel amplifiers (SPA). The proposed system achieves a noise efficiency factor (NEF) of 1.25 and 0.81, measured with and without the ADC, respectively, which are the lowest among the AFEs reporting *in vivo* or *in vitro* results.

10:45 AM
C24.3 A 1V Time-Division-Multiplexed 4-Channel 3rd-Order NS-SAR ADC with 178.8dB FoM$_{DR}$ and 5.8GΩ Input Impedance for Wearable Dry-Contact Biopotentials, Geunha Kim[1], Sehwan Lee[2], Taeryoung Seol[3], Kyuhyeon Park[1], Jeongyoon Wie[1], Minoo Lee[1], Seokhan Jeong[1], Yeonjae Shin[1], Sangkyu Lee[2], Byeongwoo Koo[2], Michael Choi[2], Junghyup Lee[1]
[1]EECS, Daegu Gyeongbuk Institute of Science and Technology, [2]Samsung Electronics Co., Ltd., [3]Georgia Institute of Technology

This paper presents a time-division-multiplexed 4-channel 3rd-order NS-SAR ADC achieving 178.8dB FoM$_{DR}$, 5.8GΩ input impedance, and a 1.72V$_{PP}$ linear input range for multi-channel dry- and non-contact biopotential acquisition in wearable applications. Implemented in a 0.18-µm CMOS process, the ADC consumes 4.17µW/ch from a 1V supply.

11:10 AM
C24.4 A 4.32uW/Ch 256-Channel Neural Probe Integrating in-situ ADC and Stimulation for Closed-Loop DBS, Shihui Sun[1], Yi Zhuo[1], Weixiong Qiu[1], Yunxu Zhao[1], Tianyu Yang[1], Zhaofeng Huang[2], Linxiao Shen[1], Yacong Zhang[1], Zhongjian Chen[1], Zhihong Li[1], Wengao Lu[1]
[1]School of Integrated Circuits, Peking University, [2]Lingfeng Vision Chip Beijing Technology Co.

This work presents a 256-channel neural probe for closed-loop DBS with per-channel low-power in-situ dual-loop-oscillator ADCs. Two operating modes are supported: neural recording with baseline stabilization and stimulation with synchronous monitoring. The 0.0010 mm^2/channel probe consumes 4.32 µW/channel and achieves 3.68 µVrms input-referred noise over a 10 kHz bandwidth. The area-energy efficiency FoM (0.21fJ·mm^2/step) and NEF (2.69) represent state-of-the-art performance.

11:35 AM
C24.5 A dynamic supply temporal interference stimulation ASIC achieving 70% peak power saving and 21V compliance with accurate charge balancing, Haoming Xin[1], Roland van Wegberg[1], Stefano Stanzione[1], Vojkan Mihajlovic[1], Carolina Mora Lopez[2], Mario Konijnenburg[1], Nick Van Helleputte[2]
[1]imec-Netherlands, [2]imec

A Temporal Interference Stimulation (TIS) ASIC featuring a novel dynamic supply (DS) and common-mode voltage generation technique is proposed. Charge balancing (CB) strategy that works seamlessly with the proposed DS generation technique is also integrated for safety. Fabricated in 130nm BCD technology, the proposed ASIC uniquely offers HV compliance, high power saving and accurate CB. The ASIC is fully validated both electrically and in-saline.

Session JFS4: More-than-Moore Designs & Robotics

9:55 AM, Tapa 2
Co-Chairs: Alvin Loke, Intel Corporation and Koji Nii, TSMC Design Technology Japan, Inc.

9:55 AM

JFS4.1 Architecture Partitioning and Optimization in the Era of Chiplet Computing, Vamsi Krishna Alla[1], Alan Smith[2], Tyrone Huang[2], Michael Steffen[2], Alexander Kaganov[2], Wonjun Jung[2]
[1]AMD, [2]Advanced Micro Devices

Traditional transistor scaling is slowing, while AI/HPC workloads demand rapidly increasing compute throughput and memory bandwidth. This gap shifts the scaling emphasis to system-level innovation through chiplet partitioning and advanced packaging. We present a taxonomy of 2.5D, 3D, and hybrid integration, and discuss chiplet design considerations including architectural partitioning, die-to-die PHY selection, power delivery, thermal management, and cross-die verification. The AMD Instinct™ MI355X GPU serves as a case study, employing eight 3D hybrid-bonded TSMC N3P accelerator complex dies on two merged TSMC N6 I/O dies with eight 12-high HBM3E stacks via a 2.5D silicon interposer. The MI355X delivers 1.9Å~ peak compute and 1.5Å~ HBM bandwidth and capacity compared to the MI300X within the same form factor.

10:20 AM

JFS4.2 3DIC Design Methodology for High-Performance Computing, Chih-Wei Jim Chang[1], Zhe-Wei Jiang[1], King-Ho Tam[1,1], Yun-Han Lee[1], Lee-Chung Lu[1]
[1]Design Technology Platform, Taiwan Semiconductor Manufacturing Company (TSMC)

The escalating requirements of contemporary workloads, such as high-performance computing, artificial intelligence, data center operations, and cloud computing have necessitated an IC design paradigm shift from the SoC architecture to the more sophisticated multi-die 3D integrated system architecture. This work examines the various design methodology advancements critical for the successful implementation of the state-of-the- art 3DIC integration.

10:45 AM

JFS4.3 Integrated Imaging Solutions for Robotic Navigation and Manipulation: Visual-Inertial SLAM and Neuromorphic Tactile Sensing, Ryoji Eki[1], Tkaaki Kato[1], Valentina Cavinato[2], Agis Politis[2], Stefan Isler[2]
[1]Sony Semiconductor Solutions Corporation, [2]Sony Advanced Visual Sensing AG

This paper presents imaging solutions for robotic Navigation and Manipulation. For Navigation, we demonstrate enhanced Simultaneous Localization and Mapping (SLAM) technology achieving centimeter-level accuracy for precise localization. For Manipulation, we introduce an Event-based Vision Sensor(EVS) based tactile sensor utilizing neuromorphic vision for high-precision marker tracking with low latency and power consumption. We propose a 3D-stacked CMOS sensor integrating EVS with on-chip AI processing, achieving reduced data volume and compact size for robotic applications.

11:10 AM

JFS4.4 Enabling the Robotic Revolution: Bridging Performance Gap between Present and Future, Qijing Huang[1], Wenqi Jiang[2], Christos Kozyrakis[2,3], Jason Clemons[2]
[1]NVIDIA, [2]NVIDIA, Santa Clara, CA 95051, USA, [3]Stanford University

The robotics industry has reached an inflection point due to the recent advancements of robotics foundation models which enable general-purpose action generation in various physical environments. However, a significant performance gap remains between these models and the constraints of edge hardware. This paper characterizes the foundation models for robotics and analyzes the hardware bottlenecks across three critical dimensions: model architecture diversity, real-time latency constraints, power and thermal limits. By analyzing state-of-the-art models like Physical Intelligence's π0, NVIDIA GR00T N1.6, and NVIDIA DreamZero on modern platforms, we demonstrate that achieving the desired frequency for smooth and responsive humanoid dexterity under stringent conditions necessitates an 8.3× increase in compute throughput, a 19.7× leap in memory bandwidth and a shift toward 4×

higher power efficiency for untethered operations. We conclude by calling on the VLSI community to help provide the innovations needed to move from research prototypes to ubiquitous, reliable humanoid deployment.

11:35 AM
JFS4.5 Device-to-Architecture Co-design of Monolithic 3D FPGA using 2D transistor, Hyeong-Seok Jang[1], Munhyeon Kim[2], Luhing Hu[1], Jaewoo Shim[1], Joonyun Kim[1], Minsu Seol[1], Sang Won Kim[1]
[1]Samsung Advanced Institute of Technology, [2]Department of Electrical and Information Engineering, Seoul National University of Science and Technology

This work presents a measurement-driven monolithic 3D (M3D) field-programmable gate arrays (FPGAs) enabled by vertically integrated SRAM using two-dimensional (2D) semiconductor devices. We propose a new FPGA architecture in which the block RAM (BRAM) based on transition metal dichalcogenide (TMD)-channel transistor is monolithically stacked directly above Si CMOS logic blocks. The feasibility of the approach is experimentally validated through integration of logic gates by using MoS_2 NMOS and WSe_2 PMOS transistors and translated into SPICE models for both circuit- and system-level evaluation. Based on the measured models, a stacked TMD-channel-based BRAM macro is designed for the first time and validated in terms of stability and functionality. When evaluated using Verilog-to-Routing (VTR)-based benchmarks, the proposed approach achieves 59% area reduction and 13% critical path delay improvement over a conventional FPGA. This work demonstrates that M3D memory integration can provide tangible system-level benefits for next-generation FPGA architectures.

Session T12: Memory (NAND/NOR/DRAM)

9:55 AM, Tapa 1
Co-Chairs: Jixin Yu, Sandisk and Vita Pi-Ho Hu, National Taiwan University

9:55 AM
T12.1 Deposition Process-Based Charge Trap Layer Segmentation for 3D Flash Memory Achieving Improved Cell Reliability While Maintaining Program/Erase Characteristics, Fumie Kikushima[1], Yosuke Murakami[1], Hideomi Aoike[1], Naoya Yoshitaka[1], Kazuhiro Yamauchi[1], Yuma Hizume[1], Katsunori Miyachi[1], Tatsuo Ogura[1], Atsushi Fukumoto[1], Junichi Kaneyama[1], Yusuke Sasaki[1], Takumi Moriyama[2], Naohiro Hosoda[2], Shunsuke Takuma[2], Masanori Tsutsumi[2], Tatsunori Isogai[1], Hiroshi Takeda[1]
[1]Advanced Memory Development Center, Kioxia Corporation, [2]Technology Development Engineering, Sandisk Corporation

In this study, we propose a 3D flash memory incorporating a pseudo-segmented charge trap layer (CTL) structure formed by a deposition process-based CTL segmentation scheme. By keeping part of the CTL continuous, program slope degradation is suppressed, and the simulated CTL thickness margin—which maintains the program slope and improves cell-to-cell interference (C2C)—is larger than that of fully segmented CTL cells. The demonstrated pseudo-segmented CTL cells achieve a 29% reduction in C2C and a 44% reduction in lateral charge loss (LCL). The word line (WL) pitch scalability for unstressed cells is estimated to be 8 nm for triple-level cells and 10 nm for quadruple-level cells.

10:20 AM
T12.2 Innovative performance of 6th generation 11nm DRAM cell transistor for HBM4 application, Shinwoo Jeong[1], Jiuk Jang[1], Chulmin Kim[1], Yuseok Choi[1], Youngwoo Son[1], Dongsoo Woo[1], Hoin Ryu[1], Kyo-Suk Chae[1], Byung-hyun Lee[1], Sangjoon Hwang[2]
[1]DRAM PA, Samsung Electronics Co., Ltd., [2]Samsung Electronics Co., Ltd.

The performance of 11nm—6th generation in 1x-nm-class—dynamic random access memory (DRAM) cell transistor is demonstrated in High Band Width Memory (HBM) products and its ultra-scaled device technology and electrical superiority are proved. The evaluated performance of DRAM retention characteristics shows that the suppressed leakage current under Sub-45Å Gate Oxide is attributed to the fully exploited dual work function gate

node of Buried Channel Array Transistor (BCAT) on the <100> flat-zone oriented silicon wafer (FZ<100>). In particular, despite of the dramatically decreased capacitance of cell storage module (Cs), the final output of the cell transistor under operating and stand-by modes exhibits novel and innovative performance characteristics. Furthermore, we have achieved the overall outstanding cell refresh interval time (tREFi) performance at high temperature of 120℃ during Chip on Wafer (CoW) test for 12-High stacked HBM devices.

10:45 AM
T12.3 Uniform and Direction-Controllable Metal-Induced Lateral Crystallization for 3D NAND Flash Memory via $NiSi_2$(111)-Oriented Seed Layer Transfer, Ohhyuk Kwon[1], Junyoung Lee[2], Jaemin Jung[3], Siyoung Yang[3], Kwangmin Park[3], HanMei Choi[4], Sejun Park[2], Hyunsang Hwang[1]
[1]Department of Materials Science and Engineering(MSE), Pohang University of Science and Technology, [2]Flash Process Architecture Team, Memory Division, Samsung Electronics Co., Ltd., [3]Advanced Process Development Team, Semi conductor R&D Center, Samsung Electronics Co., Ltd., [4]Flash Process Development Team, Semiconductor R&D Center, Samsung Electronics Co., Ltd.

This paper reports a growth-direction-controllable Metal-Induced Lateral Crystallization (MILC) technique to overcome the self-blocking issue caused by the stochastic collision of randomly oriented whiskers in high-aspect-ratio channels. Utilizing a seed layer transfer process, we effectively suppressed random nucleation and successfully guided MILC whiskers along a predefined orientation. This method guarantees stable and continuous crystallization by minimizing whisker collisions. We validated the proposed technology in a 155-layer 3D NAND structure, demonstrating significant improvements in crystallinity. This work highlights the critical role of whisker directionality and provides essential guidelines for implementing MILC in future 3D NAND architectures.

