Sunday Workshop 1

Impact of Atomic Layer Processing and Selective Area Patterning on Device Fabrication and Performance [Shunju III]

Sunday, June 9, 19:00-22:00

Organizer: E. A. Joseph, IBM Research

Session 1
19:00 Effect of ALE on Semiconductor Device Properties, G.-Y. Yeom, Sungkyunkwan Univ.
19:25 Surface Reaction Analyses for Atomic Scale Processing by Beam Experiments, K. Karahashi, Osaka Univ.
19:50 Selective and Self-Limited Thin Film Processes for the Atomic Scale Era, R. Clark, TEL
20:15 Break

Session 2
20:30 Plasma-Based Selective Atomic Layer Deposition and Etching to Enable 5nm and Beyond Device Technology, E. Kessels, Eindhoven Univ. of Tech.
20:55 Selective Deposition: The Devil is in the Defects, H. J. Yoo, Intel
21:20 Creating 3D Nanoscale Structures by Area-Selective Deposition, A. Delabie, KU Leuven / imec

Sunday Workshop 2

Two Dimensional Materials and Applications [Le Bois]

Sunday, June 9, 19:00-22:00

Organizer: I. Radu, imec

Session 1
19:00 Path towards Scaling: 2D Materials, I. Radu, imec
20:00 How to Understand Interface Properties in 2D Heterostructure FETs, K. Nagashio, The Univ. of Tokyo
20:30 Electronic, Thermal, and (Some) Unusual Applications of 2D Materials, E. Pop, Stanford Univ.
21:00 (Opto-)Electronics: from 2D Materials to Devices, T. Mueller, TU Vienna
21:30 Workshop Wrap-Up, All

Sunday Workshop 3

Low Thermal Budget Dopant Activation for Sequential-3D Integration [La Cigogne]

Sunday, June 9, 19:00-22:00

Organizers: R. Choi, Inha Univ.
J.-M. Shieh, Narlabs-NDL
P. Batude, Leti

Session 1
19:00 Advanced Annealing Processes for Fin/GAA FETs Fabrications, Y.-. Lee, NDL - Narlabs
19:25 Various Junction Formation Techniques for Monolithic 3D Integration, R. Choi, Inha Univ.
19:50 SPER Optimized Junction for High Performance Devices within 500°C Thermal Budget, P. Batude, CEA-Leti
20:15 Break

Session 2
20:30 Low Thermal Budget Pulsed Laser Thermal Annealing for 3D Sequential Integration, K. Huet, Screen LASSE
20:55 Review on SPER Process in Si and SiGe (Damage Formation, Dopant Activation and Stability, Impact on Stress Relaxation), F. Cristiano, CNRS-LAAS
Short Course 1
CMOS Technology Enablers for Pushing the Limits of Semiconductors: Materials to Packaging [Shunju II, III]
Monday, June 10, 8:25-16:50

Organizers: M. Tada, NEC Corp.
N. Ramaswamy, Micron Technology, Inc.

8:25 Introduction
9:20 Emerging Interconnect Technologies for Nanoelectronics, K. Saraswat, Stanford Univ.
10:10 Break
10:40 Advanced Process Technologies Required for Future Scaling and Devices, R. D. Clark, TEL Technology Center
11:30 DTCO in 2019: The Precious Metal Stack and the Route to Better Designs, B. Cline and D. Prasad, Arm Ltd.
12:20 Lunch
14:00 Recent STT-MRAM Technology: From Lab to Fab, Y. J. Song, Samsung Electronics Co., Ltd.
14:50 Break
15:10 Emerging Logic Devices for Future Computing, S. Salahuddin, Univ. of California, Berkeley
16:00 Overview in Three-Dimensionally Arrayed Flash Memory Technology, R. Katsumata, Toshiba Memory Corp.

Short Course 2
Advanced 5G Circuits, Systems and Applications [Suzaku I]
Monday, June 10, 8:25-16:50

Organizers: H.-J. Song, POSTECH
A. Loke, TSMC

8:25 Introduction
8:30 5G Real and Future, T. Nakamura, NTT Docomo, Inc.
9:20 mmWave RFIC Technologies for 5G Infrastructure Applications, S.-G. Yang, Samsung Electronics Co., Ltd.
10:10 Break
11:30 Multi-Band, Low-IPN LO Generation for 5G and Beyond, J. Choi, UNIST
12:20 Lunch
13:10 Acoustic Filter for 5G Smartphones, H. Nakamura, Skyworks
14:00 Substrate Material and Packaging Technology for 5G Millimeter Wave Communication, K. Sudo, Murata Manufacturing Co., Ltd.
14:50 Break
15:10 Beamforming Circuits, Systems, and Operations for 5G MIMO Systems, H. Wang, Georgia Institute of Tech.
16:00 Built-In Test and Calibration of Phased Arrays, B. Floyd, NC State Univ.
Short Course 3
Opportunities and Challenges at the Intersection of Security and AI [Suzaku II]

Monday, June 10, 8:25-16:50

Organizers: M. Hashimoto, Osaka Univ.
X. Zhang, IBM
K. Maekawa, Renesas Electronics Corp.
N. Ramaswamy, Micron Technology, Inc.

8:25 Introduction

8:30 Introduction to Artificial Intelligence and Security, R. Aitken, Arm Research

9:20 Deep Learning Processors: Turning Challenges into Opportunities, H.-J. Yoo, KAIST

10:10 Break

10:40 AI Computing Architectures and Hardware, J. L. Burns, IBM Research

11:30 Nonvolatile Circuits for AI Edge Applications, M.-F. Chang, National Tsing-Hua Univ.

12:20 Lunch

13:10 RRAM Fabric for Neuromorphic and Reconfigurable Compute-In-Memory Systems, W. D. Lu, Univ. of Michigan

14:00 Circuit Design Resistant to Side Channel Attacks, N. Homma, Tohoku Univ.

14:50 Break


16:00 Introduction to Electromagnetic Information Security, Y. Hayashi, Nara Institute of Science and Technology

Demo Session & Reception [Suzaku I, II, III]

Monday, June 10, 17:30-19:30

Organizers: S. Otani, Renesas Electronics Corp.
K. Tateiwa, TowerJazz Panasonic Semiconductor Co., Ltd.
R. Aitken, ARM Ltd.
V. Narayanan, IBM

T2-5

T8-4

JFS2-3

JFS3-4
Energy-Efficient Continual Learning in Hybrid Supervised-Unsupervised Neural Networks with PCM Synapses, S. Bianchi*, I. Muñoz-Martín*, G. Pedretti*, O. Melnic*, S. Ambrogio** and D. Ielmini*, *Politecnico di Milano, Italy and **IBM Research, USA

C5-1
Joint Evening Panel Discussion

The Semiconductor Industry at a Tipping Point: What's Next? [Shunju II, III]

Monday, June 10, 20:00-21:30

Organizers: P. Yue, Hong Kong Univ. of Science and Technology
M. Kobayashi, The Univ. of Tokyo
K. Okada, Tokyo Institute of Technology
E. Naviasky, Cadence
G. Yeric, ARM Ltd.

Moderator: K. Makinwa, Delft Univ. of Technology

Panelists: B. Nauta, Univ. of Twente
Z. Wang, Tsinghua Univ.
F. Yinug, SIA
A. Piovaccari, Silicon Labs
S. Sumida, Woodside Capital Partners
J. Chang, TSMC

The field that has transformed the world and which we have annually gathered to celebrate is undergoing a metamorphosis. Economics no longer inexorably points down Moore's curve, price per gate has levelled or is rising. The leading edge nodes have become the territory of the very few companies that dare to use them. Simultaneously, the number of startups has shrunk by orders of magnitude. So where are we going? Together with you, a panel of experts with backgrounds ranging from academia, industry association, and companies from start-ups to established will attempt to provide some insights into our future.
SESSION 1

Joint Opening and Plenary Session 1 [Shunju I, II, III]

Tuesday, June 11, 8:00-10:00

8:00-
Joint Welcome and Opening Remarks
M. Masahara, AIST
M. Ikeda, The Univ. of Tokyo
C.-P. Chang, Applied Materials, Inc.
K. Chang, Xilinx Inc

8:40-
Plenary
Chairpersons: K. Takeuchi, Chuo Univ.
T. Palacios, MIT

C1-1 - 8:40 (Plenary)
Virtual Cyborg: Beyond Human Limits, M. Inami, The Univ. of Tokyo, Japan

The social revolutions have accompanied innovation of the view of the body. If we regard the information revolution as establishment of a virtual society against the real society, it is necessary to design a new view of body “JIZAI body (Virtual Cyborg)”, which can adapt freely to the change of social structure, and establish a new view of the body.

In this talk, we discuss how we understand of basic knowledge about the body editing for construction of JIZAI body (Virtual Cyborg) based on VR, AR and Robotics. Superhuman Sports: Applying Human Augmentation to Physical Exercise.

This talk will also present Superhuman Sports, a form of “Human-Computer Integration” to overcome somatic and spatial limitation of humanity by merging technology with the body. In Japan, official home of the 2020 Olympics and Paralympics, we hope to create a future of sports where everyone, strong or weak, young or old, non-disabled or disabled, can play and enjoy playing without being disadvantaged.

T1-1 - 9:20 (Plenary)
Managing Moore’s Inflection: DARPA’s Electronics Resurgence Initiative, W. Chappell, DARPA, USA

In June 2017, the DARPA Microsystems Technology Office (MTO) announced the upwards of $1.5 billion Electronics Resurgence Initiative (ERI) to ensure far-reaching improvements in electronics performance well beyond the limits of traditional scaling. The gains that came as electronics technology sprinted forward according to Moore’s Law were not guaranteed but realized through ingenuity and close collaboration between commercial industry, academia, and government. The present moment, beyond his law, is where Gordon Moore had true prescience. ERI is building on the long tradition of successful partnerships to foster the environment needed for the next wave of U.S. and allied semiconductor innovation.

SESSION 2

Tuesday, June 11, 10:30-12:35

Chairpersons: K. Miyashita, Toshiba Electronic Devices & Storage Corp.
G. Jurczak, Lam Research

T2-1 - 10:30

In this paper, we report the reliability characterization of 7nm FinFET technology, in which the highly scaled 6th generation of FinFETs and 256Mbit SRAM cells was newly developed by utilizing EUV. The single EUV patterning of MOL and BEOL resulted in significantly improved reliability distribution as compared to the previous nodes with multiple patterning techniques. The successful demonstration on product reliability including SRAM, Logic HTOL, and SER as technology evaluation was performed, indicating the 7nm technology+EUV is ready for high volume manufacturing.

