2021 Symposium on VLSI Technology

Short Course 1 (Technology)

Advanced Process and Device Technology Toward 2nm-CMOS and Emerging Memory

Moderators: K. Endo, AIST
S. Datta, Univ. of Notre Dame

Live Q&A Session **Monday, June 14, 7:00-8:30 (JST)**

CMOS Device Technology for the Next Decade, J. Cai, TSMC

Nanosheet Device Architectures to Enable CMOS Scaling in 3nm and Beyond: Nanosheet, Forksheet and CFET, N. Horiguchi, imec

Extension of Cu Interconnects and Considerations for Post-Cu Alternative Metals in Advanced Nodes, K. Motoyama, IBM Corp.

Contact Module Engineering for Advanced CMOS Technologies: Key Concepts, Engineering Techniques and Device Integration Challenges, N. Breil, Applied Materials, Inc.

Metrology Challenges Towards 2nm Technology Node, M. Ikoda, Hitachi High-Tech Corp.

Emerging Memories and the Applications, H.-T. Lue, Macronix International Co., Ltd.

Key Device Technologies and Challenges for 3D Emerging Non-Volatile Memory, M. Saitoh, KIOXIA Corp.

DRAM: Challenges and Opportunities, K. Hamada, Micron Memory Japan, G.K.

Short Course 2 (Joint)

Enabling a Future of Even More Powerful Computing

Moderators: K. Yoshioka, Toshiba Corp.
S. H. Kang, Qualcomm

Live Q&A Session **Monday, June 14, 7:00-8:30 (JST)**

Acceleration of Tomorrow's Computational Challenges, G. Loh, Advanced Micro Devices, Inc.

3D-Structured Monolithic and Heterogeneous Devices for Post-5G System Applications, Y. Hayashi, Keio Univ.

Accelerated Computing: Latest Advances and Future Challenges, B. Keller, NVIDIA Corp.

Deep Learning Supercomputers - From Compilers to Custom Silicon, G. Watanabe, Preferred Networks

Hardware for Next Generation AI, D. Nikonov and A. Khosrowshahi, Intel Labs

Re-Engineering Computing with Neuro-Inspired Learning: Algorithms, Hardware Architecture, and Devices, K. Roy, Purdue Univ.

Quantum Computing with Superconducting Circuits, M. Brink, IBM Corp.

Digital Annealing Technology for Solving Combinatorial Optimization Problems, K. Kanda, Fujitsu

Short Course 3 (Circuits)

Advanced Circuits and Systems for Internet-of-Things (IoT) Sensors

Moderators: P.-H. Cheng, NCTU
D. Griffith, TI

Live Q&A Session Monday, June 14, 7:00-8:30 (JST)

CMOS Sensor for IoT: First Frontier, S. Pietri, NXP Semiconductors N.V.

Non-CMOS Based Sesors for IoT, M. Zevenbergen, imec

Capacitive Power Management Circuits for Miniaturized Energy Harvesting IoT Systems, M.-K. Law, Univ. of Macau

Getting the Most Out of a Little: Ultra-Low Power Circuit Techniques for the IoT, D. A. Hall, Univ. of California, San Diego

Low Power and Energy-Efficient Digital CMOS for Mixed-Signal Sensor Interfaces,
S. Bampi, Federal Univ. of Rio Grande do Sul

IC-Chip Level Physical Attack Protections for IoT Security, M. Nagata, Kobe Univ.

Image Sensor Technologies for Computer Vision Systems to Realize Smart Sensing,
A. Nose, Sony Semiconductor Solutions Corp.

Design Considerations for Battery-Free and Crystal-Less Bluetooth-LE Sensors for Low-Cost Labels,
A. Yehezkely, Wiliot Ltd.

Forum

Technologies for Post COVID-19 Era

Moderators: C. V. Hoof, imec

Live Q&A Session Saturday, June 19, 7:00-8:30 (JST)

The Growing Importance of Indoor Air Quality Monitoring: How CO2 Sensors Help to Reduce the Risk of Covid-19 Transmission, L. Beaurenaut, Infineon Technologies AG

Going from a 5G/6G Vision to Real Implementation, F. Tillman, Ericsson

Digital Annealer: Technology for Solving Combinatorial Optimization Problems in Real World, J. Koyama, Fujitsu

Challenges and Opportunities for Sub-7nm In-Memory/Near-Memory Computing, AI Accelerators and Hardware Security, R. Krishnamurthy, Intel Labs

NTT DOCOMO's View on 5G Evolution and 6G, T. Asai, Docomo

Data Analytics Approach for Smart Manufacturing and Optimizing Equipment Condition, T. Moriya, Tokyo Electron Ltd.

SESSION 1

Opening and Plenary Session 1 [Room 1]

Tuesday, June 15, 6:50-8:50

6:50-
Opening Remarks
S. Yamakawa, Sony Semiconductor Solutions Corp.
K. Takeuchi, The Univ. of Tokyo

Chairpersons: Y. Oike, Sony Semiconductor Solutions Corp.
B. Nikolić, Univ. of California, Berkeley

C1-1 - 7:20 (Plenary)
Fugaku and A64FX: the First Exascale Supercomputer and its Innovative Arm CPU, S. Matsuoka, Riken, Japan

Fugaku is the first exascale supercomputer in the world, designed and built primarily by Riken Center for Computational Science (R-CCS) and Fujitsu Ltd., but involving essentially all the major stakeholders in the Japanese HPC community. The name ‘Fugaku’ is an alternative name for Mt. Fuji, and was chosen to signify that the machine not only seeks very high performance, but also a broad base of users and applicability at the same time. The heart of Fugaku is the new Fujitsu A64FX Arm processor, which is 100% compliant to Aarch64 specifications, yet embodies technologies realized for the first time in a major server general-purpose CPU, such as 7nm process technology, on-package integrated HBM2 and terabyte-class SVE streaming capabilities, on-die embedded TOFU-D high-performance network including the network switch, and adoption of so-called ‘disaggregated architecture’ that allows separation and arbitrary combination of CPU core, memory, and network functions. Fugaku uses 158,974 A64FX CPUs in a single socket node configuration, making it the largest and fastest supercomputer ever created, signified by its groundbreaking achievements in major HPC benchmarks, as well as producing societal results in COVID-19 applications.