11:10 AM
T12.4 A 4Gb 32-Layer 3D NOR Flash Test Chip with Innovative Design Methods Enabling Sub-100ns Low-Latency Read and Robust Retention, Hang-Ting Lue[1], Teng-Hao Yeh[1], Chih-Chang Hsieh[1], Tzu-Hsuan Hsu[1], ChihWei Hu[1], Chia-Jung Chiu[1], Guan-Ru Lee[1], Ken-Hui Chen[2], Chin-Hung Chang[2], Keh-Chung Wang[1], Chih-Yuan Lu[1]
[1]Emerging Central Lab., Macronix International Co., Ltd., [2]Product Design, Macronix International Co., Ltd.

We develop a 4Gb 32-layer 3D NOR Flash test chip to achieve low latency read (<100ns) and good high-temperature retention after P/E cycling stress. To address the fundamental issues of large parasitic capacitance loading accompanied with high-layer stacked 3D memory that impacts read latency, and the retention challenges of charge-trapping memories to fulfill the automotive-grade, we have several innovative design methods. The large bitline capacitance loading in 3D array causes spurious transient displacement current, and we have invented an ideal capacitive matching design to cancel the displacement current to meet low-latency read. The charge-trapping memories often generate shallow-trapped electrons especially after P/E cycling stress. We found that the “recovery programming” built in the erase algorithm provides an effective shallow-trap reduction mechanism, providing much better data retention. TThe future prospect of 3D NOR is potential for the high-density and high-speed non-volatile memory (NVM) chiplets to support high-performance on-device AI computing.

Session T13: Device Physics and Reliability

9:55 AM, Honolulu 1
Co-Chairs: Suraj Cheema, MIT and Ryuta Tsuchiya, Hitachi, Ltd.

9:55 AM
T13.1 Workload-Derived AC Bias Temperature Instability and Mechanism Contributions in BEOL Oxide Channel DRAM Access Transistors, Hyun Jae Lee[1], Faaiq Waqar[1], Hyeonwoo Park[1], Chengyang Zhang[1], Jaewon Shin[1], Eknath Sarkar[1], Hwan Kim[2], Changik Im[2], Min Ji Hong[2], Daewon Ha[2], Asif Khan[1], Shimeng Yu[1], Suman Datta[1]
[1]Electrical and Computer Engineering, Georgia Institute of Technology, [2]Semiconductor Research and Development, Samsung Electronics Co., Ltd.

This work presents a workload-driven bias temperature instability (BTI) modeling framework that translates execution-based DRAM row-access traces (derived from Ramulator) into time-resolved AC-BTI waveforms for back-end-of-line (BEOL) oxide channel FETs. By mapping AC-to-DC BTI degradation within a single time window, threshold voltage (V_{th}) evolution is tracked without hidden recovery. This enables physics-based identification of BTI mechanisms based on their time-exponents. Our framework resolves workload-dependent relative contributions of competing mechanisms, enabling reliable BTI projection and workload-aware controller design guidance.

10:20 AM

T13.2 Elucidating the Role of Channel Materials and Interlayers in Ferroelectric NAND, Jung-Kyun Kim[1], Sijung Yoo[1], Sanghyun Jo[1], Seung Dam Hyun[1], Yoonsang Park[1], Taek Jung Kim[1], Kihong Kim[1], Kyung Mee Song[1], Donghoon Kim[1], Kyooho Jung[1], Jee-Eun Yang[1], Younjin Jang[1], Na-Rae Han[1], Byeong Gyu Chae[1], Jung Yeon Won[1], Hyangsook Lee[1], Jinseong Heo[1], Yongsung Kim[1], Sangwook Kim[1], Duk-Hyun Choe[1]
[1]Samsung Advanced Institute of Technology

Oxide-semiconductor (OS) channels are being actively explored for ferroelectric NAND (FeNAND), but their impact relative to Si channels remains poorly understood under fully identical device conditions. Here, we systematically compare Si- and OS-channel FeNAND devices, explicitly isolating the channel-induced effects within an identical top-gated structure and process. We find that OS channels exhibit a strongly asymmetric switching behavior, leading to a reduction of the memory window by more than half and its complete confinement to the negative Vth regime. We further show that introducing high-k channel interlayer (Ch.IL) effectively recovers the degraded memory window. Moreover, the Ch.IL is found to be essential for ensuring stable operation in OS-channel FeNAND. These results highlight critical design considerations for channel and interlayer engineering in FeNAND devices.

10:45 AM

T13.3 New Insights into Multi-Level Cell Behavior in FeFET: Localization Effect Aware Write-Verify Scheme, Omkar Phadke[1], Halid Mulaosmanovic[2], Stefan Duenkel[2], Sven Beyer[2], Janak Sharda[1], Shimeng Yu[1]
[1]Electrical and Computer Engineering, Georgia Institute of Technology, [2]Globalfoundries

Localization of random ferroelectric (FE) domain distribution and interface trap density (D_{IT}) in the ferroelectric field effect transistor (FeFET) is identified as a new challenge for enabling robust Multi-Level Cell (MLC) operation. FE localization leads to dynamic dependence of V_{TH} on the V_{DS}, leading to read-out errors using the classical notion of utilizing V_{TH} distribution as the criteria for MLC. This work presents a new MLC scheme that utilizes the dynamic drive strength of FeFET, featuring a characteristic time constant τ to discharge the bitline as a more appropriate MLC criteria. Key contributions include: 1) Experimental statistical characterization of scaled FeFET on 28nm Foundry platform for classical and new MLC schemes; 2) First report of D_{IT} localization that exhibits a “crossover” effect in I_D-V_G characteristics; 3) Modelling of FE and D_{IT} localization; 4) Validation of the new MLC scheme showing minimized errors across a dynamically varying bitline voltage swing.

11:10 AM

T13.4 First Demonstration of CVD Monolayer WSe_2 PFETs with Ultra-Short Channel Length of 16 nm: Achieving Record I_{on} of 1532 μA/μm at V_{ds} = -1 V and Ballistic Efficiency of 81% , Lei Sun[1], Tingting Gao[2], Lin Xu[3], Kai Wang[1], Kaustav Banerjee[3], Yang Chai[2], Lain-Jong Li[4], Xuefei Li[1]
[1]Huazhong Univeristy of Science and Technology, [2]The Hong Kong Polytechnic University, [3]University of California, Santa Barbara, [4]National University of Singapore

In this work, we report an oxygen p-doping and full contact structure to reduce the contact resistance (R_c) of monolayer p-type tungsten diselenide (WSe_2) and demonstrate high-performance p-type WSe_2 field-effect transistors (FETs) with a scaled channel length (L_{ch}) of 16 nm. The controllable oxygen doping and the full contact structure greatly increase the contact quality. The co-optimized R_c is as low as 352 Ω·μm at 4.2 K, which is ~two times lower than that of the control sample with top contact. Consequently, 16-nm-channel WSe_2 pFETs deliver a

record saturation current of 1532 μA/μm at a drain voltage of -1.0 V and a ballistic ratio of 81% at room temperature. This work presents a major breakthrough in p-type two-dimensional (2D) transistors in terms of contact optimization and device performance, filling the research gap in 2D p-type transistors.

11:35 AM
T13.5 The Thermal Paradox of Oxygen Vacancies: Screening vs. Depolarization Setting the Retention in W-Doped Indium Oxide FEFETs, Sharadindu Gopal Kirtania[1], Hyeonwoo Park[1], Emmanuel Quezada[1], Chengyang Zhang[1], Shimeng Yu[1], Asif Khan[1], Suman Datta[1]
[1]School of Electrical and Computer Engineering, Georgia Institute of Technology

This study investigates the physical mechanisms governing retention in W-doped indium oxide (IWO) FEFETs. The devices achieve record 7-day real-time retention at 25 °C and 85 °C, but exhibit pronounced retention loss in the PRG/LVT state at 125 °C. A comparative investigation of MFM and MFSM capacitor test structures isolates the dominant bottleneck as the depolarization field (E_{Dep}), which emerges from finite charge screening in the amorphous-oxide semiconductor (AOS) channel and its limited conduction-band density of states (DoS). Temperature-dependent measurements further show that thermally activated electron emission from donor-like oxygen vacancies can initially enhance screening, yet at higher temperature it increases the population of ionized V_O^{2+}, strengthens E_{Dep}, and accelerates depolarization-driven polarization back-switching. Finally, this work benchmarks state-of-the-art AOS FEFETs using retention, memory window (MW), and current window (CW) metrics, and extracts device-level design rules for improving thermal reliability in scaled AOS-based non-volatile memories.

Session T14: Emerging NVM: RRAM, MRAM, PCM

9:55 AM, Honolulu 3
Co-Chairs: Johannes Muller, GlobalFoundries and Tuo-Hung Hou, National Yang Ming Chiao Tung Univ

9:55 AM
T14.1 Decoupling Dual Degradation Pathways in OTS Endurance for High-Reliability Selector-Only Memory, Young-Sun Song[1], Kyu-Dong Park[1], Wonmoo Lee[1], Kyungpung Lee[1], Hunmo Yang[1], Bon-Jae Koo[1], Gabriel Jang[1], Chung-Hyun Lee[1], Woo-Hyun Park[1], Jae-Hyun Park[1], Zhe Wu[1], Se-Chung Oh[1], Kwang-Min Park[1], Seung-Hun Lee[1], Seung-Ju Yoon[1], Taek-You Kim[1], Sung-Ho Lee[1], Munbo Shim[1], Dae-Sin Kim[1], Seung-Pil Ko[1], Daewon Ha[1], Sujin Ahn[1], Jaihyuk Song[1]
[1]Samsung Electronics Co., Ltd.

We reveal two distinct degradation pathways for Ovonic threshold switch endurance: irreversible threshold voltage broadening from fatigue-driven elemental loss and partially reversible probabilistic reset-misread caused by filament instability. Their opposite stress dependences establish a unified framework validated by multi-scale experiments and simulations. Guided by this model, composition and bias co-optimization suppress both modes, achieving >10^8-cycle endurance in a fully integrated 14-nm Selector-only memory.

10:20 AM
T14.2 4nm FinFET CMOS Logic Embedded 1T Contact RRAM, Wei-Hwa Lin[1], Yen-Yu Liu[1], Yue-Der Chih[2], Yih Wang[2], Jonathan Chang[2], Ya-Chin King[1,3], Chrong Jung Lin[1,3]
[1]Institute of Electronics Engineering, National Tsing Hua University, [2]Design Technology Platform (DTP), Taiwan Semiconductor Manufacturing Company (TSMC), [3]College of Semiconductor Research, National Tsing Hua University

An ultra-high density 4nm FinFET CMOS logic embedded 1T Contact RRAM (CRRAM) is firstly developed in advanced TSMC 4nm FinFET CMOS logic platform with a small unit size of 0.0103μm^2/cell for high performance SOC applications. The 4nm FinFET 1T CRRAM consists of a 4nm N-type core FinFET device in serial of built-in FinFET contact TMO to form 1T CRRAM cell. The TiO_xN_y-based TMO serves as memristive storage node and formed by contact silicidation and ILD ALD process. In addition, the CRRAM is fully compatible with 4nm CMOS logic

FinFET process without extra mask. Furthermore, the unique contact TMO structure can be adapted to the next generations of CMOS FinFET and NSFET platforms. Moreover, the CRRAM shows stable and high operation performance, good reliability, and high scalability, that is, the ultra-high density 4nm FinFET 1T CRRAM is a very promising embedded memory solution for future HPC, AI, and SoC applications.

10:45 AM

T14.3 Scalable and stackable 2S1M SOT-MRAM memory cell enabled by two-terminal write and read SNGCT chalcogenide selectors, Elia Ambrosi[1], Saverio Ricci[1], Cheng-Chen Kuo[1], Mingyuan Song[1], Cheng-Hsien Wu[1], Chang-Hsien Lin[2], Wen Hsiang Lu[1], Chen-Yu Hu[1], Chengxing He[1], Xiaheng Huang[1], Sam Vaziri[1], Chen-Feng Hsu[1], Hengyuan Lee[1], Yi-Ju Chen[2], Chao-Cheng Lin[2], Chiao-Yun Lo[3], Hsin-Han Lee[3], Guan-Long Chen[3], Yu-Chen Hsin[3], Wei-Yen Woon[1], Jeng-Hua Wei[3], Kai-Shin Li[2], Tuo-Hung Hou[2], Kun-Lin Lin[2], Chien-Nan Jimmy Liu[2], Xinyu Bao[1]
[1]Corporate Research, Taiwan Semiconductor Manufacturing Company (TSMC), [2]Taiwan Semiconductor Research Institute (TSRI), [3]Industrial Technology Research Institute (ITRI)

SOT-MRAM cell size reduction is a key challenge in its conventional 2T1M configuration. An attractive approach is the replacement of 3-terminal Si transistors with 2-terminal BEOL selectors. In this work we report a 2S1M SOT-MRAM device with write and read threshold selectors based on the amorphous chalcogenide SNGCT. The device prototype structure and write/read operations are demonstrated for the first time. SNGCT selector can supply sufficient current for SOT-MRAM write. Endurance of 1e9 write cycles test-limited
is demonstrated, with low error rate. Non-destructive read with stable read current and wide 30μA sense window is obtained.