T2-2 - 10:55

Electromigration (EM) and TDDDB reliability of Cu interconnects with a barrier/wetting layer as thin as 2 nm employing a PVD-reflowed through-CO self-forming barrier (tCoSFB) is demonstrated to meet the required specifications for 7 nm BEOL. The resulting Cu EM lifetime is 2000X longer than Cu interconnects with a standard scaled barrier/wetting layer. This tCoSFB Cu EM and TDDDB reliability performance were equivalent to pure Co metal interconnects, but with a 50% lower line resistance even down to 30 nm pitch dimensions. However, the annealing process for PVD-reflow Cu seed that enhances EM reliability caused Cu agglomeration at dual damascene line-end vias, leading to poor via-chain yield. Resolving this geometry-sensitive via-fill problem was identified as key to extending Cu manufacturability to 7 nm and beyond. We propose, and show preliminary data, for Cu/tCoSFB metallization with CVD Co via pre-fill as potential solution.
T2-3 - 11:20
3D Multi-Chip Integration with System on Integrated Chips (SoIC™), M.-F. Chen, C.-C. Hu, W.-C. Chiou and D. C. H. Yu, TSMC, Taiwan

The electrical characterization of System on Integrated Chips (SoIC™), an innovative 3D heterogeneous integration technology manufactured in front-end of line with known-good-die is reported. Chiplet integration of devices including foundry leading edge 7nm FinFET technology with SoIC™ illustrates its advantages in high bandwidth density and high power efficiency, as compared with 2.5D and conventional 3D-IC with micro-bump/TSV.

T2-4 - 11:45

Building compact and efficient reinforcement learning (RL) systems for mobile deployment requires departure from the von-Neumann computing architecture and embracing novel in-memory computing, and local learning paradigms. We exploit nano-scale ferroelectric tunnel junction (FTJ) memristors with inherent analogue stochastic switching arranged in selectorless crossbars to demonstrate an analogue in-memory RL system, which, via a hardware-friendly algorithm, is capable of learning behavior policies. We show that commonly undesirable stochastic conductance switching is actually, in moderation, a beneficial property which promotes policy finding via a process akin to random search. We experimentally demonstrate path-finding based on reinforcement, and solve a standard control problem of balancing a pole on a cart via simulation, outperforming similar deterministic RL systems.

T2-5 - 12:10

Here we show a hardware prototype of a monolithic three-dimensional (3D) imaging system that integrates computing layers directly in the back-end-of-line (BEOL) of a conventional silicon imager. Such systems can transform imager output from raw pixel data to highly processed information. To realize our imager, we fabricate 3 vertical circuit layers directly on top of each other: a bottom layer of silicon pixels followed by two layers of CMOS carbon nanotube FETs (CNFETs) (comprising 2,784 CNFETs) that perform in-situ edge detection in real-time, before storing data in memory. This approach promises to enable image classification systems with improved processing latencies.

Diversity Luncheon [Le Bois]
Tuesday, June 11, 12:45-13:55

SESSION 3

Focus Session - Quantum & Neuromorphic Computing [Shnju II, III]
Tuesday, June 11, 14:00-15:40

Chairpersons: K. Endo, AIST
M. Vinet, CEA-LETI, MINATEC

T3-1 - 14:00 (Invited)
Superconductive Parametric Devices, T. Yamamoto, NEC Corp., Japan

Superconducting parametric amplifier, originally developed more than half a century ago, gained renewed interests in some experiments of superconducting quantum electronics about a decade ago, and now has become an indispensable tool in the field of superconducting quantum information processing. More recently, there are several proposals, where a parametric oscillator, which is a parametric amplifier pumped above the threshold, is used as a quantum bit and the network of the parametric oscillators solve some computational tasks. Here, we briefly introduce the research activity on the development of the superconducting parametric devices, including our results, with some historical backgrounds.

T3-2 - 14:25 (Invited)

T3-3 - 14:50

This paper presents the first monolithic 3D two-layer reconfigurable SRAM macro capable of executing multiple Compute-in-Memory (CiM) tasks as part of data readout. Fabricated using low cost FinFET based 3D-IC, the SRAM offers concurrent data read from both layers and write from layer 2 with 0.4V V_{ddmin}. 12.8x improved computation latency is achieved as compared to near memory computation of successive Boolean operations.
T3-4 - 15:15

Extremely Compact Integrate-and-Fire STT-MRAM Neuron: A Pathway Toward All-Spin Artificial Deep Neural Network,
*, **National Chiao Tung Univ. and **Industrial Technology Research Institute of Taiwan, Taiwan

This work reports the complete framework from device to architecture for deep learning acceleration in an all-spin artificial neural network (ANN) built by highly manufacturable STT-MRAM technology. The most compact analog integrate-and-fire neuron reported to date is developed based on the back-hopping oscillation in magnetic tunnel junctions. This novel device is unique because it performs numerous essential neural functions simultaneously, including current integration, voltage spike generation, state reset, and 4-bit precision. The device itself is also a stochastic binary synapse, and thus eases the implementation of the compact all-spin ANN with high accuracy for online training.

SESSION 4

Ferroelectric I [Shunju I]

Tuesday, June 11, 14:00-15:40

Chairpersons: M. Kobayashi, The Univ. of Tokyo
N. Ramaswamy, Micron Technology Inc.

T4-1 - 14:00

Energy-Efficient Edge Inference on Multi-Channel Streaming Data in 28nm HKMG FeFET Technology, S. Dutta, W. Chakraborty, J. Gomez, K. Ni, S. Joshi and S. Datta, Univ. of Notre Dame, USA

We present a system implementing extremely energy-efficient inference on multi-channel biomedical-sensor data. We leverage FeFETs to perform classification directly on analog sensor signals. We demonstrate: (i) voltage-controlled multi-domain ferroelectric polarization switching to obtained 8 distinct transconductance (g_m) states in a 28nm HKMG ferroelectric FET (FeFET) technology [1], (ii) 30x tunable range in g_m over the bandwidth of interest, (iii) successful implementation of artifact removal, feature extraction and classification for seizure detection from CHB-MIT EEG dataset with 98.46% accuracy and <0.375/hr. false alarm rate for two patients, (iv) ultra-low energy of 47 fJ/MAC with 1,000x improvement in area compared to alternative mixed-signal MAC.

T4-2 - 14:25

Fundamental Understanding and Control of Device-To-Device Variation in Deeply Scaled Ferroelectric FETs, K. Ni, W. Chakraborty, J. Smith, B. Grisafe and S. Datta, Univ. of Notre Dame, USA

In this work, we present a comprehensive Kinetic Monte Carlo (KMC) modeling based statistical framework to evaluate the device-to-device variation of thin-film HfO_2 ferroelectric FET (FeFET). We conclude that the closing of the memory window in a FeFET array with device scaling can be attributed to: 1) limited number of domains; 2) variation among domains; 3) intrinsic stochasticity of individual domain switching. To enable further scaling of FeFET, co-optimization approaches from material, process, and device operation to control variation are proposed: i) increase the number of domains through material/process optimization (e.g. decrease of deposition temperature, etc.); ii) improve the uniformity of domains (e.g. minimizing the domain size variation and defect distribution, etc.); iii) increase the pulse amplitude/width to ensure deterministic switching of individual domains.

T4-3 - 14:50

Experimental Demonstration of Ferroelectric HfO_2 FET with Ultrathin-Body IGZO for High-Density and Low-Power Memory Application, F. Mo, Y. Tagawa, C. Jin, M. Ahn, T. Saraya, T. Hiramoto and M. Kobayashi, The Univ. of Tokyo, Japan

We have experimentally demonstrated a ferroelectric HfO_2 FET with memory operation by introducing ultrathin IGZO as a channel material. Ultrathin-body IGZO ferroelectric FET (FeFET) shows high mobility with deposited channel material, nearly ideal subthreshold slope, and controllable memory characteristics with the use of back-end compatible process. These results are attributed to the properties of IGZO channel: junctionless FET operation, nearly-zero low-k interfacial layer on metal-oxide channel and good capping effect for realizing ferroelectric phase formation with HfZrO_2. IGZO FeFET will open a new path for high-density memory application.

T4-4 - 15:15

Ferroelectric and Anti-Ferroelectric Hafnium Zirconium Oxide: Scaling Limit, Switching Speed and Record High Polarization Density, X. Lyu, M. Si, X. Sun, M. A. Capano, H. Wang and P. D. Ye, Purdue Univ., USA

The ferroelectric (FE) and anti-ferroelectric (AFE) properties of hafnium zirconium oxide (HZO) are investigated systematically down to 3 nm. The ferroelectric polarization, switching speed and the impact of atomic layer deposited (ALD) tungsten nitride (WN) electrodes are studied. Record high remnant polarization (P_r) on FE HZO and record high saturation polarization (P_s) on AFE HZO are achieved with WN electrodes, especially in ultrathin sub-10 nm regime. A high dielectric constant of 30.4 is achieved on AFE HZO. The polarization switching speed of FE and AFE HZO, associated with C-V frequency dispersion, are also studied. For the first time, it is found polarization switching speed is faster in AFE HZO than FE HZO, suggesting AFE-FET could be more promising for high speed memory devices.
Focus Session - 3D Integration and Packaging [Shinju II, III]

Tuesday, June 11, 16:00-18:05

Chairpersons: Y. Masuoka, Samsung Electronics Co., Ltd
M. Delaun, Analog Devices, Inc.

T5-1 - 16:00 (Invited)
The Future of Advanced Package Solutions, D.-W. Kim and T. Hwang, Samsung Electronics Co., Ltd., Korea

As the 4th industry revolution emerges into the semiconductor industry, high computing power and high data bandwidth are required for semiconductor devices. These demands lead to the adoption of the advanced packaging technology. For mobile application, fan-out technologies are used for smart phones due to small form factors and thermal performances. For server applications, 2.5D and 3D technologies are employed for cloud and artificial intelligence in terms of high memory bandwidth and a big die. However, there are two significant issues to resolve for advanced packaging. One is a thermal issue and the other is an electrical issue. Novel thermal materials and package structures are expected to improve the thermal performances. Redistribution substrate and through silicon via will reduce electrical loss for high speed signals. In this paper, we will investigate how the packaging technologies evolve in the future.

T5-2 - 16:25 (Invited)
Heterogeneous Integration Roadmap - Driving Force & Enabling Technology for Systems of the Future, W. Chen* and B. Bottoms**, *ASE and **3MTS, USA

Our industry has reinvented itself through multiple disruptive changes in market, products, and technology. We are at the triple inflection point, brought about by tech company disruption, Moore’s Law plateauing, and the explosive growth of the digital economy. Continued progress will require a new vision for electronic innovations. This paper shows examples of Heterogeneous Integration (SiP) for two market segments, Smart Phone and HPC Server and describes the purpose and organization of the Heterogeneous Integration Roadmap.

T5-3 - 16:50

The fan-out packaging technology has recently been adopted in mobile application processors due to its advantages in form factor, fine pitch traces, and efficient thermal dissipation. This paper demonstrates heterogeneous integration on a fan-out redistribution layer (RDL) interposer. The package has a full-reticle size Si die and two HBMs. Si die and memory modules are attached to a fanout RDL and are then attached to a multilayer substrate. This advanced package meets both electrical and mechanical requirements. The fanout RDL interposer is comprised of polymer and copper traces, and it is relatively mechanically flexible. Such flexibility enhances C4 joint integrity, and allows the new package to scale up its size to meet more complex functional demands.