11:10 AM

T14.4 Ultra-Low Drift (4 mV/dec.) and Excellent Write Endurance (> 3×10^{10}) in Heterostructure Te-Based Selector-Only Memory (SOM) with 2D WS_2 Interlayer, Jangseop Lee[1], Seungkwon Hwang[1,2], Dongmin Kim[1], Laeyong Jung[1], Yoori Seo[1], Yu Bin Park[3], Byeongjin Park[2], Seongmo Kang[2], Yonghun Kim[2], Tae Hoon Lee[3], Hyunsang Hwang[1]
[1]POSTECH, [2]Korea Institute of Materials Science (KIMS), [3]Kyungpook National University

We report a heterostructured Te-based selector-only memory (SOM) utilizing a WS_2 van der Waals interlayer on nanoscale W plugs. By engineering a dangling-bond-free interface, we successfully controlled interface trap dynamics to enable efficient carrier injection. This architecture notably reduced forming voltage and suppressed drift characteristics (4 mV/dec.). Furthermore, the interlayer served as a robust diffusion barrier, enabling enhanced endurance exceeding 10^{10} cycles. Multiscale simulations clarify the defect-modulated switching mechanism driving this performance enhancement.

11:35 AM

T14.5 Automotive-Grade 5 nm Embedded MRAM with Robust Read Margin from -40℃ to 150℃ via Co-Optimized MTJ and Integration, Hyunsung Jung[1], Myoungsu Son[1], Wanjin Chung[1], Jaechul Shim[1], Youngkeol Kim[1], Jaehak Yang[1], Hideo Sato[1], Seongcheol Noh[1], Daeshik Kim[1], Jinwoo Choi[1], Hyukjun Kwon[1], NaYoung Ji[1], KyungChul Lee[1], Jinho Park[1], Cheol Kim[1], Whan Kyun Kim[1], Junho Jeong[1], Jeong-Heon Park[1], Shinhee Han[2], Byeoungju Lee[2], Gyuseong Kang[2], Seungpil Ko[1], Sangjin Hyun[1], Sujin Ahn[1], Jaihyuk Song[1]
[1]R&D center, Samsung Electronics Co., Ltd., [2]Foundry Business, Samsung Electronics Co., Ltd.

We demonstrate the first 5nm embedded MRAM (eMRAM) with 2T1R cell architecture that meets automotive-grade reliability requirements. By implementing a co-optimized process flow, we simultaneously enhanced key device metrics: tunneling magnetoresistance (TMR) by 15%, breakdown voltage by 15%, and suppressed bitline leakage current by over >6 orders of magnitude, significantly improving device reliability and scalability. The device shows no hard failures in both read and write operations, maintains robust read margin across –40 °C to 150 °C, and meets the automotive grade raw bit error rate (BER) requirement. This work enables 5nm eMRAM for automotive SoCs targeting AEC-Q100-level reliability, overcoming key scaling challenges in photolithography, dielectric integrity, and inter-cell leakage.

Joint Luncheon | 12:15 PM -1:15 PM, Coral 4-5

Innovative Neurotechnologies – A Journey from the Lab to the Clinic and Back, Madjid Hihi, Ph.D., CEA-Leti, Grenoble (France)
Establishing translational ecosystems that unite neuroscientists, VLSI circuit designers, materials engineers, and clinicians is essential for the development of scalable and reliable neurotechnology. Institutional settings such as CEA-Leti's Clinatec biomedical center in France demonstrate how long-term research, preclinical validation, and clinical evaluation can be combined within a unified framework. These environments support iterative development of fully implantable systems, their application in clinical settings, and their eventual transfer to industry.

Neurotechnologies developed at Clinatec in Grenoble include a Brain-Computer Interface for which a wireless electrocorticography device approved for long-term clinical use has been developed to benefit motor-disabled patients: **the WIMAGINE® system [1].**

World premieres have been successfully conducted at Clinatec by coupling WIMAGINE® with exoskeletons [2], as well as electrical spinal cord stimulators to close the loop in a so-called brain-spine interface [3]. These two applications have the objective to compensate **spinal cord injury patient mobility** and have paved the way to a successful **industrial transfer** in 2024 to ONWARD Medical [4]. The upcoming application targeted by the Clinatec team and their collaborators, such as at EPFL, is the use of WIMAGINE® coupled to motor rehabilitation systems, for **stroke** patient neurorehabilitation.

In this context, **onboard neural decoding** is emerging as a new technological frontier for restoring motor function and enabling effective neurorehabilitation. Advances in microelectronics, high-density electrode design, and embedded artificial intelligence are now converging toward autonomous neuroprosthesis, enabling the decoding of motor intentions and the driving of complex effectors in real time.
These advances will bring us closer to seamless, lifelong, and intuitive restoration of movement—where the boundary between brain and machine becomes physiologically continuous—and will in turn **transform how the brain interfaces with machines and therapeutic systems**.

[1] C. S. Mestais *et al.*, "WIMAGINE: Wireless 64-Channel ECoG Recording Implant for Long Term Clinical Applications," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 1, Art. no. 1, Jan. 2015, doi: 10.1109/TNSRE.2014.2333541.
[2] A. L. Benabid *et al.*, "An exoskeleton controlled by an epidural wireless brain–machine interface in a tetraplegic patient: a proof-of-concept demonstration," *The Lancet Neurology*, vol. 18, no. 12, pp. 1112–1122, Dec. 2019, doi: 10.1016/S1474-4422(19)30321-7.

Technical Sessions Block 9 (1:30 PM – 3:10 PM)

Session C25: Energy Efficient Wireline Transceivers and Clocking Circuits

1:30 PM, Tapa 2
Co-Chairs: Anand Vasani, Broadcom and Hisakatsu Yamaguchi, 1FINITY Inc.

1:30 PM
C25.1 A 2.05pJ/bit, 212.5Gbps DSP based Transceiver with 55dB Reach in 3nm FinFET, M. Gambhir[1], A. Mostafa[1], R. Chen[1], F. Chu[1], Z. Guo[1], X. Han[1], M. Hasan[1], A. Hassan[1], E. Hsiao[1], A. Lahiri[1], Z. Li[1], P. Liu[1], F. Lu[1], K.-M. Lu[1], P. Ramakrishna[1], M. Shannon[1], U. K. Shukla[1], A. K. Singh[1], M. Singh[1], Y.-P. Su[1], D. Visani[1], D. Zhou[1], H. Wang[1], K. Chang[1]
[1]Marvell Semiconductor Inc.

This work presents 212.5Gbps ultra-long reach PAM4 transceiver implemented in 3nm FinFET technology. 4-lane transceiver achieves pre-FEC BER of 2e-7 for 54.7dB bump-bump loss and 4e-9 for 49.5dB loss. Per-lane analog power efficiency at 212.5Gbps is 2.05pJ/bit which represents the lowest power as well as the longest reach published for this data rate. DAC current-mode driver-based transmitter achieves SNDR of 36dB, RLM of 0.97 and rms jitter of 50fs.

1:55 PM
C25.2 A 0.06pJ/b Current-Conveyor-based Current-Domain Transceiver with DC-Balanced Multi-Transmission-Line PAM Symbol Encoding, Seunghee Han[1], Seokman Kim[2], Sanghyun Han[2], Yeonghwan Kim[2], Joo-Young Kim[1]
[1]School of Electrical Engineering (EE), Korea Advanced Institute of Science and Technology (KAIST), [2]LeadingUI

This paper presents a current-domain transceiver using DC-balanced PAM symbol encoding across multiple transmission lines (TLs). By minimizing TL voltage swing to a few millivolts and transmitting information through controlled TL currents using an RX-side current conveyor (CC), the proposed architecture significantly reduces switching power and circuit area. The symbol encoding constrains a constant aggregate PAM level across a TL group, thereby reducing data-dependent simultaneous switching noise (SSN). CC-based current mirroring and ordering-based decoding remove the need for resistive I-V conversion and reference thresholds. A 4-TL PAM-4 group prototype fabricated in 14nm CMOS achieves 22.9Gb/s/pin data rate with 0.006mm^2 silicon area per TL and 0.06pJ/b energy efficiency. Limited by the CC current-feedback response rather than RC time constants, the proposed transceiver is expected to scale favorably with advanced technology nodes.

2:20 PM
C25.3 A 16.8–33.6Gb/s, 0.93pJ/b Direct-Frequency-Tracking Digital CDR with an SSC Profile Replicator for Per-Lane Data-Rate Variation and High Jitter Tolerance under Deep SSC, Jongwha Kim[1], Yongwoo Jo[2], Seongyun Min[1], Juntak Chun[1], Yuhwan Shin[1,3], Hyunjun Song[1], Seohyeon Kwak[1], Kwanseo Park[4], Jaehyouk Choi[1]
[1]Seoul National University, [2]Samsung Electronics Co., Ltd., [3]Korea Advanced Institute of Science and Technology (KAIST), [4]Yonsei University

This work presents a direct frequency-tracking (DFT) CDR that supports flexible per-lane data rates and improves jitter tolerance (JTOL) under deep spread-spectrum clocking (SSC). A ΔΣM-based fractional-*N* PLL directly tracks frequency offset, while an SSC profile replicator (SSC-PR) eliminates SSC-induced phase error in a type-II loop. The proposed CDR in 28nm CMOS demonstrates wide data-rate flexibility from 16.8 to 33.6Gb/s and improved JTOL under up to 7000ppm SSC. Operating at 0.7V without power-hungry phase interpolators (PIs), the highly digital CDR achieves an excellent energy efficiency of 0.93pJ/b.

2:45 PM
C25.4 A 1.39pJ/bit 224Gb/s PAM-4 Long-Reach SerDes Receiver in 3nm FinFET, Euhan Chong[1], Jacob Pike[1], Tae Young Goh[1], Reza Nikjah[1], David Berton[1], Paul Madeira[1], Masoume Akbari[1], Christian Bourget[1], Vinh Tuan Tran[1], Tung Nguyen[1], Sadok Aouini[1], Shahab Oveis Gharan[1], Naim Ben-Hamida[1]
[1]Ciena

This work presents a 1.39pJ/bit 224Gb/s PAM-4 DSP-based Serdes receiver fabricated in 3nm FinFET technology, achieving a BER of 3e-7 over a 38dB channel.

Session C26: UWB, IoT, and RF Enablers

1:30 PM, Honolulu 1
Co-Chairs: Ayman Shabra, MediaTek and Ibrahim Abdo, NTT, Inc.

1:30 PM

C26.1 A 3.25-Gb/s 0.74-pJ/bit IR-UWB Transmitter using Edge-Reference Differential Phase Shift Keying (ER-DPSK), Donggun Lee[1], Kyoungseok Song[1], Sangho Rhee[1], Tae Wook Kim[1]
[1]Department of Electrical and Electronic Engineering, Yonsei University

This paper presents a high-data-rate, low-power impulse-radio ultra-wideband (IR-UWB) transmitter employing edge-reference differential phase-shift keying (ER-DPSK). ER-DPSK uses the rising edge of each pulse as an inherent phase reference and encodes data in the phase difference between adjacent pulses. ER-DPSK eliminates the need for power-hungry phase-locked loops (PLLs) and preamble-based synchronization. ER-DPSK is compatible with pulse-position modulation (PPM), pulse-width modulation (PWM), and frequency-shift keying (FSK), enabling high-order hybrid modulation. The transmitter achieves 3.25 Gb/s with an energy efficiency of 0.74 pJ/bit. To the best of our knowledge, this is the first IR-UWB transmitter to reach the 3-Gb/s class while breaking the 1-pJ/bit energy-efficiency barrier.

1:55 PM

C26.2 A Cross-Band Wi-Fi/BLE Passive Tag Demonstrating Battery-Free Standard-Mask Communication with Smartphones and Battery-Free AoA Localization, Qijing Xiao[1], Changgui Yang[1], Junhong Sun[1], Weixiao Wang[1], Yunshan Zhang[1], Xin Hu[1], Kai Huang[1], Bo Zhao[1]
[1]College of Integrated Circuits, Zhejiang University

This paper demonstrates a Wi-Fi/BLE passive crystal-less tag achieving battery-free cross-band backscatter communication with smartphones. The proposed cross-band technique suppresses the image component of backscattered signal by more than 32dB, which eliminates the self-jammer and enables battery-free angle-of-arrival (AoA) localization. Moreover, the polar-modulated pre-distortion mask-shaping technique makes this chip fully compliant with both Wi-Fi and BLE standard masks.

2:20 PM

C26.3 A Switched Inductor-Capacitor Power Amplifier Supporting 100MHz 256-QAM, Yuncheng Zhang[1], Duo Li[1], Zihang Zhang[1], Hiroyuki Sakai[1], Kazuaki Kunihiro[1], Kenichi Okada[1]
[1]Institute of Science Tokyo

This paper presents a switched inductor–capacitor power amplifier (SLCPA) that enhances power efficiency. By inserting inductors into the capacitor array, the proposed SLCPA suppresses capacitor charging currents, mitigating capacitor-related losses, and improving efficiency. A prototype fabricated in 65 nm CMOS shows a 1.29× efficiency over Class-B PAs at 6-dB PBO. This design supports up to 100 MSymbol/s 256-QAM operation, achieving 18.8% DE, 17 dBm output power, and –29 dB error magnitude vector (EVM) without applying digital predistortion (DPD).

2:45 PM

C26.4 An IEEE 802.15.4z-Compliant IR-UWB System-on-Chip with 27 dB Self-Interference Cancellation and Reconfigurable TIA for Automotive Radar and Ranging , Hyun-Gi Seok[1], Sinyoung Kim[1], Wonjun Jung[1], Junhyeong Kim[1], Eunhye Park[1], Yeongdae Kim[1], Jaekeun Lee[1], Sumin Kang[1], Jonghoon Myeong[1], Duyong Seo[1], Honggul Han[1], Hyeonuk Son[1], Junyoung Jang[1], Hoon Kang[1], Chiyoung Ahn[1], Hyukjun Sung[1], Wan Kim[1], Hyun-Chul Park[1], Chan-Hong Park[1], Joonsuk Kim[1]
[1]Samsung Electronics Co., Ltd.

This work presents a UWB SoC with a TX spillover cancellation enabling shared-antenna full-duplex operation. The proposed PVT-robust FIR filter based SIC achieves 27 dB average rejection (σ = 1.4 dB), improving the RX noise figure from 21 dB to 4 dB. A reconfigurable TIA supports both of low-power ranging and wide bandwidth radar modes using a feed-forward architecture. Child presence detection is demonstrated in a real vehicle using a certified infant dummy and human subjects, including worst-case legroom scenarios.

Session C27: Next-Generation Wireless Baseband Processors

1:30 PM, Honolulu 3
Co-Chairs: Sophia Shao, University of California, Berkeley and Youngjoo Lee, Korea Advanced Institute of Science and Technology (KAIST)

1:30 PM
C27.1 A 17.6Gb/s, 39.1pJ/b Tiled Beamspace Training and Detection Processor for mmWave Massive MIMO, Jiyoon Han[1], Wei Tang[1], Cheng-Hsun Lu[1], Jungho Lee[1], Upamanyu Madhow[2], Zhengya Zhang[1]
[1]Electrical Engineering and Computer Science, University of Michigan, [2]Electrical and Computer Engineering, University of California, Santa Barbara

A 3.17mm^2 tiled beamspace training and detection processor is designed for mmWave massive MIMO. It integrates channel estimation, weight calculation, and detection using a 2D beamspace transform and adaptive windowing, achieving 17.6Gb/s throughput at 39.1pJ/b for a 256×8 uplink, with 461ns channel-estimation latency. Compared to recent work, this design provides up to 9.76× higher throughput, 6× higher area efficiency, and 1.78× lower training latency.

1:55 PM
C27.2 A 39pJ/b 7.3Gbps 1.3mm^2 Multi-Subcarrier Massive MU-MIMO-OFDM Detector Exploiting Beamspace Sparsity and Frequency-Domain Correlation in 22FDX, Abhishek Kumar[1], Seyed Hadi Mirfarshbafan[1], Oscar Castañeda[1], Christoph Studer[1]
[1]Dept. of Information Technology and Electrical Engineering, ETH Zürich

We present the first multi-subcarrier massive multi-user (MU) multiple-input multiple-output (MIMO) orthogonal frequency-division multiplexing (OFDM) data detector reported in the open literature. By exploiting the channel's beamspace sparsity and frequency-domain (FD) correlation, we achieve up to 3× area and power reduction. Our design supports U = 8 user equipments and B = 64 basestation antennas, computes soft outputs for QPSK to 256-QAM, and processes 16 subcarriers in parallel. The fabricated 22FDX ASIC has a core cell area of 1.3mm2, consumes 286mW, and delivers a throughput of 7.3Gbps at 0.8V core voltage, achieving best-in-class energy efficiency of 39pJ/b.

2:20 PM
C27.3 A 5.69mm^2 6.22Gbps User-Correlation-Based Hybrid Detector for 1024x64 Massive MIMO Systems, Chen-Chien Kao[1], Tianyu Wei[1], Wei Tang[1], Zhengya Zhang[1]
[1]Electrical Engineering and Computer Science, University of Michigan

This work presents a massive multi-user MIMO (MU-MIMO) detector for next-generation wireless systems. The 5.69mm^2 detector employs user grouping and hybrid detection, leveraging a mixed-precision architecture to achieve a 6.22Gb/s throughput with 1259mW. The chip supports the largest configuration to date, with 1024 base station (BS) antennas and 64 users using 256-QAM. Compared to the state of the art, the chip achieves up to 16.1× higher area efficiency and up to 8.4× lower normalized energy.

2:45 PM
C27.4 A 2.35mm^2 118.3Gbps 5.14pJ/bit Dual-Mode LDPC Decoder for 5G/6G, Yuqing Ren[1], Yuqi Wang[1], Leyu Zhang[1], Wenqing Song[1], Ludovic Blanc[1], Andreas Burg[1]
[1]EPFL

A dual-mode high-throughput LDPC decoder for 5G/6G is presented. The 2.35mm^2 decoder in 16nm FinFET delivers a peak throughput of 118.3Gbps with an energy efficiency of 5.14pJ/bit, supporting up to 10-core parallel 5G-LDPC decoding as well as spatially-coupled LDPC codes for 6G with code lengths up to 5x of the longest 5G-NR standard codes. Compared with the state-of-the-art, this work achieves 12.4x in peak throughput, 1.7x in area

efficiency, and 2.3x in energy efficiency. In 6G mode, it further provides up to 1dB coding gain advantage over 5G-LDPC codes.

Session JFS5: HPC Connectivity

1:30 PM, Tapa 3
Co-Chairs: Samuel Palermo, Texas A&M University and Meng-Fan (Marvin) Chang, National Tsing Hua University (NTHU) & TSMC

1:30 PM
JFS5.1 Membrane III-V Photonic Devices on Silicon Photonics Platform for Short-Distance Optical Interconnections, Shinji Matsuo[1], Tatsurou Hiraki[1], Tadashi Minotani[1,1], Takuma Aihara[1], Takuro Fujii[1], Yoshiho Maeda[1], Suguru Yamaoka[1,1], Yoshiya Shikama[1], Norio Sato[1], Tomonari Sato[1]
[1]NTT, Inc.

We have developed a membrane III–V electro-absorption modulator (EAM) integrated with a DFB laser on a silicon photonics platform. The membrane structure reduces the junction capacitance, enabling a 3-dB bandwidth exceeding 67 GHz and 150-Gbit/s NRZ operation. We also propose and demonstrate an optical chiplet using a membrane EAM array. By integrating substrate-embedded CMOS drivers, the optical chiplet achieves 64-Gbit/s PAM4 modulation with an energy cost of 0.75 pJ/bit.

1:55 PM
JFS5.2 A 3-nm FinFET CMOS PAM-4 Retimer for 800-Gb/s and 1.6-Tb/s Optical Modules, Andy Fan, Amrutha Iyer, Amber Tan, Beppe Cusmai, Chandrashekar Ramanna, Chengtao Song, Charlie Zhu, Hao Lo, Jinq Horng Teo, Jeffrey Wang, Kishore Karthe Ravi Prakash, Minsoo Choi, Naveen Shivashankar, Srinivasan Balakrishnan, Stark Chen, Saman Jafarlou, Shahram Mahdavi, Sagar Ray, Siyao Yu, Vivek Gurumoorthy, Wen Yan Neo, Yichao Wang, Zhuochao Sun, Zisong Wang, Cindra Abidin, Stephen Jantzi, Ken Chang
Marvell

This work presents a PAM-4 retimer implemented in 3-nm FinFET CMOS for operation beyond 100 GBd per lane. The design integrates DAC-based transmitters and ADC-based receivers operating up to 115 GS/s. The line-side transmitter employs digital pre-distortion, FIR pre-emphasis, and an optimized high-swing output stage to mitigate parasitic-induced distortion and group-delay variation. The receiver incorporates a wideband analog front end followed by a track-and-hold and a time-interleaved SAR ADC. Measured results demonstrate a 115 GBd PAM-4 eye with approximately 3-Vpp differential swing, RLM = 0.983, and TDECQ = 0.91.

2:20 PM
JFS5.3 A Monolithic Optical IQ Coherent Receiver in Unmodified Bulk CMOS, Samir Nooshabadi[1], Debjit Sarkar[1], David Baum[1], Craig Ives[1,2], Ali Hajimiri[1]
[1]California Institute of Technology, [2]now with Applied Materials

A monolithic optical coherent receiver with quadrature detection is presented for the first time in an unmodified bulk CMOS chip. Networks of optical phase shifters and splitters distribute modulated light across the chip using subtractive photonic waveguides operating at 780 nm, and integrated avalanche photodiodes and transimpedance amplifiers enable balanced detection and amplification in the electronic domain. All components are monolithically integrated in an unmodified 180 nm bulk CMOS process with simple post-processing steps, paving the way for low-cost optoelectronic imaging and sensing systems utilizing coherent detection.

2:45 PM
JFS5.4 A 28nm 0.7 pJ/bit Optically-Programmable, Fully-Digital SRAM-based Receiver Array with On-Chip Skew Calibration for Energy-Efficient Wireless Data Transfer, Yifan He[1], Sungjin Park[1], Hangyeol Mun[1], Xiaofeng

Hu[1], Peter McMahon[2], Alyosha Christopher Molnar[2], Jae-sun Seo[1]
[1]Cornell Tech, [2]Cornell University

This paper presents an Optically-Programmable, fully-digital SRAM-based Receiver (OP-SR) array to address the energy overhead of off-chip memory access. The 28nm chip demonstrates optical wireless data transfer using spatially modulated light at 0.7pJ/bit energy with on-chip skew calibration.

Session T15: Advanced Device Technology

1:30 PM, Tapa 1
Co-Chairs: Gilbert Dewey, Intel and Jin Cai, TSMC

1:30 PM
T15.1 Ultra Low-Power SRAM Technology on High Performance FinFET Platform for HPC, Mobile, and High-Reliability Applications, Seunhan Park[1], Yohwan Park[1], Tae Hyung Kim[1], Jinkyu Lee[1], Hongjeon Kang[1], Youngjae Seo[1], Hyungjin Kim[1], Sunghyuk Ha[1], Dongjun Kim[1], Jeongmin Choi[1], Jongki Jung[1], Jongjin Ro[1], Youngmin Park[1], Jongho Lee[1], Kyuwon Choi[1], Dukho Hong[1], Jongmoo Kim[1], Hyojong Cho[1], MinYong Jeong[2], Aliaksei Ivaniukovich[2], Taejung Kim[1], Jaehun Jeong[1], Ho Lee[1], Hyun-Jo Kim[1], Yuri Yasuda-Masuoka[1], Ja-Hum Ku[1]
[1]Foundry Business, Samsung Electronics Co., Ltd., [2]Semiconductor R&D Center, Samsung Electronics Co., Ltd.

Extreme Low-Power FinFET SRAM technology has been developed on the top of the latest advanced high performance 4nm FinFET Technology. The technology offers the world-best low-leakage SRAM, achieving a 0.52x bit-cell leakage current compared to the previous FinFET node. Additionally, SRAM V_{min} is successfully reduced by 70mV. To attain these advancements, total SRAM transistor design from Fin/SD to Gate stack is performed. As a result, in compiler level, significant leakage reduction, such as 0.35x in standby and 0.31x in retention mode, are achieved, as well as Performance +50% and Power -35%.