T5-4 - 17:15
1 Kbit 6T SRAM Arrays in Carbon Nanotube FET CMOS, P. S. Kanhaiya, C. Lau, G. Hills, M. Bishop and M. M. Shulaker, Massachusetts Institute of Technology, USA

We experimentally demonstrate the first static random-access memory (SRAM) arrays based on carbon nanotube (CNT) field-effect transistors (CNFETs). We demonstrate full 1 Kbit 6 transistor (6T) SRAM arrays fabricated with CNFET CMOS (totaling 6,144 p- and n-type CNFETs), with all 1,024 cells functioning correctly without any per-unit customization. We demonstrate robust operation by writing and reading multiple patterns to the Kbit arrays and characterize single-cell SRAM variability (write and read margins) and repeat cycling of cells. Due to low-temperature BEOL-compatible processing, CNFET SRAM enables new opportunities for digital systems, since: (1) CNFET SRAM can be fabricated directly on top of computing logic, and (2) buried power rails (i.e., as in our demonstration where the power rails are fabricated underneath the FET) can potentially enable smaller-area SRAM layouts.

T5-5 - 17:40

3D sequential integration is shown to be compatible with a back gate implementation suitable for dynamic Vth tuning of the FDSOI top tier devices. The back gate is inserted seamlessly into the 3D sequential process flow during the top Si layer transfer, providing a close proximity to the top tier device, as well as a uniform and high quality thermal back oxide. A threshold voltage tuning of ~103mV/V and ~139mV/V is obtained in p- and nMOS top tier junction-less devices, respectively, over a back gate bias range of +/-2V. BTI reliability measurements show no detrimental impact of the back gate bias. Backgating can therefore be used to enhance the Ion performance with no reliability penalty. The buried metal line is also shown to lower crosstalk by metal shielding insertion between top and bottom tier metal lines, with a reduction larger than 10dB up to 45GHz.
T6-1 - 16:00
We present a scaling study toward 1Znm node 3D Cross-point PCM (XPCM) for Storage Class Memory (SCM) applications. The low operation current, and low metal line loading resistance are desired to avoid a wide operation voltage distribution in a cross-point array. For the first time, AC threshold voltage (Vth) of 1S1R OTS-PCM was studied, which will impact the operation scheme. To achieve Tera bits per chip density, six layers 1Znm 3D XPCM with OTS showing high Vth and low leakage current, and scalable periphery circuit are required.

T6-2 - 16:25
Ultra-Thin (<10nm) Dual-Oxide (Al₂O₃/TiO₂) Hybrid Device (Memory/Selector) with Extremely Low I_{off} (<1nA) and I_{reset} (<1nA) for 3D Storage Class Memory, C. Sung, J. Song, D. Lee, S. Lim, M. Kwak and H. Hwang, POSTECH, Korea
We demonstrate ultra-thin ALD-processed dual-oxide (Al₂O₃/TiO₂) hybrid device with memory and selector characteristics by engineering the stability of metal filament in Al₂O₃ and TiO₂ layer. The optimized hybrid memory device shows outstanding performances such as low off current (<1nA), low reset current (<1nA), and high on/off ratio (>10⁴). Inserting a Ti buffer layer which has a low electrode potential value, we observed excellent uniformity and retention property. Finally, an outstanding read/write margins and ultra-low power consumption are confirmed through array simulations of the proposed hybrid memory device.

T6-3 - 16:50
This work presents an optimized model for resistance evolution in PCM cells with Ge-rich GeSbTe (GST) alloy as active material. Unlike conventional Ge₂Sb₂Te₅, the low-resistance (set) state of Ge-rich GST shows a resistance drift to high resistance R, similar to the high resistance (reset) state, which could be a potential risk for data reliability. We develop a Monte Carlo (MC) model which predicts the time evolution of R at the statistical level of a memory array at various temperature T. The model is validated against variable temperature annealing, such as the soldering profile in embedded PCM, supporting the good reliability of Ge-rich GST at high T.

T6-4 - 17:15
We have demonstrated, for the first time, a combination of outstanding linearity of analog programming with matched PCM pairs, small analog programming noise, an extremely low resistance drift (R-drift) coefficient (0.005, median) and high endurance for a CVD-based confined phase change memory (PCM) with a thin metallic liner. In-depth analysis of linear analog programming is also presented. MNIST simulations using a pair of these confined PCM devices as a synaptic element yield a high test accuracy of 95%.
Technology Evening Panel Discussion
What Will the Foundries of the Future Do? [Shunju I]

Tuesday, June 11, 20:00-21:30

Organizers:  M. Kobayashi, The Univ. of Tokyo
G. Yeric, ARM Ltd.

Moderator:  A. DeVilliers, TEL

Panelists:  K. Mistry, Intel Corp.
A. Yu, GLOBALFOUNDRIES Inc.
C. Chidambaram, Qualcomm Technologies, Inc.
P. Jung, Samsung
T. Ohba, Tokyo Tech. Univ.
K. Zhang, TSMC

Conventional process node scaling has been extremely successful for many decades, but the challenges below 5 nm may require us to question our assumptions about the foundries of the future if we want the semiconductor industry to thrive in the next decade:

- Is EUV enough for patterning requirements? What is a key stopper?
- How long can FinFET/nanowire/nanosheet push the silicon MOSFET? Will something different be required?
- How specifically can “More-than-Moore” provide value beyond scaling?
- How much can 3D chiplets provide?
- Are there opportunities for other fab styles? Minimal fab vs. giga fab?

In this panel, we will discuss technology challenges and opportunities as well as new deliverable values to market, in order to provide perspective on the foundries of the future.

SSCS & EDS Young Professionals and Students Micro-Mentoring and Career Coaching Session [La Cigogne]

Tuesday, June 11, 18:15-19:15

SESSION 7
Remarks, Awards and Plenary Session 2 [Shunju I, II, III]

Wednesday, June 12, 8:00-10:00

8:00-
Remarks and Award Ceremony
M. Masahara, AIST
M. Ikeda, The Univ. of Tokyo
C.-P. Chang, Applied Materials, Inc.
K. Chang, Xilinx Inc

8:40-
Plenary
Chairpersons:  S. Yamakawa, Sony Semiconductor Solutions Corp.
B. Ginsburg, Texas Instruments

C10-1 - 8:40 (Plenary)
Computational Directions for Augmented Reality Systems, S. Rabii, E. Beigne, V. Chandra, B. De Salvo, R. Ho and R. Pendse, Facebook Inc., USA

Augmented reality (AR) is a set of technologies that will fundamentally change the way we interact with our environment. It represents a merging of the physical and the digital worlds into a rich, context aware and accessible user interface delivered through a socially acceptable form factor such as eyeglasses. One of the biggest challenges in realizing a comprehensive AR experience is managing power consumption to ensure both adequate battery life and a physically comfortable thermal envelope. This presentation reviews advanced concepts in minimizing power in data transfer across components, leveraging highly efficient accelerators while maintaining programmability, and the potential of emerging nonvolatile memories for low power computing.
SESSION 8

AI I [Shnju II, III]

Wednesday, June 12, 10:30-12:10

Chairpersons: S. S. Chung, National Chiao Tung Univ.
V. Narayanan, IBM

T8-1 - 10:30
Inference of Long-Short Term Memory Networks at Software-Equivalent Accuracy Using 2.5M Analog Phase Change Memory Devices, H. Tsai, S. Ambrogio, C. Mackin, P. Narayanan, R. M. Shelby, K. Rocki, A. Chen and G. W. Burr, IBM Research, USA

We report accuracy for forward inference of long-short-term-memory (LSTM) networks using weights programmed into the conductances of >2.5M phase-change memory (PCM) devices. We demonstrate strategies for software weight-mapping and programming of hardware analog conductances that provide accurate weight programming despite significant device variability. Inference accuracy very close to software model baselines is achieved on several language modeling tasks.

T8-2 - 10:55

A machine learning method is demonstrated for building a model of gait data for identification purpose. In contrast to ‘deep learning’ algorithms, our algorithm allows for a very fast, ‘one shot’ or ‘few shots’ learning. The algorithm is completely implementable in oxygen vacancy-based RRAM, and uses an intrinsic SET-dependent RESET probability in TaOx filamentary OXRRAM as local learning rule. Equal error rate down to ~10% is reached using less than 1Mb memory. Our algorithm relates to Non-Linear Dynamic Systems (NDS), allowing for interpretation and minimized hardware dimensioning.

T8-3 - 11:20

This work presents a hybrid CMOS-RRAM integration of spiking nonvolatile computing-in-memory (nvCIM) processing engine (PE) that includes a 64Kb RRAM macro and a novel in situ nonlinear activation (ISNA) module. We integrate the computing controller and nonlinear activation function on-chip to compute convolutional or fully-connected neural network. ISNA merges A/D conversion and activation computation by leveraging its nonlinear working region. This eliminates the need for additional circuits to realize nonlinearity and reduces area by 43.7 times w.r.t. the ADC scheme. The activation precision of ISNA can be configured from 1 to 8 bits to balance throughput, accuracy and power efficiency. The measurement of 4-layer LeNet shows such optimization improves 23.1% of computing speed via compromising a 2.5% relative accuracy drop. The proposed nvCIM PE achieves 16.9 TOPS/W power efficiency and a maximum spike frequency of 99.24 MHz.

T8-4 - 11:45

We demonstrate the first fully-printed resistive random access memory (RRAM) on flexible substrate using 2D layered dichalcogenides, exhibiting ultra-low switching voltage down to 0.18 V and an on/off ratio up to 10⁷. The novel switching medium is printed by formulating multilayer molybdenum disulfide (MoS₂) into 3D-printable ink. Both volatile and non-volatile resistive switching are achieved in a single device by varying current compliance, which enables the implementation of electronic synapse with neuromorphic functionality including short-term plasticity (STP) and long-term plasticity (LTP).
SESSION 9
Ge & SiGe FET [Shunju I]

Wednesday, June 12, 10:30-12:35

Chairpersons: K. Tomida, Sony Semiconductor Solutions Corp.  
S-C. Song, Qualcomm Inc.

T9-1 - 10:30
A Record $G_m$/SSSat and PBTI Reliability in Si-Passivated Ge nFinFETs by Improved Gate Stack Surface Preparation


We have demonstrated Ge nFinFETs with a record high $G_m$/SSSat and PBTI reliability by improving the RMG high-k last process. The SiO$_2$ dummy gate oxide (DGO) deposition and removal processes have been identified as knobs to improve electron mobility and PBTI reliability even with a nominally identical Si-passivated Ge gate stack. Surface oxidation of Ge channel during the DGO deposition is considered to impact the final gate stack. By suppressing the Ge channel surface oxidation, increasing mobility with decreasing fin width is obtained, whereas PBTI reliability, $D_{IT}$ of scaled fin as well as high-field mobility are improved by extending the DGO in-situ clean process, resulting in the record $G_m$/SSSat of 5.4 at 73 nm Lg.