1:55 PM
T15.2 10-nm-Gate-Length Nanosheet CNFET with Self-Aligned Extension Doping for Ambipolar Leakage Suppression, Shengman Li[1,2], Mary Vaughan Petty[1], Nathaniel Safron[3], Aslan Wei[3], Shreyam Natani[4], Donglai Zhong[3], Sang Young Lee[3], Aylee A Wu[1], Prabhakar Bandaru[4], Andrew Kummel[4], Iuliana Radu[3], Gregory Pitner[3], Subhasish Mitra[1], H.-S. Philip Wong[1,3]
[1]Stanford University, [2]National University of Singapore (NUS), [3]Taiwan Semiconductor Manufacturing Company (TSMC), [4]University of California, San Diego

In this work, we demonstrate a self-aligned nanosheet CNFET technology that fundamentally suppresses ambipolar leakage through localized extension doping. Doped extensions between the gate and contacts enable aggressive gate-length scaling down to 10 nm while achieving a high on-current of 1.2 mA/μm and a low off-state current of 10 nA/μm at V_{DS}=−0.7V, representing over 10× leakage reduction compared with prior CNFET reports. Systematic device comparisons reveal that highly doped extensions reduce leakage by ~10× across both low and high drain bias, mitigate short-channel effects, and maintain high I_{max}/I_{min} over more than 5 decades of gate-length scaling. Further, CNT diameter reduction from 1.4 nm to 1.1 nm increases the bandgap, yielding an additional 100× leakage reduction at V_{DS}=−0.7V and SS of 95 mV/dec.

2:20 PM
T15.3 Interfacial-Layer-Free NCFETs Achieving 2mV Hysteresis, 27mV/dec SS, 2mV/V DIBL, and Record 2088μA/μm I_{ON} at V_{DS} = V_{OV} = 0.5V, Min-Kuan Lin[1], Jhih-Yuan Liang[1], Ying-Qi Liu[1], Wei-Jen Chen[1], Jui-Yu Hsu[1], Xuan-Hao Wu[1], Jie-Ni Dai[2], Yan-Jyun Chen[1], Bo-Hui Yu[1], Zhi-En Fan[3], Yi-Ming Mei[1], Ming-Hao Lee[1], Ming-Wen Chu[1], Pin Su[2], Chenming Hu[2,4], C. W. Liu[1]
[1]National Taiwan University, [2]National Yang Ming Chiao Tung University, [3]National Tsing Hua University, [4]University of California, Berkeley

To reduce V_{DD}, not only a small subthreshold swing (SS) of 27mV (minimum) and 53mV (average from I_{OFF} to V_T) mV/dec with only 2mV hysteresis, but also record high I_{ON} per footprint (2088μA/μm @ V_{DS}=V_{OV}=0.5V) are achieved by negative capacitance, with low CET (capacitance equivalent thickness) of 0.43nm and no interfacial layer (IL). IL-free stacked ferroelectrics around $Ge_{0.95}Si_{0.05}$ nanosheets accomplish all the above with negligible DIBL (2mV/V) and demonstrate a promising path to future low power AI chips.

2:45 PM
T15.4 Double-Flip Sequential CFET: Superior 3T Library Efficiency and PPA, Sheng Yang[1], Jürgen Bömmels[1], Halil Kükner[1], Fabian M. Bufler[1], Lynn Verschueren[1], Hans Mertens[1], Anne Vandooren[1], Naoto Horiguchi[1], Anita Farokhnejad[1], Geert Hellings[1]
[1]imec

Double-flip sequential CFET (sCFET) architecture enables split-gate devices and dense BM0 routing, supporting 3-track standard cells with high library efficiency. Hybrid orientation and asymmetric nanosheet widths in sCFET further optimize device-level PPA, offering scaling potential.

Session T16: Power Devices

1:30 PM, Honolulu 2
Co-Chairs: Nandu Mahalingam, Texas Instruments and Shiro Ozaki, NTT, Inc.

1:30 PM
T16.1 First Demonstration of Ga_2O_3 Reverse Blocking Transistors with $BV^2/R_{on,sp}$ Figure-of-merit Over 1 GW/cm² and Dual Blocking of 8 kV by Double Super-junction Design, Chenlu Wang[1], Sihan Sun[1], Hong Zhou[1], Yao Zhu[1], Xiaodong Zhang[2], Guoliang Peng[1], Min Zhou[1], Kun Zhang[1], Tao Zhang[1], Chunfu Zhang[1], Zhihong Liu[1], Yue Hao[1], Jincheng Zhang[1]
[1]Xidian University, [2]Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences

This work reports the first demonstration of reverse blocking n-Ga_2O_3 transistor by constructing p-NiO/n-Ga_2O_3 double super-junction (DSJ) MOS junction-FETs (DSJ-MOSJFETs). The DSJ-MOSJFET is beneficial for boosting on-state performance, leading to a large gate swing of 9.5 V and low specific on-resistance ($R_{on,sp}$) of 5.67 mΩ·cm². Meanwhile, DSJ technique also enhanced the breakdown voltage (BV) under the charge balance condition, resulting in a BV of 2.4 kV at gate-to-p-NiO spacing (L_{GP}) of 4 μm. Thus, the average electric field (E_{AV}) is pushed to be 6.1 MV/cm and the $BV^2/R_{on,sp}$ power-figure of merit (P-FOM) is yielded to be over 1 GW/cm². The repeated measurement and stress tests with both shift in V_{TH} less than 0.1 V illustrate the strong stability for this work. Combined with the both forward and reverse BV delivered to be > 8 kV, this work shows the great promise for future high-power and high-voltage power conversion applications.

1:55 PM
T16.2 6.5 kV Enhancement-Mode GaN Monolithic Bidirectional Switch (MBDS) Achieving Record Power Figure of Merit, Junjie Yang[1], Jingjing Yu[1], Hao Chang[1], Yunhong Lao[1], Jiawei Cui[1], Han Yang[2], Xuelin Yang[2], Xiaosen Liu[3], Maojun Wang[1], Bo Shen[2], Jin Wei[1]
[1]School of Integrated Circuits, Peking University, [2]School of Physics, Peking University, [3]School of Integrated Circuits, Tsinghua University

This work reports 6.5-kV GaN monolithic bidirectional switch featuring symmetric gate termination extensions (GTE-MBDS). The introduced GTE effectively modulates the electric field distribution under bidirectional stress to enhance breakdown voltage and to harden stability. Thus, the GTE-MBDS with an L_{GG} of 62 μm exhibits symmetric breakdown voltages (BVs) of −7874 V and 7884 V. With a specific ON-resistance (R_{SP}) of 25.97 mΩ·cm², GTE-MBDS achieves a high figure of merit (FOM = BV^2/R_{SP}) of 2.39 GW/cm². Both the BV and FOM represent record values among reported MBDS in SiC, GaN and Ga_2O_3. Furthermore, for the first time, this work reports the dynamic R_{ON} performance of GaN MBDS up to 2-kV stress, demonstrating a decent dynamic R_{ON}/static R_{ON} ratio of

1.43 after 2-kV stress at 150 °C. In addition, negligible conduction degradation was observed after holding the GTE-MBDS for 10^3 s at 3 kV and 150 °C.

2:20 PM
T16.3 AlN Buffered K/Ku Bands GaN RF Power HEMTs with in-situ SiN Silicidation Enabled Low R_C = 0.03 Ω·mm at 500 K and Record High P_{out} = 25 W/mm, Kun Zhang[1], Hong Zhou[1], Qifeng Lyu[2], Naiqian Zhang[2], Chenrui Zhang[3], Jianfei Liu[3], Yixu Yao[3], Chenlu Wang[1], Min Zhou[1], Zhihong Liu[1], Sen Huang[3], Yi Pei[2], Yue Hao[1], Jincheng Zhang[1]
[1]Xidian University, [2]Dynax Semiconductor Company Limited, [3]Institute of Microelectronics of the Chinese Academy of Sciences

Ku (12-18 GHz)/K (18-26.5 GHz) bands are typically used for radar, satellite communication and navigation et al., which are however limited by maximum output power density (P_{out}) of RF devices, mainly constrained by low breakdown voltage (BV) and increased high temperature (T) contact resistance (R_C). This work reports a record P_{out} = 25/24 W/mm at 18/20 GHz, which is twice of previous highest P_{out} value of 12 W/mm for Ku/K bands. This breakthrough is realized by combining a thick AlN buffer with enhanced carrier confinement and source field-plate (S-FP) structure for increased BV, and an in-situ SiN layer-assisted Ohmic silicidation enabled low R_C of 0.03 Ω·mm at T = 500 K. Those results indicate a significant progress for Ku/K bands GaN RF power devices to achieve high-power and high-efficiency RF applications.

2:45 PM
T16.4 First demonstration of f_{max} > 700 GHz in L_g = 45 nm $In_{0.13}Al_{0.83}Ga_{0.04}N$/GaN HEMTs for future 6G applications, Seung-Woo Son[1], Wan-Soo Park[1], In-Geun Lee[1], Su-Min Choi[1], Sang-Pyeong Son[1], Se-Hun Kim[1], Ji-In Byeon[1], Min-Kyu Song[1], Ho-Kyun Ahn[2], Dong-Hee Shin[3], Young-Kyun Noh[3], Sang-Kuk Kim[4], Jacob Yun[4], Ted Kim[4], Tae-Woo Kim[5], Jae-Hak Lee[1], Kyounghoon Yang[6], Dae-Hyun Kim[1]
[1]Kyungpook National University, [2]Electronics and Telecommunications Research Institute (ETRI), [3]IVWorks, [4]QSI, [5]Texas Tech University, [6]Korea Advanced Institute of Science and Technology (KAIST)

We report InAlGaN/GaN HEMTs with L_g = 45 nm that achieve record high-frequency characteristics. This outstanding performance is enabled by the successful integration of selective S/D regrowth and a high-aspect-ratio T-gate process using *mix-and-match* wafer-level integration. To gain insight into carrier transport and identify the key factors governing such performance, we employed methodologies based on L_g-scaling driven g_m modeling and delay time analysis. The L_g = 45 nm device exhibits a combination of R_{ON} = 0.33 Ω·mm, g_{m_max} = 1.03 S/mm, f_{max} = 742 GHz and f_{avg} = 497 GHz. To the best of our knowledge, this work represents the first demonstration of f_{max} > 700 GHz and the highest reported f_{avg} for any GaN HEMT technology.

Technical Sessions Block 10 (3:25 PM – 5:30 PM)

Session C28: Oversampled Noise Shaping ADCs

3:25 PM, Tapa 2
Co-Chairs: Erik Olieman, NXP Semiconductors and Tomohiro Nezuka, MIRISE Technologies Corporation

3:25 PM
C28.1 A 120-MHz BW CT 0-4 MASH DS ADC Employing a 2xTI SAR ADC and a Digital 1-0 MASH DSM DAC in 22nm FDSOI, Jonathan Ungethüm[1], John G. Kauffman[1], Maurits Ortmanns[1]
[1]Institute of Microelectronics, University of Ulm

A continuous-time (CT) 0-4 multi-stage noise shaping (MASH) Delta-Sigma (DS) ADC achieving high linearity without DAC calibration in a 120MHz BW is presented. The ADC uses a dual-return-to-zero (DRTZ) DAC with single-ended LSBs and requires a single 0.8V supply. The prototype fabricated in 22nm FDSOI operates at 2.4GHz sampling frequency and achieves 70.2dB SNDR and 93.3dB SFDR.

3:50 PM

C28.2 A 1MHz-BW 90.1dB-SNDR CT Zoom ADC with Mismatch Error Shaping Preserving Full-Scale Input Range, Junghyun Yoon[1], Zhaonan Lu[2], Yong Lim[1,3], Zhichao Tan[2], Youngcheol Chae[1,3]
[1]Yonsei University, [2]Zhejiang University, [3]XO Semiconductor

This paper presents a high-resolution, continuous-time (CT) zoom ADC with mismatch error shaping (MES) preserving the full-scale input range. To achieve a high SFDR over 100dB at a 1MHz signal bandwidth, this work introduces a partial-feedback coarse ΔΣM, a bitwise subtraction technique for MES, and a tri-level RDAC with a dummy DAC. Fabricated in a 65nm CMOS, it achieves 104.8dB SFDR, 90.1dB SNDR, and 92.5dB DR in a BW of 1MHz, while consuming only 1.7mW. This corresponds to a state-of-the-art FoM of 180.2dB.

4:15 PM

C28.3 A 91.9dB-SNDR 100kHz-BW DT ΔΣ ADC Employing Thermal Noise Suppression Techniques with Gain-Switched Dynamic Self-Quenching FIA, Ximing Wang[1], Jiaao Yu[1], Yunjie Chen[1], Haoming Zhang[1], Yuyang Zhu[1], Tetsuya Iizuka[1]
[1]The University of Tokyo

This paper presents a 3rd-order fully dynamic ΔΣ ADC employing thermal noise suppression (TNS) techniques. By suppressing both the sampling noise and amplifier noise of the discrete-time (DT) integrator, the proposed TNS improves the SNR of the ΔΣ ADC by 3.4dB. A 2-stage floating inverter amplifier (FIA) is employed to achieve stable gain for sampling noise cancellation (SNC) as well as high gain for accurate closed-loop operation. In addition, to enhance the output swing range of the FIA while maintaining power efficiency, a dynamic self-quenching FIA (DSQFIA) is proposed. The ADC prototype is fabricated in a 65nm CMOS process and achieves an SNR of 92.7dB, an SNDR of 91.9dB, and a DR of 93.7dB over a 100kHz bandwidth. With a power consumption of 199.2μW, the Schreier figure-of-merit (FoM) of 178.9dB is achieved.