T9-2 - 10:55
High Performance Strained Germanium Gate All Around P-Channel Devices with Excellent Electrostatic Control for Sub-30nm L$_G$


This paper demonstrates high performance strained p-type double stacked Ge Gate-All-Around (GAA) devices at significantly reduced gate lengths (L$_G$~25nm) compared to our previous work. Excellent electrostatic control is maintained down to L$_G$=25nm by using extension-less scheme, while the performance is kept by appropriate spacer scaling and implementation of highly B-doped Ge or GeSn as source/drain (S/D) material.

T9-3 - 11:20
SiGe Channel CMOS: Understanding Dielectric Breakdown and Bias Temperature Instability Tradeoffs


Breakdown and bias temperature instability for n/pFETs are studied on a wide composition of SiGe channels on different strain relaxation buffers. This study represents the first in-depth look at AC/DC PBTI trends of low Ge% SiGe nFinFETs. Dielectric breakdown is shown to be largely independent of channel composition over the region studied. Finally, we calculate the end-of-life performance benefit compared to Si, demonstrating the potential benefit of CMOS SiGe as a technology element.

T9-4 - 11:45
Channel Strain Dependence of Tinv in Strained Si and Si$_{1-x}$Ge$_x$ FETs: Internal Strain-Induced Modification of Chemical Oxidation


Self-limiting chemical oxidation behaviors on SiGe channels with biaxial and uniaxial strain are studied to understand the role of the internal channel strain in the interfacial layer (IL) formation. Biaxial strain accelerates chemical oxidation on Si and SiGe regardless of Ge content in the channel, forming a thicker IL in planar FETs. On the other hand, chemical oxidation is not sensitive to uniaxial strain thanks to the less strain in the out-of-plane direction, forming a thinner IL in FinFETs. The universal relationship of Tinv and channel strain from both planar FETs and FinFETs is discussed.

T9-5 - 12:10
Improvement of SiGe MOS Interface Properties with A Wide Range of Ge Contents by Using TiN/Y$_2$O$_3$ Gate Stacks with TMA Passivation

T.-E. Lee, K. Kato, M. Ke, M. Takenaka and S. Takagi, The Univ. of Tokyo, Japan

We demonstrate the low interface trap density at SiGe p-MOS interfaces with TiN/Y$_2$O$_3$ gate stacks by employing the Trimethylaluminum (TMA) passivation. PMA temperature is optimized to maximize scavenging of GeO$_x$. The impact of the gate electrode among TiN, Al, Au and W is studied for Y$_2$O$_3$/SiGe interfaces. The TiN/Y$_2$O$_3$/SiGe interfaces with PMA at 450°C minimize interface trap density ($D_{it}$), hysteresis and gate leakage current. TMA passivation is found to further improve the interfacial properties. The record-low minimum $D_{it}$ of $2.7\times10^{10}$, $5.4\times10^{10}$, $1.7\times10^{11}$, $2.0\times10^{11}$, $7.4\times10^{11}$ and $4.2\times10^{12}$ eV·cm$^{-2}$ are obtained for SiGe MOS interfaces with the Ge contents of 0.13, 0.22, 0.32, 0.38, 0.49, and 0.62, respectively.
T10-1 - 14:00
7nm Mobile SoC and 5G Platform Technology and Design Co-Development for PPA and Manufacturability,

We report on Qualcomm Snapdragon SDM855 mobile SoC and world's first commercial 5G platform using industry-leading 7nm FINFET technologies. SDM855 exhibits >30% CPU performance gain over the previous generation thanks to a new design architecture enabled by dual poly pitch process integration. Low voltage operation and tight spread in power consumption has been achieved through process and design co-development, delivering a high performance and low power solution for both mobile and AI applications. Extending the 7nm technology with 2nd-year process enhancement demonstrates up to 50mV CPU Vmin reduction without any change to de-sign rules, which paves the road for an integrated 5G mobile platform with >10Gbps connectivity.

T10-2 - 14:25
Accurate High-Sigma Mismatch Model for Low Power Design in Sub-7nm Technology,

High-sigma yield simulation analysis based on accurate SPICE mismatch model is required for high volume product design. Especially for the low power design in sub-7nm technology, the non-Gaussian behavior of the transistor drain currents (Id) is intensifying due to large mismatch variation. To achieve reliable high-sigma simulation, SPICE mismatch model needs to accurately reflect the non-Gaussian Id distribution obtained from the silicon data. Gaussian distribution modeling of channel resistance factor (Rch) and source/drain external resistance (Rext) is proven to be effective to model the skewed Gaussian distribution shape of massive silicon Ids data.

T10-3 - 14:50
Sub-10 nm Advanced FinFET Design for Different Applications in Various Vdd and Temperature Operation Ranges,

An advanced FinFET design is identified to improve both variation and minimum operation voltage (Vmin) in various temperature and supply voltage (Vdd) ranges, using sub-10 nm FinFET transistors. Through a clarification of each electrical parameter impact on both variation and operation voltage, a suitable FinFET design is successfully demonstrated to reduce Idoff variation by 0.4x, as well as Ioff variation by 0.8x in various Vdd ranges. This paper also provides Tr. design to improve Vmin by 35 mV with the switching energy 0.87x reduction.

T10-4 - 15:15
Fin Bending Mitigation and Local Layout Effect Alleviation in Advanced FinFET Technology Through Material Engineering and Metrology Optimization,

In advanced FinFET devices, STI gap fill and ILD stress are responsible for fin defects, fin bending as well as device performance degradations due to the local layout effect (LLE). In this paper, for the first time, we look at different ways to modulate the stress from the Flowable CVD (FCVD) films either by additional UV treatment and/or ion beam treatment (hot Helium implantation). By leveraging in-line e-beam metrology capabilities of PROVision for massive measurements of critical dimensions (CDs), the process impact on fin spacing and LLEs are characterized and analyzed. Significant improvement for LLE is observed for nFET device which correlates to fin bending improvement. In addition, ~5% drive current gain for pFET is observed after ILD stress optimization.
SESSION 11

Embedded Memory [Shunju I]

Wednesday, June 12, 14:00-15:40

Chairpersons: K. Tateiwa, TowerJazz Panasonic Semiconductor Co., Ltd.
Y. Pan, Lam Research

T11-1 - 14:00
High-Speed and Ultra-Low Power IoT One-Chip (MCU + Connectivity-Chip) on a Robust 28-nm Embedded Flash Process,

Based on robust 28-nm embedded flash (eFlash) process, IoT One-chip for high-speed and low power applications which MCU-chip (10Mb eFlash) and connectivity-chip (BLE/Zigbee) are integrated for the first time. By introducing new devices on 28-nm low-power eFlash process, high-speed (>40MHz random read), ultra-low power (<3uA sleep mode current, 10/13mA RF current at Tx/Rx mode) and robust reliability (-40~125°C stable operation, 100K cycle endurance, 150C/RT retention up to 200K hours) are achieved. LDD-first IO transistor with low Vth (~0.5V) for low-Vdd (~1.0V) operation and ultra-low leakage (ULL) SRAM bit-cell (0.1x vs normal) supporting low sleep mode chip current are applied to extend battery life-time. Stable endurance and high (low)-temperature retention after cycling stress are achieved by robust split-gate type eFlash cell.

T11-2 - 14:25
Turning Logic Transistors into Secure, Multi-Time Programmable, Embedded Non-Volatile Memory Elements for 14 nm FINFET Technologies and Beyond,

Described is a secure, multi-time programmable memory (MTPM) solution for the 14 nm FINFET node and beyond, which turns as-fabricated standard logic transistors into embedded non-volatile memory (eNVM) elements, without the need for any process adders or additional masks. These logic transistors, when employed as eNVM elements, are dubbed "Charge Trap Transistors" (CTTs). Outlined are the technological breakthroughs required for employing logic transistors as an MTPM. An erase technique, called "Self-heating Temperature Assisted eRase" (STAR), is introduced which enables 100% erase efficiency, as compared to < 50% erase efficiency using conventional methods, in turn enabling MTPM functionality in CTTs. For the first time, hardware results demonstrate an endurance of > 10^4 program/erase cycles. Data retention lifetime of > 10 years at 125°C and scalability to 7 nm have been confirmed.

T11-3 - 14:50
Embedded PUF on 14nm HKMG FinFET Platform: A Novel 2-Bit-Per-Cell OTP-Based Memory Feasible for IoT Security Solution in 5G Era,

In this work, a novel concept of 2-bit-per-cell (2B/C) is introduced to realize high-density OTP PUF from a new scheme of dielectric breakdown. This PUF shows 10^x of large window, good immunity to high-temperature disturbance, and excellent retention under 125°C baking, which are particularly for automotive applications. In terms of security, this PUF exhibits near ideal normal distribution of hamming distance and narrow distribution of hamming weight. The bit error rates are low, 0.78% at 25°C and 1.95% at 150°C, benchmarked on a 256-bit array. Finally, the security test of this PUF against the hackers attack from the machine learning process has been proved to have high security. Overall, the proposed 2B/C OTP PUF demonstrated great potential for IoT security in 5G era.

T11-4 - 15:15
Novel Quad Interface MTJ Technology and Its First Demonstration with High Thermal Stability and Switching Efficiency for STT-MRAM Beyond 2Xnm,

We have proposed novel quad-interface magnetic tunnel junction (MTJ) technology which brings forth an increase of both thermal stability factor Δ and switching efficiency Δ/ICO by a factor of 1.5-2 compared with conventional double-interface MTJ technology. We successfully fabricated the quad-interface MTJ using 300nm process based on novel low damage integration process including PVD, RIE and so on. By developing the quad-interface MTJ, we have achieved about two times larger Δ and Δ/ICO. Moreover, we have achieved about two times larger TMR ratio/RA by the stack development specific for the quad-interface MTJ technology. The developed quad-interface MTJ technology regarded as post-double-interface MTJ technology will become an essential technology for the scaling of the STT-MRAM beyond 20nm.
Technology / Circuits Joint Focus Session 1

New Computing [Suzaku III]

Wednesday, June 12, 14:00-15:40

Chairpersons: M. Yamaoka, Hitachi, Ltd.
A. Wang, Psikick

JFS1-1 - 14:00 (Invited)
A Cloud-Ready Scalable Annealing Processor for Solving Large-Scale Combinatorial Optimization Problems, M. Hayashi, T. Takemoto, C. Yoshimura and M. Yamaoka, Hitachi, Ltd., Japan

This paper presents a CMOS annealing processor (CMOS-AP) that accelerates ground state searches of the Ising model. The main feature of this processor is its inter-chip connection interface for making a larger chip. A credit card sized compute node integrating two CMOS-APs was also developed as an interface with existing computer systems. The compute node can handle up to 61,952 spins at a time. A performance evaluation using the node improved the CPU speed by 55 times in solving a minimum vertex cover problem, one of the NP-hard combinatorial optimization problems. Finally, we describe a cloud interface for the compute node to make the CMOS-APs more useful and to promote application development for it.

JFS1-2 - 14:25

A Sparse Matrix-Matrix multiplication (SpMM) accelerator with 48 heterogeneous cores and a reconfigurable memory hierarchy is fabricated in 40 nm CMOS. On-chip memories are reconfigured as scratchpad or cache and interconnected with synthesizable coalescing crossbars for efficient memory access in each phase of the algorithm. The 2.0 mm x 2.6 mm chip exhibits 12.6x (8.4x) energy efficiency gain, 11.7x (77.6x) off-chip bandwidth efficiency gain and 17.1x (36.9x) compute density gain against a high-end CPU (GPU) across a diverse set of synthetic and real-world power-law graph based sparse matrices.

JFS1-3 - 14:50
Spoken Vowel Classification Using Synchronization of Phase Transition Nano-Oscillators, S. Dutta, A. Khanna, W. Chakraborty, J. Gomez, S. Joshi and S. Datta, Univ. of Notre Dame, USA

The paradigm of biologically-inspired computing endows the components of a neural network with dynamical functionality, such as self-oscillations, and harnesses emergent physical phenomena like synchronization, to learn and classify complex temporal patterns. In this work, we exploit the synchronization dynamics of a network of ultra-compact, low power Vanadium dioxide (VO\textJane) based insulator-to-metal phase-transition nano-oscillators (IMT-NO) to classify complex temporal patterns for speech discrimination. We successfully train a network of four capacitively coupled IMT-NOs to recognize spoken vowels by tuning their oscillation frequencies electrically according to a real-time learning rule and achieve high recognition rates of 90.5% for spoken vowels. Such an energy-efficient compact hardware with a small number of functional elements are a promising technology option for edge artificial intelligence.

JFS1-4 - 15:15

In this paper, we present an integrated circuit which supports Full-HD photorealistic refocusing. In contrast to the conventional single-image blurring, it provides physically-correct bokeh effect by rendering and then averaging hundreds of novel views from five images taken in different perspectives. To address the huge requirement of DRAM bandwidth and computing power, we adopt a block-based multi-rate framework and further propose two techniques: four-direction view generation and highly-parallel view rendering. The former provides a compact system architecture to save 32% of SRAM area and 92% of DRAM bandwidth without noticeable quality degradation. The latter efficiently generates 5.4G novel pixels per second to provide high-quality refocusing. This chip is fabricated in 40nm CMOS process, and the core area is 3.61 mm\textsuperscript{2}. It consumes 250mW when operating at 200MHz and 0.9V to support Full-HD photorealistic refocusing up to 40 fps.
T12-1 - 16:00

In this work, we propose a novel split-gate FeFET (SG-FeFET) with two separate external gates to dynamically modulate the memory window (MW) for non-volatile memory and neuromorphic applications. During read operation, only one gate is turned on to decrease the area ratio ($A_{FE}/A_{IL}$) of ferroelectric layer to insulator layer, which increases MW and read current ratio ($I_{Read_1}/I_{Read_0}$). During write operation (program/erase), both two gates are turned on to increase $A_{FE}/A_{IL}$, which decreases MW, thereby resulting in lower write voltage ($V_{write}$). Compared to FeFET, SG-FeFET (1) Demonstrates lower $V_{write}$ (1.85V) and 59.5% reduction in write energy at fixed $I_{Write}/I_{Read_0}$ (2) Exhibits lower read energy (-11.3%) and higher $I_{Read_1}/I_{Read_0}$ (8.6E6) at fixed $V_{Write}$. (3) Allows random access and eliminates half-select disturb; (4) Preserves higher endurance due to lower $V_{Write}$ and charge trapping. SG-FeFET as synaptic device also exhibits superior symmetry and linearity for potentiation and depression processes.

T12-2 - 16:25
Bio-Inspired Neurons Based on Novel Leaky-FeFET with Ultra-Low Hardware Cost and Advanced Functionality for All-Ferroelectric Neural Network, C. Chen, M. Yang, S. Liu, T. Liu, K. Zhu, Y. Zhao, H. Wang, Q. Huang and R. Huang, Peking Univ., China

For the brain-inspired neuromorphic computing, various emerging memory devices, including FeFET, have been applied to develop the artificial synapses, while the artificial neurons are still mostly CMOS-implemented and suffer from high-hardware-cost issue, especially when expanding advanced functions. In this work, a novel leaky-FeFET (L-FeFET) based on partially crystallized Hf$_0.5$Zr$_0.5$O$_2$ layer is designed to mimic biological neurons. For the first time, we propose and experimentally demonstrate a capacitor-less L-FeFET neuron for basic leaky-integrate-and-fire function with ultra-low hardware cost of only one transistor and one resistor. Furthermore, a new hybrid L-FeFET-CMOS neuron is implemented to expand advanced spike-frequency adaptation with almost half of hardware cost compared with CMOS neuron. This work provides a highly-integrated and inherently-low-energy implementation for neuron and the possibility for all-ferroelectric neural networks.

T12-3 - 16:50

We will explore pure logic FinFET devices to realize the functionality of linear weight tuning capability as electric synapses. The unit cell of this new FinFET synapse is composed of two identical FinFETs in series, one serves as control and the other one as storage. This new FinFET synapse exhibits ideal linearity with nearly infinity training cycles (>10$^{12}$), much lower programming voltage, 0.85V, and faster speed, 2.5ns. It can also analogically increase or decrease the transistor $V_{th}$ to vary the drain conductance. For its analog performance, it performs excellent linearity and a wide tuning-window (20x) of weight-tuning. 1kb synaptic array has been designed. The Spice-simulated results have shown that new FinFET synaptic array can expand the array size to 64x64, given 300x of SNR, w.r.t. that of RRAM array. Finally, the training accuracy can achieve 97.43% accuracy as high as the GPU one does.

T12-4 - 17:15

Biologically plausible mechanism like homeostasis compliments Hebbian learning to allow unsupervised learning in spiking neural networks[1]. In this work, we propose a novel ferroelectric-based quasi-LIF neuron that induces intrinsic homeostasis. We experimentally characterize and perform phase-field simulations to delineate the non-trivial transient polarization relaxation that underlines the Q-LIF behavior. Performing network level simulations with our Q-LIF neuron model exhibits a 2-3x reduction in firing rate compared to traditional LIF neuron while maintaining iso-accuracy of 84- 85% across varying network sizes. Such an energy-efficient hardware for spiking neuron can enable ultra-low power data processing in energy constraint environments like health monitoring with wearable systems.
SESSION 13
Process [Shunju I]

Wednesday, June 12, 16:00-17:40

Chairpersons: T. Tsunomura, Tokyo Electron Ltd.
B. Colombeau, Applied Materials, Inc.

T13-1 - 16:00

In this paper, we present for the first time a Gate Cut Last integration scheme completed within the Replacement Metal Gate (RMG) module. This novel gate cut (CT) technique allows the scaling of gate extension length past the end fin which reduces parasitic capacitance, leakage and performance variation. In addition, we demonstrate that CT in RMG is a promising alternative integration process that can enable scaling for future logic technology nodes. Device, circuit and reliability results are shown to compare this novel CT in RMG process to the conventional gate cut method.

T13-2 - 16:25

More-Moore logic device technology roadmap suggests Lateral/Vertical Gate-All-Around (LGAA / VGAA) device architectures beyond FinFETs for further scaling and performance. At extremely scaled gate pitches, parasitic resistance significant impacts the performance of the devices. Direct partition measurement of the resistance components in FinFETs has been established. Stacked LGAA devices at further scaled gate pitches exhibit high S/D series resistance and contact resistance which can be partitioned with similar Kelvin measurement. VGAA transistors have a very different structure from FinFETs or LGAAs. Their asymmetric bottom and top S/D results in high S/D series resistance and contact resistance values. Separate partition of the resistances at the bottom and top is needed. In this paper, VGAA test structures and measurement methodology is introduced for partitioning the resistance components.

T13-3 - 16:50

We demonstrate a novel self-aligned gate contact (SAGC) scheme with conventional oxide/nitride materials that allows superior process integration for scaling while simplifying the SRAM cross-couple wiring. We show that the key feature to avoid both gate-contact (CB) to source-drain local interconnect (LI) shorts and the LI contact (CA) to gate shorts is the shape of the LI cap. A trapezoid-shaped oxide (SiO\textsubscript{2}) LI cap with an appropriate taper angle eliminates shorting between the contacts in the gate and source-drain region. We further demonstrate that this oxide LI cap is fully compatible with Cobalt (Co) metallization with a novel selective tungsten (W) growth process. Additionally, this process enables the SRAM cross-couple (XC) in the same metallization level, eliminating the need for an upper level wiring and greatly simplifying routing in the SRAM cell.

T13-4 - 17:15

A novel ladder transmission line method (LTLM) that features eliminated parasitic resistance from contact metal and access electrodes, simple fabrication process, and 2×10\textsuperscript{-10} Ω-cm\textsuperscript{2} resolution for highly-accurate extraction of specific contact resistivity ($\rho_c$) in the $10^{-10}$ to $10^{-8}$ Ω-cm\textsuperscript{2} regime is demonstrated. The current distribution and extraction of $\rho_c$ in LTLM are verified by TCAD and numerical resistive-resistor-network method, respectively. The extraction error caused by the current spreading and crowding in LTLM are modeled, and design guidelines to achieve 10\textsuperscript{-10} Ω-cm\textsuperscript{2} resolution for $\rho_c$ extraction are provided. By applying LTLM to the Ni p’-Ge\textsubscript{0.98}Sn\textsubscript{0.02} contact, a record-low $\rho_c$ down to 4.0×10\textsuperscript{-10} Ω-cm\textsuperscript{2} was obtained. LTLM is insensitive to variation of metal resistance, unlike the refined TLM (RTLM) which could overestimate $\rho_c$ by at least tens of times.
Integrated Power Management and Microcontroller for Ultra-Wide Power Adaptation Down to nW, L. Lin, S. Jain and M. Alioto, National Univ. of Singapore, Singapore

This paper presents a power management unit (PMU) driving a microcontroller, and controlling a power knob that enables adaptation to the sensed power availability over an ultra-wide range, well beyond voltage scaling. Conventional battery-powered operation is augmented with pure harvesting. Wide power adaptation is enabled by comparator delay self-biasing and zero-current switching scheme shared among all power modes with single-cycle convergence.