4:40 PM

C28.4 0.144mm² Fully-Integrated-Onchip Fast-Converging Background Calibrations for 3-Stage 70.5dB-DR CT Pipelined ADC in 28nm CMOS, Kai Xing[1,2], Zi-Xuan Xu[1], Xiang-hui Pan[1], Yan Zhu[1,2], Rui P. Martins[1], Chi-Hang Chan[1,2]
[1]University of Macau, [2]University of Macau Advanced Research Institute

This paper presents a fully-integrated-on chip (FIO) background calibration technique for continuous-time Pipeline (CTP) ADC. The calibration leverages the split ADC architecture and tunable negative-R (NR) to facilitate a fast-convergence and low power OTA in the residue amplifying filter (RAF). The introduced digital pre-filters reduce the number of taps in the digital reconstruction filter (DRF) to 6 and enables FIO. The calibration is verified in a 70.5dB DR, 100MHz CTP ADC, which converges within 2E5 samples.

5:05 PM

C28.5 Robust Configurable 1.5MHz / 2.5MHz IF Quadrature CT ΔΣ Modulator Using SAQB and QDNC for 175.4 dB FoMs in 14 nm FinFET, Seong-Eun Cho[1], Ye-Dam Kim[1], Jongmi Lee[1], Yong Lim[1,2], Yong-Sik Kwak[1], Kyoung-Jun Moon[1], Sangmin Yoo[1]
[1]Samsung Electronics Co., Ltd., [2]Yonsei University

This paper presents a quadrature continuous-time delta–sigma modulator (QCTDSM) with low intermediate frequency (Low-IF) applications, featuring a configurable IF switching between 1.5MHz and 2.5MHz. The proposed architecture employs a Single-Amplifier Quadrature Biquad (SAQB) together with a proposed Quadrature Digital Noise Coupling (QDNC) technique, enabling sixth-order noise shaping with a compact, efficient loop structure. By utilizing the digitally operated QDNC scheme, the proposed modulator achieves improved robustness against process variations. As a result, the modulator achieves FoMs of 175.2 dB and 175.4dB for 2MHz and 4MHz BW, respectively. Measurement results from 50 fabricated chips demonstrate performance variations confined within

±3dB, validating the robust characteristics of the proposed design. Implemented in a 14 nm FinFET process, the prototype operates at 1.05V, consumes 914mW, and exhibits ±1.65dB of STF peaking.

Session C29: Advanced Memory Design

3:25 PM, Tapa 3
Co-Chairs: Mahmut Ersin Sinangil, Nvidia and Hoesam Jeong, SK Hynix

3:25 PM
C29.1 A 2nm 37.4 Mbit/mm^2 Dual-Rail SRAM with Row-Access Aware Read Tracking and Write Assist Circuits Enabling 2.28 pJ/Access Energy Efficient Operation, Ryo Takamatsu[1], Makoto Yabuuchi[1], Masaya Hamada[1], Tomotaka Tanaka[1], Shu-Huan Hsu[2], Yao-Yi Liu[2], Yuichiro Ishii[1], Koji Nii[1], Isabel Wang[2], Hong-Chen Cheng[2], Hung-Jen Liao[2], Tsung-Yung Jonathan Chang[2]
[1]TSMC Design Technology Japan, Inc., [2]Taiwan Semiconductor Manufacturing Company (TSMC)

This paper proposes an energy efficient high-density (HD) 6T single-port SRAM employing row-access aware read tracking and write assist (WA) circuits. In the read operation, the excessive bitline (BL) swing is narrowed by 45% at the worst condition, reducing the read dynamic power by up to 8.7%. It is also reduced by 8.5% on average thanks to the row-access aware read tracking and its optimal wordline (WL) deactivating timing. In the write operation, the WA circuit switches the negative BL boost capacitance by being aware of row-access, reducing up to 15% write dynamic power on average. 539-kbit dual-rail multi-array SRAM macros are demonstrated in 2-nm nanosheet technology [1], achieving a bit-density of 37.42 Mbit/mm^2. Measured silicon data show 0.35 – 1.10 V wide low voltage operation at 125 ℃ and an energy efficient read/write operation with 2.28 pJ/access.

3:50 PM
C29.2 Dual-Phase Offset Compensation and Conditional Latch Gating for a Fully Single-Ended DRAM Core Data-Line Architecture Operating at 0.75 V, Changyoung Lee[1], Youngseok Park[1], Hyeonjin Park[1], Donghak Shin[1], Keonwoo Park[1], Seryeong Yoon[1], Dongkyu Lee[1], Minsoo Kim[1], Gijong Sung[1], Kyuchang Kang[1], Sangyun Kim[1], Hyunchul Yoon[1], Bokyeon Won[1], Incheol Nam[1], Younghun Seo[1], Seungjun Bae[1], Youngsoo Sohn[1], Sangjun Hwang[1]
[1]Samsung Electronics Co., Ltd.

This paper proposes a fully single-ended DRAM core data-line architecture for low-power operation and improved backend routing efficiency. And also, this paper proposes a dual-phase offset compensation (DPOC) scheme to prevent charge transfer (CT) gain degradation caused by over-compensation, along with a conditional latch gating (CLG) technique to suppress CSL-induced latch disturb under low supply voltage. The proposed design was implemented in a test chip fabricated using a 10-nm-class DRAM process. At 0.75 V operation, the proposed scheme achieves a 99.4% reduction in fail bit count (FBC) compared to the case without CLG, while reducing IDD0 current by 27%.

4:15 PM
C29.3 A 25.3% Area-Efficient 1CPP Mask ROM with Robust Power Mesh on 2-nm NanoSheet Logic CMOS, Ryota Watanabe[1], Kazumasa Uno[1], Hidemitsu Kojima[1], Ayumu Yamada[1], Hiromasa Otsubo[1], Ted Chu[2], Tsung-Hsien Huang[2], Hong-Chen Cheng[2], Koji Nii[1], Yih Wang[2], Tsung-Yung Jonathan Chang[2]
[1]TSMC Design Technology Japan, Inc., [2]Taiwan Semiconductor Manufacturing Company (TSMC)

An area-efficient physical layout for a 1T ROM array with robust VSS mesh and coding procedure are proposed in this study. A 1.8-Mbit 1CPP ROM macro on 2-nm Nanosheet logic CMOS technology shows area-shrink by 25.3% from conventional physical layout, as well as access time, maximum operational frequency and active power improve by 14.8%, 10.6%, and 8.4%, respectively, while keeping the peak IR drop amplitude comparable to the baseline.

4:40 PM

C29.4 A 2nm Nanosheet 128Kb Anti-Fuse One-Time-Programmable Memory Featuring Core-Transistor-Only High Voltage Program Circuit and Fault-Tolerant Design, Gu-Huan Li[1], Wan-Hsueh Cheng[1], Chen-Wei Liang[1], Shaun Chou[1], Perng-Fei Yuh[1], Li-Yu Yeh[1], Yen-Hsiang Huang[1], Yi-Ching Liu[1], Chungi Huang[1], Duncan Yang[1], Yih Wang[1], Tsung-Yung Jonathan Chang[1], Lien-Jung Hung[1], Ping-Wei Wang[1], Chin-Shan Hou[1], Hsin-Jou Chuang[1], An-Shun Teng[1], Li-Ling Luo[1], Jordan Hsu[1]

[1]Taiwan Semiconductor Manufacturing Company (TSMC)

This paper presents a 128K-bit anti-fuse One-Time-Programmable (OTP) memory fabricated in a 2nm nanosheet CMOS process, utilizing a core-transistor-only architecture. A novel Decoded-stack High Voltage Scheme (DSHVS) is introduced, enabling reliable 5V programming despite the core-transistor-only constraint. This DSHVS effectively mitigates peripheral device reliability risks by maintaining all core transistors operating below 1V, far from their 2V-3V breakdown region. Validated across the full industrial temperature range (-40°C to 125°C), the design demonstrates robust performance with no functional failures post programming. Furthermore, an effective Triple Modular Redundancy (TMR) read scheme significantly enhances the minimum read voltage to 0.7xVDD, providing an additional 50% read margin. This achieves low macro ppm defect rates, eliminating the need for Error Correction Code typically employed in conventional 2-bit redundancy schemes.

5:05 PM

C29.5 A 16 Mb 28 nm FD-SOI STT-MRAM Achieving 2.0 ns Read Access Time at 0.9 V via a PVT-Variation-Tolerant Dual-Reference Sensing Scheme, Youjin Choi[1], Donguk Seo[2], Mingyu Kim[2], Heeyeon Kim[1], Kyungjun Lee[2], Jaerok Kim[2], Donggyu Kim[2], Shin Han[2], Yoonmyung Lee[2]

[1]Foundry Business, Samsung Electronics Co., Ltd., [2]Depatrment of Electrical and Computer Engineering, Sungkyunkwan University

A 16 Mb STT-MRAM is presented in 28 nm FD-SOI. To address PVT-induced read margin degradation, a read interface combines a dual-reference sense amplifier (DRSA) with an array-adaptive dual-reference (AADR) scheme. This technique extends the read window by tracking local resistance variations. The proposed sensing scheme expands the normalized zero-BER read window by 2.00×, 2.34×, and 2.62× compared to a conventional scheme at −40 °C, 25 °C, and 125 °C, respectively. The chip achieves a 2.0 ns read access time at 0.9 V/25 °C and 3.3 ns at 125 °C with BER = 0.

Session JFS6: Power Management Devices & Design

3:25 PM, Honolulu 3

Co-Chairs: Jonathan Douglas, Advanced Micro Devices and Sanghyeon Kim, KAIST

3:25 PM

JFS6.1 Backside-Compatible Oxide Vertical Channel-All-Around Transistors and Cylindrical Capacitors for On-Chip Voltage Conversion at High Power Density, Hung-Chun Chou[1], Song-Hyeon Kuk[1], Chengyang Zhang[1], Jungyoun Kwak[1], Emmanuel Quezada[1], Eknath Sarkar[1], Suman Datta[1], Shimeng Yu[1]

[1]Georgia Institute of Technology

This work presents monolithic integration of indium-gallium-oxide (IGO) vertical channel-all-around transistors (VCAAT) and vertical cylindrical flying capacitors (VCCaps) for high-efficiency point-of-load (PoL) active backside power delivery network (BSPDN) design at 3V rating, achieving a low specific on-resistance ($R_{on,sp}$) of 7.57 $\mu\Omega \cdot mm^2$ and a high capacitance density of 162 $fF/\mu m^2$. With calibrated device models from experiments, circuit-level simulation shows 80.2% efficiency at high power density (1 W/mm^2) with a proposed 4:1 ladder-based DC-DC converter for 3V to 0.75V step-down conversion, improving the power density by 4 times compared to recent works for the active BSPDN at the comparable voltage ratings and efficiency.

3:50 PM

JFS6.2 A 1.58-W/mm^2 Extreme-Power-Density Ka-Band GaN PA with Electro-Thermal-Mechanical Co-Design in CMOS+GaN Heterogeneous Integration, Ruitao Matthew Ma[1,2], Hongyu Bruce Bao[1,2], Shuo Sarah Feng[1,2], Zhiyu Liu[1,2], Yi Pei[3], Man Hoi Wong[1,2], Chik Patrick Yue[1,2]

[1]Hong Kong University of Science and Technology (HKUST), [2]High5 Semiconductor (Hong Kong) Limited, [3]Dynax Semiconductor, Inc

This paper presents a comprehensive electro-thermal-mechanical co-design methodology for CMOS+GaN heterogeneous integration, validated using a Ka-band GaN power amplifier (PA) featuring a record 1.58 W/mm^2 power density. To mitigate reliability risks from such extreme heat flux, we propose a reliability-aware packaging strategy on an Aluminum Nitride (AlN) substrate. We optimize the Residual Copper (RC) rate to resolve the fundamental trade-off between thermal dissipation and mechanical stress. Fabricated in a 150-nm GaN-on-SiC process, the compact 2.30-mm^2 PA chiplet delivers a saturated output power of 35.6 dBm (3.6 W) with 32.0% peak PAE. This work demonstrates that the proposed co-design framework guarantees system robustness even under aggressive power scaling.