A 10mm$^2$ Light-Dose Sensing IoT$^2$ System with 35-to-339nW 10-to-300klx Light-Dose-to-Digital Converter, I. Lee*, E. Moon*, Y. Kim**, J. Phillips* and D. Blaauw***, *Univ. of Michigan and **CubeWorks, Inc., USA

This paper presents a 10mm$^2$ Internet-of-Tiny-Things (IoT$^2$) system that measures light dose using custom photovoltaic cells and a light-dose-to-digital converter (LDDC). The LDDC nulls diode leakage for temperature stability and creates headroom without power overhead by dual forward-biased photovoltaic cells. It also adaptively updates the current mirror ratio and accumulation weighting factor for a low, near-constant power consumption. The system can operate energy-autonomously at >500lx light level. The LDDC achieves a 3-sigma inaccuracy of ±3.8% and o/u of 2.4% across a wide light intensity range from 10lx to 300klx while consuming only 35 - 339nW.


Ppm-level hydrogen and ammonia in air were recognized by low-power, integrated sensors consisting of catalytic metal nanosheets. Thermal energy necessary for catalytic reactions were given by Joule heating not by external heaters. The thermal-aware design of sensors reduces the power consumption to 0.14 mW. The low-power and small-area properties enable large-scale, on-chip integration of molecular sensors, which will be useful in IoT era. A sensor array was successfully connected to a platform with wireless connectivity.


We demonstrate a record-high performance monolithic trantenna (transistor-antenna) using 65-nm CMOS foundry in the field of a plasmonic terahertz (THz) detector. By applying ultimate structural asymmetry between source and drain on a ring FET with source diameter ($d_S$) scaling from 30 to 0.38 micrometer, we obtained 180 times more enhanced photoresponse ($\Delta u$) in on-chip THz measurement. Through free-space THz imaging experiments, the conductive drain region of ring FET itself showed a frequency sensitivity with resonance frequency at 0.12 THz in 0.09 - 0.2 THz range and polarization-independent imaging results as an isotropic circular antenna. Highly-scalable and feeding line-free monolithic trantenna enables a high-performance THz detector with responsivity of 8.8 kV/W and NEP of 3.36 pW/Hz$^{0.5}$ at the target frequency.

In quest to execute emerging deep learning algorithms at edge devices, developing low-power and low-latency deep learning accelerators (DLAs) have become top priority. To achieve this goal, data processing techniques in sensor and memory utilizing the array structure have drawn much attention. Processing-in-sensor (PIS) solutions could reduce data transfer; computing-in-memory (CIM) macros could reduce memory access and intermediate data movement. We propose a new architecture to integrate PIS and CIM to realize low-power DLA. The advantages of using these techniques and the challenges from system point-of-view are discussed.

Computational Memory-Based Inference and Training of Deep Neural Networks, A. Sebastian*, I. Boybat**, M. Dazzi***, I. Giannopoulos****, V. Jonnalagadda*, V. Joshi*****. G. Karunaratne******, B. Kersting*******, R. Khaddam-Aljameh*******, S. R. Nandakumar*******, A. Petropoulos********, C. Piveteau*********., T. Antonakopoulos*********, B. Rajendran*********, M. Le Gallo* and E. Eleftheriou*, *IBM Research, **EPFL, ***ETH Zürich, Switzerland, ****New Jersey Institute of Technology, USA, *****RWTH, Germany and ******Univ. of Patras, Greece

In-memory computing is an emerging computing paradigm where certain computational tasks are performed in place in a computational memory unit by exploiting the physical attributes of the memory devices. Here, we present an overview of the application of in-memory computing in deep learning, a branch of machine learning that has significantly contributed to the recent explosive growth in artificial intelligence. The methodology for both inference and training of deep neural networks is presented along with experimental results using phase-change memory (PCM) devices.

A Ternary Based Bit Scalable, 8.80 TOPS/W CNN Accelerator with Many-Core Processing-in-Memory Architecture with 896K Synapses/mm², S. Okumura, M. Yabuuchi, K. Hijioka and K. Nose, Renesas Electronics Corp., Japan

A Processing-In-Memory (PIM) accelerator with ternary SRAM is proposed for low-power, large-scale deep neural network (DNN) processing. The accelerator consists of Ternary Neural Arithmetic Memory (TNAM) which is capable of bit-scalable MAC (multiply and accumulation) operation in accordance with target accuracy and power limit. An ADC less readout circuits to reduce analog-digital conversion power and a system-level variation avoidance technique utilizing features of TNAM are also proposed. A test chip with large-scale PIM is fabricated and successfully operate convolutional neural networks (CNNs) with 8.8TOPS/W and highest accuracy and area density among recent SRAM-type PIMs are obtained.

Energy-Efficient Continual Learning in Hybrid Supervised-Unsupervised Neural Networks with PCM Synapses, S. Bianchi*, I. Muñoz-Martin*, G. Pedretti*, O. Melnic*, S. Ambrogio** and D. Ielmini*, *Politecnico di Milano, Italy and **IBM Research, USA

Artificial neural networks (ANNs) can outperform the human ability of object recognition by supervised training of synaptic parameters with large datasets. Contrarily to the human brain, however, ANNs cannot continually learn, i.e. acquire new information without catastrophically forgetting previous knowledge. To solve this issue, we present a novel hybrid neural network based on CMOS logic and phase change memory (PCM) synapses, mixing a supervised convolutional neural network (CNN) with bio-inspired unsupervised learning and neuronal redundancy. We demonstrate high classification accuracy in MNIST and CIFAR10 datasets (98% and 85%, respectively) and energy-efficient continual learning of up to 30% of non-trained classes with 83% average accuracy.
**SESSION 14**

**GeSn Device [Shunju I]**

Chairpersons:  N. Sugii, Hitachi, Ltd.  
N. Collaert, imec

**T14-1 - 8:30**


We report the first demonstration of high performance germanium-tin (GeSn) multiple-quantum-well (MQW) photodiode (PD) on a 200 mm GeOI platform realized using a low temperature wafer bonding process. Record-low leakage of 25 mA/cm² was achieved for GeSn PDs using this new architecture. Both Ge p- and n-FinFETs were also realized on the GeOI platform to substantiate the promising monolithic integration of all GeOI-based photonic components with Ge CMOS on this architecture via top-down processing approach. This work paves way for advanced optoelectronic integrated circuits (OEIC) operating at 2 um band and beyond using GeSn as photo detection material for communication and sensing applications.

**T14-2 - 8:55**

Record Low Contact Resistivity (4.4×10⁻²⁰Ω-cm³) to Ge Using In-Situ B and Sn Incorporation by CVD with Low Thermal Budget (≤400°C) and without Ga, F.-L. Lu*, C.-E. Tsai*, C.-H. Huang*, H.-Y. Ye*, S.-Y. Lin* and C. W. Liu**, *, National Taiwan Univ. and **Taiwan Semiconductor Research Institute, Taiwan

The record low contact resistivity of 4.4×10⁻²⁰Ω-cm³ is achieved in Ti metal contact to in-situ B-doped GeSn using B (>1×10¹⁷cm⁻³) and Sn (>12%) segregations at the Ti and GeSn:B interface. Sn incorporation into Ge lowers the Schottky barrier height of holes. Increasing B doping at the Ti and GeSn:B interface reduces the hole tunneling distance. Thanks to the low growth temperature (305°C) of the chemical vapor deposition using Ge₂H₆, the GeSn:B with the bulk active [B] of 2.1×10¹⁷cm⁻³ (∼the solid solubility of B in Ge equivalent to 5.5×10¹⁶cm⁻³) and the bulk [Sn] of 4.9% is successfully grown. Without the needs of the previously reported Ga dopants and the high temperature annealing for dopant activation, the record low contact resistivity is achieved with all the process temperatures ≤400°C.

**T14-3 - 9:20**

First Vertically Stacked, Compressively Strained, and Triangular GeₐSnₜSn₀.₉₀ pGAAFETs with High Ion of 19.3μA at VGS=VGS=0.5V, Gm of 50.2μS at VGS=0.5V and Low SS of 84mV/Dec by CVD Epitaxy and Orientation Dependent Etching, Y.-S. Huang*, H.-Y. Ye*, F.-L. Lu*, Y.-C. Liu*, C.-T. Tu*, C.-Y. Lin*, S.-Y. Lin*, S.-R. Jan* and C. W. Liu**, *, National Taiwan Univ. and **Taiwan Semiconductor Research Institute, Taiwan

The natural etching stop on (111) facets yields the small dangling bond density and roughness, enabling low SS and high Ion on (111) sidewalls of the GAA channels. In addition, the ~2% uniaxial compressive strain and [Sn]=9% in the channel can reduce the hole effective mass. As a result, 50% improvement of Ion of 120μA/μm (perimeter), and 71% improvement of Gm of 31μS/μm are achieved than our previous 3 stacked GeSn (001) nanosheets. Record high Ion of 19.3μA per stack at VGS=0.5V and record Gm of 50.2μS per stack at VGS=0.5V among all GeSn FinFETs and GAAFETs are achieved. The SS, as low as 84mV/dec is also obtained, 22% reduction than our previous work.

**T14-4 - 9:45**


For the first time, complementary FinFETs and tunneling FinFETs with fin with (Wfin)of 20 nm and fin height (Hfin)of 40 nm were co-integrated on the same substrate, enabled by the formation of high quality GeSn-on-insulator (GeSnOI) substrate with 200 mm wafer size. Decent electrical characteristics were realized for both GeSn n- and p-FinFETs and n- and p-tunneling FinFETs. We also performed simulation study to show the promise of the GeSnOI platform that is not only able to suppress the off-state leakage current and improve the Ion/Ioff ratio of tunneling FETs, but also capable of providing the powerful flexibility of using back bias to achieve superior electrical characteristic beyond the benefits of incorporation of Sn into Ge for tunneling FETs.

**Technology / Circuits Joint Focus Session 4**

The Future of Memory [Shunju II, III]

Chairpersons:  H.-T. Lue, Macronix International Co., Ltd.  
G. Hemink, Western Digital Corp.

**JFS4-1 - 10:30**

Circuit and Systems Based on Advanced MRAM for Near Future Computing Applications, S. Fujita, S. Takaya, S. Takeda and K. Ikegami, Toshiba, Japan

Recently MRAM technologies have been intensively developed. This paper describes novel solutions using advanced MRAM for near future computing applications. Three beneficial applications with MRAM are presented: energy saving, data reliability and performance improvement.
JFS4-2 - 10:55

We demonstrated a cross-point memory array composed of 40nm Ag ionic memory cell with sub-μA and selectorless operation and 10-year data retention, making it a promising candidate for terabit-scale high-density memory application. Discontinuous conductive path with large and dense Ag clusters enabled 10-year retention even at sub-μA current with keeping high non-linearity in I-V. We implemented, for the first time, the improved cell into a 40nm cross-point array and demonstrated narrow read distribution which satisfies requirements for reliable array operation.

JFS4-3 - 11:20 (Invited)

MRAM can play a variety of on-chip memory roles in advanced VLSI technology spanning from high retention, solder-reflow-capable non-volatile memory (NVM) to dense non-volatile or high retention working RAMs. This paper describes results for a solder-reflow-capable MRAM NVM and for extensions that trade off high retention against speed, power, and density.

JFS4-4 - 11:45 (Invited)
The PCM Way for Embedded Non Volatile Memories Applications, P. Zuliani, A. Conte and P. Cappelletti, STMicroelectronics, Italy

A comparative analysis of different Resistive Memories proposed as Non Volatile Memories for embedded applications is here presented. Based on today scenario of industry-standard Floating Gate solutions, key factors as performances, reliability and technology maturity are considered when facing more innovative memory cells. In particular the race seems to be open at 28nm, where different players are proposing different memories integrated in the Back End Of the Line. Original results obtained on multi-megabits array integrating Phase Change Memories are here discussed covering cell scalability, High Temperature data retention and extended endurance capability, all in line with eNVM application requirements.

JFS4-5 - 12:10 (Late News)

We propose a field-free switching SOT-MRAM concept that is integration friendly and allows for separate optimization of the field component and SOT/MTJ stack properties. We demonstrate it on a 300 mm wafer, using CMOS-compatible processes, and we show that device performances are similar to our standard SOT-MTJ cells: reliable sub-ns switching with low writing power across the 300mm wafer. Our concept/design opens a new area for MRAM (SOT, STT and VCMA) technology development.

SESSION 15
Advanced FinFET & GAA II [Shunju I]

Chairpersons: H. Morioka, Socionext Inc. W. Rachmady, Intel Corp.

T15-1 - 10:30

For the first time, we establish a fabrication process flow of an EUV-era ultra-density 6-surrounding-gate-transistor SRAM with 0.0205 μm² unit cell area and demonstrate NMOS surrounding-gate-transistor function. In this paper, 6-surrounding-gate-transistor SRAM design layout is shown, and the fabrication process flow and key process steps are explained in detail. NMOS functional device characteristics of surrounding-gate-transistor is analyzed.

T15-2 - 10:55

This paper presents an 3D thermal impedance network approach to study self-heating effects in advanced FinFETs that are difficult to be analyzed in conventional models: (i) temperature distribution analysis for large FinFET devices used in high current drivers, (ii) transient thermal modeling for heat accumulation in GHz digital circuits, and (iii) investigation for layout methods to reduce FinFET self-heating temperature.
Moore's law, the principle that has powered semiconductor scaling for the past 50 years is nearing its end. However, the industry would like to pursue a dimensional scaling roadmap to reap the full benefit of technology innovation. Results shown on this paper demonstrate traditional dimensional scaling approaches involving multi-patterned lithography would skyrocket the manufacturing cost. Design level techniques collectively known as scaling boosters, and innovative Complementary FET (CFET) devices would help to reduce the cost of the technology nodes. To the best of our knowledge, this is the first approach where semiconductor node transitions are benchmarked based on their economic feasibility. To summarize, we have formulated a cost-driven approach that can guide the industry to continue semiconductor scaling.


The structure of the complementary FET (CFET) with NMOS stacked on top of PMOS, inherently yields standard cells and SRAM cells with 25% smaller layout area, 25% higher pin density and 2x higher routing flexibility than FinFET with same overall active footprint. Moreover, our work, based on advanced modelling, demonstrates that 4 track CFET can match and even outperform 5 track FinFET, without the need to lower S/D contact resistivity down to 5e-10Ω.cm or to elevate the channel stress up to 2 GPa. All gains in power-performance-area at circuit-level are maintained at block-level, making 4 track CFET a suitable candidate for N3 & N2 technologies.

Session 16

3D NAND [Shnju II, III]

Thursday, June 13, 14:00-15:40

Chairpersons: K. Hamada, Micron Memory Japan, Inc.
G. Hemink, Western Digital Corp.

T16-1 - 14:00


The advantage of using extremely-thin body (ETB, Tsi=3nm) device has been demonstrated in a 3D NAND Flash test chip. Net P/E memory window gain of >1.3V is observed for devices using ETB poly-Si. This substantial gain can be explained by the quantum confinement that raises effective Si bandgap and in turn reduces the tunneling barrier height. Simulation model has been validated and it shows equivalent barrier height reduction of ~0.16eV and 0.07eV for electron and hole, respectively for Tsi=3nm. However, the Idsat is degraded to only 160nA for Tsi=3nm, which is attributed to the larger effective mass or higher contact resistance. The degraded Idsat can be accommodated by lower Isense~30nA for page buffer circuit tuning. Random telegraph noise (RTN) is significantly reduced by extremely-thin body, and it shows tighter program-verify (PV) distribution in the MLC/TLC operation.

T16-2 - 14:25

For the first time, we have fabricated a confined nitrogen (SiN) trapping layer device for 3D NAND Flash and demonstrated excellent post-cycling retention performances. The key process step is to develop a uniform sidewall lateral recess in the 3D stack, followed by a SiN pull back process to isolate the SiN trapping layer in a self-aligned way. Excellent retention with only ~600mV shift of charge loss (out of initial 7V window) after 125°C 1-week high-temp baking for a post 1K cycled device was obtained. It is far superior to the control sample without confined SiN structure. Arrhenius analysis at various baking temperatures shows that the retention may pass >100 years at 60°C, and is even longer at room temperature. The device has potential to meet the low-cost long-retention archive memory applications.

T16-3 - 14:50

Right after program, stored electrons in the shallow nitride trap level can be released less than a few seconds. By setting the delay between program and reading phase to as small as 10μs, we found that several mechanisms are mixed when stored electrons are emitted during short term retention of 3-D NAND Flash. For the first time, we have confirmed that the charge loss mechanisms consist of three mechanisms and have separated each mechanism. In particular, the vertical redistribution of electrons in the charge trap layer, observed only during short term, was analyzed for the first time. Short term retention data measured at various temperatures (25-115°C) and at several program verify levels (PV3, PV5, PV7) in solid (S/P) and checkerboard patterns (C/P) were analyzed using our model. Finally, the activation energy (Ea) of each mechanism was extracted by the Arrhenius law and the magnitudes of Ea were compared.

T16-4 - 15:15
Pre-Shipment Data-Retention/Read-Disturb Lifetime Prediction & Aftermarket Cell Error Detection & Correction by Neural Network for 3D-TLC NAND Flash Memory, M. Abe, T. Nakamura and K. Takeuchi, Chuo Univ., Japan

This paper proposes two neural network techniques for 3D-TLC NAND flash memory. The first proposal is to predict data-retention/read-disturb lifetime for chip sorting during pre-shipment test. The second proposal is to detect and correct errors in aftermarket. First, in pre-shipment test, Neural Network-based Lifetime Prediction (NNLP) predicts future ECC decoding fail rate and estimates data-retention/read-disturb lifetime at higher endurance by using low endurance data. Then, NNLP sorts NAND flash memories into read-cold or read-hot applications. Chips with long data-retention lifetime are shipped to read-cold applications such as archives and cold flash storage. Chips immune to read-disturb are shipped to read-hot applications such as entertainment devices and cache of hierarchical storage. Second, in aftermarket, Neural Network-based Error Detection (NNED) detects and corrects errors. When ECC cannot correct errors, NNED detects and corrects part of errors and then BCH ECC completely corrects remaining errors. NNED decreases bit-error rate by 81.4%.

SESSION 17
Ferroelectric II [Shunju I]

Chairpersons: S. Takagi, The Univ. of Tokyo
K. Benaisa, Texas Instruments

T17-1 - 14:00
Transient Negative Capacitance as Cause of Reverse Drain-Induced Barrier Lowering and Negative Differential Resistant the Memory FETs, C. Jin, T. Saraya, T. Hiramoto and M. Kobayashi, The Univ. of Tokyo, Japan

We have investigated transient ID-IV characteristics of ferroelectric FET (FeFET) by simulation with ferroelectric (FE) model considering polarization switching dynamics. For the first time, we show transient negative capacitance (TNC) with polarization reversal and depolarization effect results in sub-60 SS, reverse drain-induced barrier lowering (R-DIBL), and negative differential resistance (NDR) without traversing the quasi-static negative capacitance (QSNC) region in S-shaped P-V characteristic. The mechanism demonstrated in this work can be a possible explanation for the previously reported negative capacitance FET (NCFET) with steep SS, R-DIBL, and NDR.

T17-2 - 14:25

This paper clarifies for the first time the origin of ferroelectricity in the Negative Capacitance Field-Effect Transistors (NCFETs) by molecular dynamics (MD) simulation. MD simulation considering atomic interactions between all atoms enables accurate predictions for the microstructure even at all interfaces. By incorporating the results from MD simulations into a kinetic model, it is able to predict the conditions of crystallization and phase transition during RTP and cooling processes that govern ferroelectricity in FETs. Our simulation reveals that the lower discrepancy between orthorhombic (o-) phase and tetragonal (t-) phase of interfacial energies, together with in-plane tensile stress, will improve t-o transition efficiency in the cooling down process, leading more ferroelectricity in NCFETs. Finally, design methodology to maintain the electric variation of NCFETs is also proposed.

T17-3 - 14:50
We demonstrated that the 2.5nm-thick HfAlO$_2$, N-type NCFET based on defect-passivated multidomain switching can achieve a minimum 9 mV/dec subthreshold swing (SS), a negligible hysteresis of 1mV, an ultralow $I_{on}$ of 135 fA/µm, a large $I_{on}/I_{off}$ ratio of 8.7×10$^7$ and a sub-60 mV/dec SS over 5 decade. For P-type NCFET, the non-hysteric steep-slope switch is still reached under the synergistic effect of gate stress, defect passivation and doping engineering. The Al doping and defect passivation play the key role for reducing trap-related leakage, enhancing NC, and stabilizing multidomain switching. The highly scaled HfAlO$_2$, CMOS NCFET shows the potential for low power logic applications.

**T17-4 - 15:15**

**Microscopic Crystal Phase Inspired Modeling of Zr Concentration Effects in Hf$_{1-x}$Zr$_x$O Thin Films**, A. K. Saha*, B. Grisafe**, S. Datta*** and S. Gupta*, *Purdue Univ. and **Univ. of Notre Dame, USA

In this paper, we theoretically and experimentally investigate the Zr concentration dependent crystal phase transition of Hf$_{1-x}$Zr$_x$O (HZO) and the corresponding evolution of dielectric (DE), ferroelectric (FE) and anti-ferroelectric (AFE) characteristics. Providing the microscopic insights of strain induced crystal phase transformations, we propose a physics based model that shows good agreement with our experimental results for 10nm Hf$_{1-x}$Zr$_x$O (with x=0 through 1). Utilizing our model, we analyze HZO-FET operation as a nonvolatile memory device for different x.