4:15 PM

JFS6.3 A 300mm Monolithic GaN MOSHEMT and Si pMOS Technology Demonstrating Energy-Efficient Multi-Thousand-Gate Digital Logic for Power Management, Samuel James Bader[1], Heli Vora[1], Po-Han Chen[2], Saket Jha[2], Dimitri Frolov[3], Isha Khandelwal[3], Shreyas Samraksh Jayaprakash[3], Pratik Koirala[1], Adedapo Oni[1], Michael Beumer[1], Prafful Golani[1], Ahmad Zubair[1], Thomas Hoff[1], Andrey Vyatskikh[1], Marko Radosavljevic[1], Jagdish Rangaswamy[3], Ibukunoluwa Momson[3], Qiang Yu[3], Said Rami[3], Soumen Sarkar[3], Hanh-Phuc Le[2], Han Wui Then[1]

[1]Intel Foundry Technology Research, Intel Corporation, [2]Department of Electrical and Computer Engineering, University of California, San Diego, [3]Intel Foundry Design Technology Platform, Intel Corporation

This work demonstrates multi-thousand-gate, standard-cell-based digital circuits realized in a monolithic GaN nMOS and Si pMOS process. First, we report a 32-bit simplex serial-peripheral interface for configuring on-chip designs. Second, we report a record power-delay-product of 6.2 aJ/stage for a 7213-stage ring oscillator (>3 orders more efficient than GaN DCFL). These circuits show that "GaN+Si" technology expands the viable scale of monolithically-integrated controls for GaN power delivery and RF applications.

4:40 PM

JFS6.4 A 4.8V 16nm Stackable Hybrid Fully Integrated Voltage Regulator with in-package Magnetic Inductors featuring Automatic Voltage and Current Sharing between Stacks, Jingshu Yu[1], Suhwan Kim[1], Harish K Krishnamurthy[1], Minxiang Gong[1], Zakir K Ahmed[1], Sheldon Weng[1], Krishnan Ravichandran[1], James W Tschanz[1], Vivek De[1]

[1]Intel Corporation

A high voltage (4.8V), high efficiency (85%), high density solution (4A/mm^2 incl. passives) Stackable Hybrid FIVR (SH-FIVR) converter operating at >25MHz with an in-package magnetic inductor (3.8nH) is proposed. The SH-FIVR utilizes a series-input and parallel-output configuration with automatic voltage and current sharing to provide a modular solution for AI accelerators. This is the highest input voltage FIVR reported to date with compelling efficiency and density.

Session T17: Advanced Packaging and 3D Integration

3:25 PM, Tapa 1

Co-Chairs: Xinyu Bao, TSMC and Yoichiro Kurita, Institute of Science Tokyo

3:25 PM

T17.1 High-Density Chiplet Integration using Face-Down COW Processes with Bumpless Interconnects for Heterogeneous 3D Systems, Hideki Kitada[1,2], Yoshiaki Satake[3,4], Wataru Doi[3], Naoko Araki[5], Hiroyuki Ryoson[6],

Norio Chujo[1], Tatsuya Funaki[3], Azusa Yagi[7], Toshiyuki Takasaki[8], Takashi Yoda[1], Tadashi Fukuda[1], Takayuki Ohba[1], Hideki Kitada[6,7,7,8,9,10,11]
[1]Institute of Science Tokyo, [2]Institute of Science Tokyo, Japan, [3]Murata Manufacturing Co, [4]Murata Manufacturing Co., [5]Daicel Corp, [6]Dexerials corp, [7]Zacros corp, [8]Panasonic Connect Co. Ltd, [9]WOW Alliance, Institute of Science Tokyo, [10]Murata Manufacturing Co. Japan, Ltd, [11]Daicel Corp.

A face-down chip-on-wafer (COW) process for high-density chiplet integration using via-last bumpless interconnections has been developed. A proprietary "waffle wafer" suppresses warpage and enables a thin CoW structure with low thermal resistance. It was demonstrated that a face-down chip bonding scheme employing a thin SiOC adhesive layer achieves high chip-to-wafer alignment accuracy, enabling the minimum chip-to-chip spacing to be reduced to 10 μm. This reduced chip spacing, together with high-density bumpless interconnections, achieves a fourfold increase in shoreline density. After back grinding and CMP, the total CoW thickness is reduced to less than 50 μm. FEM analysis shows a 30% reduction in wafer warpage, and thermal analysis shows a 52% reduction in thermal resistance.

3:50 PM
T17.2 Record-High Power Dissipation (up to 5kW) Heterogeneous Integration Package with Direct-Chips-Attached Microfluidic Module Enabling TIM-Free Liquid Cooling , Yong Han[1], Boon Long Lau[1], Lai Yee Chia[1], Sharon Pei Siang Lim[1], Yong Liang Ye[1], Hsiang Yao Hsiao[1], Huicheng Feng[1], Gongyue Tang[1], Surya Bhattacharya[1], Yee-Chia Yeo[2]
[1]Institute of Microelectronics, Agency for Science, Technology and Research (A*STAR), Singapore, [2]Agency for Science, Technology and Research (A*STAR), Singapore

A heterogeneous integration (HI) package with direct-chips-attached (DCA) microfluidic cooling module is demonstrated by directly bonding a Si cooler baseplate to multiple chips, achieving a record-high power dissipation of 5 kW for the entire package, including SoCs with power of 4.2 kW and uniform chip heat flux of 360 W/cm^2. HI packages with multiple test chips were characterized with two DCA configurations: DCA-F with Full baseplate bonded to all chips and DCA-W with Windows to expose backside of chips for direct liquid cooling to eliminate thermal interface material (TIM) for extremely high power dissipation.

4:15 PM
T17.3 An Upper-Tier ESD Protection Scheme Using Parallel-Electrode ITO TFTs for Monolithic-3D Heterogeneous Integration Achieving >3000 V HBM Tolerance, Wenting Xu[1], Jinfan Wang[1], Li Lu[2], Feng Lin[2], Yuzhen Zhang[1], Tingrui Huang[1], Wangran Wu[1], Runxiao Shi[1], Weifeng Sun[1]
[1]Southeast University, [2]Central Semiconductor Manufacturing Corporation

A monolithic 3D (M3D) heterogeneous integration architecture with three stacked tiers is presented, integrating bottom-tier Si SRAM, middle-tier IGZO DRAM, and an upper-tier ITO thin-film transistor (TFT)-based electrostatic discharge (ESD) protection layer. A parallel-electrode ITO TFT structure is introduced to enhance voltage uniformity under ESD stress, increasing the breakdown current from 0.6 A to 1.38 A under transmission-line-pulse (TLP) testing. Based on the ESD-hardened devices, an upper-tier diode-connected TFT ESD scheme protects underlying silicon transistors, achieving >3000 V human-body-model (HBM) tolerance. This performance satisfies industrial Class 2 ESD requirements while eliminating silicon area overhead.

4:40 PM
T17.4 High-Density Backside Connectivity with Remaining Bulk Silicon Substrate Using Self-Aligned Local Backside Dielectric Isolation for Frontside Active, Takushi Shigetoshi[1,2], Peng Zhao[2], Liesbeth Witters[2], Violeta Georgieva[2], Daniel Montero Alvarez[2], Rajendra Kumar Saroj[2], Pallavi Puttarame Gowda[2], Thomas Altantzis[2], Tzu-Hsien Shen[2], Nicolas Jourdan[2], Bart Kenens[2], Evi Vrancken[2], Jan Willem Maes[3], Joeri De Vos[2], Gerald Beyer[2], Eric Beyne[2], Zsolt Tokei[2], Kan Shimizu[1], Yoshihisa Kagawa[1]
[1]Sony Semiconductor Solutions Corporation, [2]imec, [3]ASM

We demonstrate a Local Backside Dielectric Isolation (Local-BDI) process that enables nanoscale Via-last Through Silicon Vias (VL-TSV) overlap with the Active region, thereby overcoming design constraints in conventional approaches. Robust dielectric isolation of TSVs was confirmed across various layouts, without compromising backside-to-frontside connectivity. Compared to prior work, the Local-BDI increases the VL-TSV overlay window by threefold and, when TSVs are used as lines, reduces line resistance by over 50% compared to Buried Power Rail (BPR) at an equivalent footprint. This approach offers a solution for high-density backside connections with remaining bulk silicon.

5:05 PM

T17.5 Multiple-Wafer (9-layer), Extreme thin (3µm-Si per stack) and Innovative Fusion-bonded Via-in-one Architecture for High Bandwidth 3D Memory, Stephen Morein[1], Prashant Majhi[2], William Gilmore[2], Chun-Lin Lu[3], Sheng-Chieh Lin[3], Chih-Hao Chuang[3], Chih-Feng Sung[3], Chun-Hsun Chen[3], Chao-Wen Huang[3], Yu-Lun Chang[3], Shou-Zen Chang[3], Chin Hung Liu[4], Hsin Nan Hcueh[4], Chun Ting Yeh[4], Ping Tse Chen[4], Wenliang Chen[4]
[1]Saimemory Corporation, [2]Intel, [3]Powerchip Semiconductor Manufacturing Corporation, [4]AP Memory

A new cost-effective architecture of connecting fusion bonded wafers using a multiple wafer via last process is presented. Via-in-one TSV is demonstrated in a stack of custom DRAM wafers reducing power consumption during data movement (<0.7pJ/bit). Each stacked memory with ultra-thin (~3µm, Lower resistance of TSV) thickness of Si substrate and the TSV in oxide trench slot (10x85µm^2 per trench) with 20µm pitch (13.7Kea/layer) in the structure could provide better signal integrity during the high speed data transfer. Consistence of normalized resistance for each TSV and contact ring (<10% variation) describes the good quality of interconnection. The O-type contact ring was decided since its contact resistance is 40% lower than C-type one. Functional validation of 9-stacked DRAM (0.95V-1.2V) and successful reliability testing confirm its robustness. This architecture can offer exceptional memory bandwidth (~0.25Tb/s/mm^2) at low data transfer power (<0.35W/mm^2), presenting a completing solution for memory-intensive workloads in AI and HPC.

Session T18: Memories for AI Applications

3:25 PM, Honolulu 1
Co-Chairs: Elisa Vianello, CEA-LETI and Tuo-Hung Hou, National Yang Ming Chiao Tung Univ

3:25 PM

T18.1 A 4Mb 3bits/cell RRAM Macro in 12nm FinFET Technology with the smallest 0.019µm^{2}Bit-cell, 100K-cycle Endurance and 10 years Retention up to 125°C, Yuan He[1], Kaimeng Liu[1], Jianshi Tang[1], Cong Sun[1], Wei Chang[1], Chengxiang Ma[1], Haotian Mao[1], Huan Qiao[1], Kun Wang[1], Falong Zhou[1], Songchun Xu[1], Yilin Shen[1], Zhen Zhu[1], Mingcheng Shi[1], Xurui Zhu[1], Kun Wang[2], Xiangchao Ma[2], Dong Wu[1], Peng Yao[1], Bin Gao[1], He Qian[1], Huaqiang Wu[1]
[1]School of Integrated Circuits, Tsinghua University, [2]R&D, Beijing InnoMem Technologies Co., Ltd

This work presents a Design-Technology Co-Optimization (DTCO) methodology to address the critical challenges of RRAM integration in advanced FinFET technology nodes. By synergistically optimizing the access transistor, RRAM cell and program algorithm, we achieve the most aggressive cell size scaling and exceptional reliability simultaneously. A 4Mb RRAM macro is successfully demonstrated on a 12nm FinFET platform features the smallest 0.019µm^2 bit-cell. It exhibits robust reliability with 100K cycling endurance and 10-year data retention at 125℃ for 1-bit per cell (SLC) operation. Furthermore, reliable 3-bit per cell (TLC) operation is realized, achieving 10^9 read-disturb immunity and state-of-the-art 10-year TLC retention at 85℃. This work provides a viable path for high-density and highly reliable RRAM integration in advanced nodes, significantly accelerating RRAM deployment for future Edge AI applications.

3:50 PM

T18.2 Monolithic 3D Integration of Stacked IGZO/Si Memtransistor-based 2T0C Gain Cell for Attention Computation and CFET SRAM DCIM with multiple spectral CNT-Sensors for End-to-end Vision Transformers

(ViTs) on Edge AI, Yanqing Li[1], Yunjiao Bao[1], Mingcheng Shi[2], Shuang Liu[1], Feixiong Wang[1], Yi Ding[2], Xiying Liu[2], Yanyu Yang[1], Yanzhao Wei[1], Yadong Zhang[1], Gaobo Xu[1], Zhaohao Zhang[1], Yuyan Wang[2], Heyi Huang[1], Jianshi Tang[2], Huaxiang Yin[1], Xiaolei Wang[1], Tianchun Ye[1]
[1]Institute of Microelectronics, Chinese Academy of Sciences(CAS), [2]Tsinghua University

We demonstrate the first memtransistor-based gain-cell (M-GC) compute-in-memory (CIM) device implemented in a monolithic three-dimensional (M3D) chip integrating Si, IGZO, and CNTs technologies. Beyond logics, Si transisitor employing dopant-segregation exhibit memtransistor character suitable for memory applications. By vertically stacking an IGZO access transistor on the LT-Si read memtransistor, an innovative hybrid 2T0C gain cell is realized, enabling low-We demonstrate the first memtransistor-based gain-cell (M-GC) compute-in-memory device implemented in a monolithic three-dimensional (M3D) chip integrating Si, IGZO, and CNTs technologies. Beyond logics, Si transistors employing dopant-segregation exhibit memtransistor characteristics suitable for memory applications. By vertically stacking an IGZO access transistor on the low-thermal-budget-Si (LT-Si) read memtransistor, an innovative hybrid 2T0C gain cell is realized, enabling low-voltage programmable multistate and stable CIM operation. The device is structurally matched with DEtection TRansformer (DETR) vision transformers, where persistent key-value (K/V) tensors of the frame are stored in memtransistors and dynamic object queries are parallelly pipelined into the storage nodes. Combined with CFET SRAM-DCIM and on-chip multiple spectral CNT-sensor, this end-to-end Q/K/V-aware device-algorithm co-design achieves an 8.93× energy-efficiency improvement over the baseline, highlighting the potential of M-GC CIM and DCIM for edge vision transformers (ViTs).voltage programmable multistate and stable CIM operation. The device is structurally matched with DETR-style vision transformers, where persistent key-value (K/V) tensors of the frame are stored in memtransistor and dynamic object queries are parallelly pipelined into the storage nodes. Combined with on-chip multiple spectral CNT-sensor, this end-to-end Q/K/V-aware device-algorithm co-design achieves an 8.93× energy-efficiency improvement over the baseline, highlighting the potential of M-GC CIM for edge vision transformers (ViTs).

4:15 PM
T18.3 First Demonstration of Heterogeneous n-IGZO/p-WSe_2 Integration as a Power-Performance-Area Booster for BEOL Complementary Gain-Cell Memory, Hong-De Lin[1], Yuan-Chun Su[1], Tzu-Chuan Su[1], Elia Ambrosi[2], Cheng-Hsien Wu[2], Cheng-Chen Kuo[2], Chih-Yu Chang[3], Hung-Li Chiang[2], Ting-Hua Wei[1], Yu-Wei Hsu[4], Ching-Shuan Huang[5], Sin-Yue Lee[1], Wei-Ning Chang[6], Jui-Han Fu[7], Vincent Tung[7], Jack Y.-C. Sun[8], Chenming Hu[8], Chih-I Wu[4], Chao-Cheng Lin[6], Tuo-Hung Hou[5], Xinyu Bao[2], Tsung-En Lee[1,8]
[1]Department of Microelectronics, National Yang Ming Chiao Tung University, [2]Corporate Research, Taiwan Semiconductor Manufacturing Company (TSMC), [3]Pathfinding, Taiwan Semiconductor Manufacturing Company (TSMC), [4]Graduate School of Advanced Technology, National Taiwan University, [5]Institute of Electronics, National Yang Ming Chiao Tung University, [6]Taiwan Semiconductor Research Institute, [7]Department of Chemical System Engineering, The University of Tokyo, [8]Industry Academia Innovation School, National Yang Ming Chiao Tung University

This work presents the first demonstration of a BEOL-compatible complementary two-transistor gain cell (2T GC) memory, featuring a heterogeneous integration of an oxide semiconductor (OS) nFET as the write transistor and a 2D-material pFET as the read transistor. At a scaled channel length (L_{CH}) of 50 nm and width (W_{CH}) below 100nm, both n-IGZO and p-WSe_2 FETs exhibit large I_{on}/I_{off} (>10^8), leading to the record-high read current (I_{RBL}~22μA) and I_{RBL} margin of 1.5×10^5 for high-speed read operation. Though aggressive cell size scaling increases the challenge of capacitive coupling, the counter-coupling inherited from the NP-GC effectively reduces the coupling-induced voltage drop at the storage node by~0.57× during the read operation. The read margin (V_{margin}) is enhanced by ~1.52× through the optimization of the proposed n-OS/p-2D 2T GC structure. This design provides a scalable pathway toward high-density multi-layered cache memory stacked on advanced CMOS technology.

4:40 PM
T18.4 FAB Gain Cell Memory: A Folded Asymmetric Boosted Design to Improve Sense Margin for a Refresh-Free Operation, Jay Sonawane[1], Sunbin Deng[1], Kiseok Lee[1], Faaiq Waqar[1], Chengyang Zhang[1], Omkar Phadke[1],

Suman Datta[1], Shimeng Yu[1]
[1]Electrical and Computer Engineering, Georgia Institute of Technology

We propose a folded asymmetric boosted 3T gain cell (GC) to enable a refresh-free, high-speed on-chip memory, using amorphous oxide semiconductor (AOS) FETs. An underlapped write FET suppresses write WL coupling, while an overlapped storage/read gating FET enables read WL driven storage node (SN) boosting through a preferential capacitive coupling. This boosting operation shifts the sensing operation out of the near-threshold regime, improving V_{TH} variation tolerance. The measurements show a 1.48 V boosted memory window (at V_{DD} = 1 V) and >300 s at 85 °C data retention (3 orders of improvement over conventional 3T GC). The projected 7nm FAB design shows 2.6 ns read delay and satisfies the refresh-free requirement for GPU last-level cache, given maximum data lifetime profiled from AI workloads.

5:05 PM
T18.5 Vertically stacked Five-Word-Line IGZO FeFETs with Dual-Gate-Enabled Stable Erase Toward High-Bandwidth Storage, Zhuo Chen[1,2], Nicolò Ronchi[1], Roman Izmailov[1,2], Kruti Trivedi[1], Subhali Subhechha[1], Devin Verreck[1,2], Yang Xiang[1], Hongwei Tang[1,2], Geert Van den bosch[1], Attilio Belmonte[1], Maarten Rosmeulen[1,2], Valeri Afanasiev[1,2], Jan Van Houdt[1,2]
[1]imec, [2]KU Leuven

Electrical erase of hole-scarce IGZO FeFETs so far relies on fringing field by aggressive channel length scaling, which is difficult to reconcile with μm-long channel in 3D vertically stacked cells. In this study, we report on IGZO FeFETs with a Dual-Gate (DG) architecture that enables stable -erase over channel lengths from 0.1μm to 10μm, together with 1e8-cycling endurance. Moreover, for the first time, we demonstrate 3D vertically stacked DG IGZO FeFET strings with 5 word lines, showing robust ferroelectric memory window. The fabricated 3D DG devices achieve fast program (100ns) and erase (10μs), providing a low-voltage and high-endurance alternative to NAND flash for data staging/offloading-oriented HBS in AI inference.

Session T19: Imagers, Sensors and RF Devices

3:25 PM, Honolulu 2
Co-Chairs: Thanh Viet Dinh, NXP Semiconductors and Masafumi Tsutsui, Tower Partners Semiconductor

3:25 PM
T19.1 A Single Exposure 137 dB Dynamic Range CMOS Image Sensor with Low Dielectric Absorption Image Lag, Eiichiro Kosugo[1], Takaya Yamanaka[1], Ken Yahata[1], Mitsuki Hayashida[1], Takuto Ohashi[1], Kei Matsumoto[1], Makoto Aoki[1], Haruyuki Yoshida[1], Keisuke Watanabe[1], Yasuo Kiyofuji[2], Satoshi Gondo[2], Yorito Sakano[1], Yusuke Oike[1]
[1]Sony Semiconductor Solutions Corporation, [2]Sony Semiconductor Manufacturing Corporation

This paper presents a 2897×1877 CMOS image sensor with a 3.0 μm pixel pitch. The proposed sensor achieves a 137 dB single-exposure dynamic range (SEDR) at 85 °C and a 28 dB junction SNR at 105 °C. The sensor employs a sub-pixel architecture with 3D-MIM technology and includes a function to reduce the dielectric-absorption image lag. The combination of high dynamic range and reduced dielectric-absorption image lag makes the sensor well-suited for automotive applications.

3:50 PM
T19.2 A fully-buried transfer gate and polysilicon wire: Noise enhancement technology for CMOS image sensors with 2-layer pixels, Changkyu Lee[1], Myunghae Seo[1], Jihun Lim[1], Sooyeon Kim[1], Minho Jang[1], Jongeun Park[1], KeunYeong Cho[1], Jinyoung Kim[1], Minkwan Kim[1], Yongkun Jo[1], YoonSeok Kim[1], Dami Park[1], Gyunha Park[1], Jiyoun Song[1], Sungmin Son[1], Yongsang Park[1], Hajin Lim[1], Jonghyun Go[1], Jaihyuk Song[1]
[1]Semiconductor R&D Center, Samsung Electronics Co., Ltd.

In this paper, we develop a fully-buried transfer gate and polysilicon wire (FBP) structure that significantly improves noise performance in a 2-layer pixel CMOS image sensor. FBP converts all the devices made of poly-Si into fully buried structures and interconnects multiple floating diffusion nodes using buried poly-Si wire, thereby simultaneously reducing process steps and improving random noise. The Conversion Gain (CG) was improved by more than 33% compared to the conventional scheme.

4:15 PM

T19.3 Record-Performance In_2O_3 FETs with Ultra-scaled Contacted Gate Pitch: Enabled by Higher-*k* HZO Linear Dielectric and Novel Ti-Seeded Contact, Jian-Yu Lin[1], Zehao Lin[1], Chang Niu[1], Chang Liu[2], Juanjuan Lu[2], Kisoo Nam[1], Haiyan Wang[1,2], Peide D. Ye[1]

[1]Elmore Family School of Electrical and Computer Engineering, Purdue University, [2]School of Materials Engineering, Purdue University

In this work, we demonstrate two strategies to enhance the performance of oxide semiconductor (OS) FETs with ultra-scaled contacted gate pitch (CGP ≤ 80 nm). First, the gate capacitance-equivalent thickness (CET) is scaled to 1.36 nm using HZO-based linear dielectric with k > 28. Second, a novel ultrathin 0.5-nm Ti-seeding layer is introduced to further reduce contact resistance (R_C). By combining CET scaling and Ti-seeded metal contact, In_2O_3 FETs achieve a record-low R_C of 29.4 Ω·μm (approaching the Landauer limit), a record-high maximum current of 4.15 mA/μm, and a record-high on-current of 1.74 mA/μm among all reported OS FETs with contact length scaling. These results establish a new benchmark for OS FETs with ultra-scaled device footprints.

4:40 PM

T19.4 First Demonstration of High-precision (15 bits) Photonic-electronic Hybrid Compute-in-memory System Based on FeFET Pockels Photonic Memory , Haochen Yu[1], Yixuan Jiang[1], Ruihao He[1], Hao Yu[1], Zihao Deng[1], Hongwu Jiang[1], Shanshi Huang[1], Zefeng Xu[1]

[1]Microelectronics Thrust, The Hong Kong University of Science and Technology (Guangzhou)

This paper presents a photonic-electronic hybrid high-precision computing-in-memory (CIM) system. By leveraging the cyclic output characteristic of Pockels photonic memory and residual number system (RNS) encoding, the proposed system is closed under multiplication and addition, significantly reducing signal-to-noise ratio (SNR) requirements. The ferroelectric field-effect transistor (FeFET) Pockels photonic memory utilizes the Pockels effect for waveguide modulation and the screening effect for in-memory multiplication. With RNS implemented in this hybrid system, the system achieves up to 15-bit computational precision, showing improvement of up to 42% in inference accuracy on transformer-based model and 10dB in super-resolution task, compared to conventional analog CIM arrays.

5:05 PM

T19.5 Monolithic Photonic BiCMOS Technology with 400-GHz SiGe-HBTs and Germanium-Fin Photodiodes, Stefan Lischke[1], Bernd Heinemann[1], Daniel Steckler[1], Anna Peczek[1], Holger Rücker[1], Christian Mai[1], Aleksandra Kroh[1], Steffen Marschmeyer[1], Yuji Yamamoto[1], Andreas Krüger[1], Falk Korndörfer[1], Thomas Lenke[1], Philipp Kulse[1], Oksana Fursenko[1], Dirk Wolansky[1], Mirko Fraschke[1], Keerthi Dorai Swamy Reddy[1], Florian Teply[1], Stefan Simon[1], Thiemo Herbel[2], Robert Huber[2], Dietmar Kissinger[2], Lars Zimmermann[1,3]

[1]Technology, IHP – Leibniz-Institut für innovative Mikroelektronik, [2]Institut für Elektronische Bauelemente und Schaltungen, Universität Ulm, [3]HFT4, FG Silizium-Photonik, Technische Universität Berlin

We present a novel monolithically integrated SiGe-photonics technology that combines top-tier SiGe-HBTs with f_t/f_{max} = 390/440 GHz and Ge-fin photodiodes with bandwidths estimated >200 GHz. This ePIC platform is enabled by adapting IHP's SG13G3 SiGe-HBT modules and leveraging Ge-fin photodiode technology. As a proof of concept, we demonstrate a monolithic photonic receiver realized in this experimental fabrication run, achieving electro-optical bandwidth >110 GHz.