**SESSION 18**

**ReRAM & Selector [Shunju II, III]**

**Chairpersons:** S.-W. Chung, SK Hynix Semiconductor Ltd.

**F. Arnaud, ST Microelectronics**

**T18-1 - 16:00**


This paper presents key specifications of RRAM-based non-volatile memory embedded into Intel 22FFL FinFET Technology. 22FFL is a high performance, ultra low power technology developed for mobile and RF applications providing extensive high voltage and analog support and high design flexibility combined with low manufacturing costs. Embedded RRAM technology presented in this paper achieves 10$^8$ cycle endurance combined with 85°C 10-year retention and high die yield. Technology data retention, endurance and yield are demonstrated on 7.2Mbit arrays. We describe device characteristics, bit cell integration into the logic flow, as well as key considerations for achieving high endurance and retention properties.

**T18-2 - 16:25**


This work proposes (1) an auto-forming (AF) scheme to shorten the macro forming time ($T_{FORM}$) and testing costs; (2) an auto-RESET (ARST) scheme to shorten page-RESET time ($T_{PAGE,RESET}$) for expanding the applications of hidden-RESET operation in standby mode, and (3) an auto-SET (ASET) scheme to shorten page-write time ($T_{PAGE}$) combined with hidden-RESET scheme. A fabricated 40nm 2Mb ReRAM macro achieved 85% reduction in $T_{FORM}$, and 99+% reduction in $T_{PAGE,RESET}$ for a page. For the first time, AF, ARST, and ASET schemes are demonstrated in silicon for ReRAM.

**T18-3 - 16:50**

**Application-Induced Cell Reliability Variability-Aware Approximate Computing in TaO$_x$-Based ReRAM Data Center Storage for Machine Learning**, C. Matsui, S. Fukuyama, A. Hayakawa and K. Takeuchi, Chuo Univ., Japan

This paper proposes Variability-Aware Approximate Computing (V-AC) for TaO ReRAM storage at data centers with 6 design strategies. For the first time, this paper shows that application-induced variability degrades the performance. To solve this problem, V-AC fully utilizes error resilience of machine learning (ML) applications and decreases bit-error rate of typical cells by removing extra data copy and enlarging BER difference among memory cells. By combining device measurement and system emulations, this paper realizes system, circuit and device co-design (SCDC). Proposed SCDC platforms combines device measurement and transaction-level-modeling system emulator. V-AC is the key enabling technology to push the limits of performance, power, chip size and scaling of ReRAM for ML. Performance, energy and cell area of ReRAM storage improves by 7.0 times, 90% and 8.5%, respectively. Moreover, lower 20% Set/Reset voltage reduces driving current specification of selector devices and will extends the scaling of future ReRAM.
In this work, we demonstrate a high performance Nb$_3$O$_x$ based selector with thermal feedback mechanism for 3D X-point application. Ultra-high endurance (>10$^{12}$), high operation speed (10ns), bidirectional operation and excellent $V_{th}$ stability were achieved. By adding a barrier layer between Nb$_3$O$_x$ film and electrode, the off-state leakage current was reduced by one order of magnitude (selectivity as high as 500). This work provides a universal selector solution for various emerging memories, including RRAM, MRAM and PCM.

Chairpersons: Y. Shiratori, NTT Corp. P. Grudowski, NXP

T19-1 - 16:00
GaN HEMTs with Breakdown Voltage of 2200 V Realized on a 200 mm GaN-on-Insulator(GNOI)-on-Si Wafer, Z. LIU*, H. Xie*, K. H. Lee*, C. S. Tan**, G. I. Ng** and E. A. Fitzgerald**, *Singapore-MIT Alliance for Research and Technology (SMART), **Nanyang Technological Univ., Singapore and ***Massachusetts Institute of Technology, USA

GaN-on-Si has revealed its great potential for next-generation power electronics applications, however, there remains a challenge in increasing the breakdown voltage ($B_{V_{th}}$) due to the limit of the GaN epi-layer thickness on large size wafers. In this work we propose a GaN-on-Insulator (GNOI)-on-Si structure to address this issue. A 200 mm GNOI-on-Si wafer was prepared through removing the original Si substrate of a GaN-on-Si wafer and bonding onto a fresh SiO$_2$/Si substrate. HEMTs were fabricated on SiO$_2$/Si substrate using channel area-selective CVD growth. Smooth and uniform WS$_x$/Si substrate. HEMTs comprising approximately 6 layers was formed by area-selective CVD growth in which a patterned tungsten-source/drain served as the seed for WS$_x$ growth. For a 40 nm gate length transistor, the device has impressive electrical characteristics: on/off ratio of ~10$^{6}$, a S.S. of ~97 mV/dec., and nearly zero DIBL.

Area-selective channel material growth for 2D transistors is more desirable for volume manufacturing than exfoliation or wet/dry transfer after large area growth. We demonstrate the first top-gate WS$_x$p-channel field-effect transistors (p-FETs) fabricated on SiO$_2$/Si substrate using channel area-selective CVD growth. Smooth and uniform WS$_x$ comprising approximately 6 layers was formed by area-selective CVD growth in which a patterned tungsten-source/drain served as the seed for WS$_x$ growth. For a 40 nm gate length transistor, the device has impressive electrical characteristics: on/off ratio of ~10$^{6}$, a S.S. of ~97 mV/dec., and nearly zero DIBL.

Reassessing InGaAs for Logic: Mobility Extraction in Sub-10nm Fin-Width FinFETs, X. Cai*, A. Vardi*, J. Grajal** and J. del Alamo*, *Massachusetts Institute of Technology, USA and **Universidad Politecnica de Madrid, Spain

We study the performance degradation of InGaAs FinFETs as they scale to sub-10 nm fin width. This is often attributed to degradation in intrinsic transport parameters. High frequency measurements, however, indicate increasingly severe oxide trapping as the fin width narrows. This is confirmed by pulsed-IV measurements. A new mobility extraction method using concurrent S-parameter and DC-IV measurements avoids the impact of oxide trapping and reveals promising mobility in thin-channel InGaAs planar MOSFETs and narrow-width FinFETs. Our study suggests that performance degradation of InGaAs FinFETs is largely an extrinsic phenomenon that can be engineered around and that the potential performance of deeply-scaled InGaAs FinFETs is significantly underestimated.

Multicolor photodetectors (PDs) by using bulk p-i-n based visible GaAs and near-infrared (IR) InGaAs PD was successfully fabricated via monolithic integration by wafer bonding and epitaxial lift-off. It showed high-performance individual operation comparable to that of bulk PDs with tight vertical alignment on a single substrate for future high-resolution multicolor PDs. At the same time, it covered a broad wavelength range from visible to IR.

Friday Forum
Enabling Technologies for Autonomous Driving [Suzaku I, II, III]
Friday, June 14, 9:00-15:35

Organizers:  T. Tanaka, Tohoku Univ.
K. Benaissa, Texas Instruments Inc.
Y. Oike, Sony Corp.
R. Kapusta, Analog Devices, Inc.

Moderator:  K. Nakamura, Analog Devices, Inc.

9:00 Opening

9:10 Inertial and Depth Sensors for Autonomous Vehicles, R. Kapusta, Analog Devices, Inc.
The pathway to achieving full autonomy begins with a cross-functional, system-level approach that surrounds the vehicle with a real-time, 360-degree safety shield. This shield is created by fusing data from high-performance inertial sensors, RADAR, and LiDAR with other sensor outputs, which ultimately gives the vehicle its ability to accurately perceive the road around it. While RADAR and LiDAR remain a premium feature in autonomous vehicles currently, they must become a standard fitment to address safety concerns.

In this talk, we will look at the sensor framework for the fully autonomous vehicle. The need for significantly improved performance and response time will be explored, along with opportunities to exploit the complementary nature of various sensors through fusion of their outputs.

9:55 Safety and Security at the Heart of Autonomous Driving, K. Khouri, NXP Semiconductors
Automotive safety and security are not only pivotal to market acceptance of autonomous vehicles, but a required rite of passage for any automotive supplier. At NXP, safety and security are part of our DNA; we have deep knowhow on these subjects and our safety & security culture is deeply embedded within the company. This means that at every stage of the design and development process, we are implementing industry best practices, complemented by NXP's unique experience and knowhow in safety & security to deliver state-of-the-art security and safety solutions. Making it easier for our customers to comply with industry-wide requirements and standards for safety and security by delivering documentation, tools and support.

10:40 Break

11:00 Electronics Technologies Evolve Automobiles!?, N. Kawahara, DENSO Corp.
Automotive electronics have been evolving and creating new control systems to realize safer and more eco-friendly vehicles. Many automotive functions are changing from mechanical to electronic control. By changing the control systems, the number of electronic parts such as sensors, electronic circuits, and actuators has been drastically increasing. And this trend will continue in the future to evolve automobiles. MEMS technologies, along with the packaging, electronic circuit, and software technologies, will become more important in the future vehicle equipped with many advanced sensors. Undoubtedly, the control system becomes more advanced with each improvement in the sensing speed or sensor accuracy.

In the presentation, the future trend of automobiles and Electronics will be discussed.

Human vision is the most essential sensor to drive vehicle. Instead of human eyes, CMOS image sensor is the best sensing device to recognize objects and environment around the vehicle. Image sensors are also used in various use cases such as driver and passenger monitor in cabin of vehicle. For these use cases, some special functionalities and specification are needed. In this session the requirements for automotive image sensor will be discussed such as high dynamic range, flicker mitigation and low noise. In the last part the key technology to utilize image sensor, such as image recognition and computer vision will be discussed.

12:30 Lunch
13:30 **The Advent of the GPU in AI/Supercomputing and its Application to Autonomous Driving**, T. Baji, NVIDIA Corp.

In the old good days, CPU performance increased almost 1.5 times / year thanks to the Moor’s Law. However by the year 2010, due to the leakage current and too complex CPU architecture, this rate becomes 1.1 times / year. On the other hand, parallel processing dedicated GPU continues to grow its performance with the rate of 1.5 times / year, and even with the Moor’s Law ending, it still continues to grow its performance by built-in accelerators. Now GPU is the most widely used accelerator in AI and Supercomputing. This GPU architecture is also applied to the most advanced autonomous driving SoC Xavier.

In this talk, GPU technologies which realize this high performance, the autonomous driving platform based on this GPU and Xavier SoC, and the end-to-end system solution that enables its functional safety and reliability will be introduced.

14:15 **Envisioning Smart Mobility Society in the Connected Future**, T. Imai, Toyota Motor Corp.

The automotive industry is changing faster today than it has in 100 years and must reconsider what our society and customers expect from us – as automotive companies. It is not only a shift from a car manufacturing & sales company to a mobility company but also a convergence of electrification, connectivity and artificial intelligence. With these exciting advances, it is our mission to provide new mobility society.

The main objectives of this session are: (1) the current state of vehicle connectivity, showing connected vehicles in Japan and how to utilize big data, and (2) our vision of the smart mobility society of the future, which is the key to realize seamless and comfortable transportation through connected vehicles with the Vehicle Control Interface and the Mobility Service Platform (MSPF).

15:05 **Panel Discussion**

15:35 **Closing**

---

Friday Evening Event [Taizo-in]

Friday, June 14, 16:15-19:35