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

T1-1 - 8:05 (Plenary)
Materials to Systems in Semiconductor Manufacturing and Beyond, O. Nalamasu, Applied Materials, Inc. and Appiked Ventures, LLC

SESSION 2

Highlight [Room 4]

Tuesday, June 15, 9:20-10:10

Chairpersons: T. Tsunomura, Tokyo Electron Ltd.
N. Ramaswamy, Micron Technology, Inc.

T2-1 - 9:20
Forksheet FETs for Advanced CMOS Scaling: Forksheet-Nanosheet Co-Integration and Dual Work Function Metal Gates at 17nm N-P Space, H. Mertens, R. Ritzenthaler, Y. Oniki, B. T. Chan, B. Briggs, A. Hikavyy, T. Hopf, G. Mannaert, Z. Tao, F. Sebaai, A. Peter, K. Vandersmissen, E. Dupuy, E. Rosseel, D. Batuk, J. Geypen, G. T. Martinez, D. Abigail, E. Grieten, K. Dehave, J. Mitard, S. Subramanian, L.-A. Ragnarsson, P. Weckx, D. Jang, B. Chehab, G. Hellings, J. Ryckaert, E. Dentoni Litta and N. Horiguchi, imec, Belgium

We report on forksheet N- and PFETs co-integrated with gate-all-around nanosheet FETs. The forksheet short-channel control is on par with nanosheets down to 22nm gate length (SS_{SAT}=66-68mV/dec). Forksheet I_{ON} and I_{OFF} characteristics are improved by post-channel-release wet clean optimization, attributed to gate stack interface trap density reduction. Dual work function metal gates are integrated at 17nm N-P space, highlighting a key benefit of forksheets for CMOS area scaling.

T2-2 - 9:30
Highly Manufacturable 7th Generation 3D NAND Flash Memory with COP Structure and Double Stack Process, J. H. Kim, Y. Yim, J. Lim, H. S. Kim, E. S. Cho, C. Yeo, W. Lee, B. You, B. Lee, M. Kang, W. Jang, Y. Kwon, K. Lee, J. Lee, M.-C. Kim, J.-Y. Lee, S. Hur, S. J. Ahn, H. Hong, Y. G. Shin, H.-S. Kim and J. Song, Samsung Electronics Co., Ltd., Korea

A novel 3D NAND Flash memory device with 17X WL (Word line) layers has been successfully developed. COP(Cell Over Peripheral) Structure has been applied, improving tR and tPROG by 11% and 20%, respectively. Compared with our previous product(6th generation), the bit density is increased by 70% through cell volume scaling and COP structure. Several breakthrough processes have been successfully combined to achieve this new structure. Double stack process, low stress W(tungsten), MSE(Multi Step Etch), and channel hole side-wall butting etc. In addition, the double stack process was applied to significantly improve the channel hole profile. As a result, better cell operation characteristics were achieved.

T2-3 - 9:40
Advancing Monolayer 2D NMOS and PMOS Transistor Integration From Growth to Van Der Waals Interface Engineering for Ultimate CMOS Scaling, C. Dorow, K. O'Brien, C. Naylor, S. Lee, A. Penumatcha, A. Hsiao, T. Tronic, M. Christenson, K. Maxey, H. Zhu, A. Oni, U. Alaan, T. Gosavi, A. Sen Gupta, R. Bristol, S. Clendenning, M. Metz and U. Avci, Intel Corp., USA

2D-material channels enable ultimate scaling of MOSFET transistors and will help Moore's Law Scaling for decades. We demonstrate the state of both n and p-MOSFETs using monolayer TMD channels of sub-1nm thickness and manufacturable CVD, MBE or seeded growth. NMOS devices on transferred MBE MoS_2 using novel contact metal show low variation, one of the lowest reported contact resistances (R_c) of 0.4 kΩ.um, low hysteresis, and good subthreshold slope (SS) of 77 mV/dec. PMOS devices using CVD WSe_2 show 89 mV/dec SS, best reported for PMOS on grown films, but on-current remains behind NMOS. Transfer-free, area-selective WS_2 transistors achieve 10 μA/μm on-current, highest reported on WS_2 using seeded growth. A new capacitance method is shown to monitor 2D material contact interface quality. Gate-oxide interface engineering through metal seeding and ALD demonstrates that a single 2D channel material can selectively make PMOS or NMOS transistors, alike Si CMOS.

T2-4 - 9:50
First Demonstration of Atomic-Layer-Deposited BEOL-Compatible In_2O_3 3D Fin Transistors and Integrated Circuits: High Mobility of 113 cm^2/Vs, Maximum Drain Current of 2.5 mA/μm and Maximum Voltage Gain of 38 V/V in In_2O_3 Inverter, M. Si, Z. Lin, Z. Chen and P. D. Ye, Purdue Univ., USA

In this work, we report the first demonstration of In_2O_3 3D transistors coated on fin-structures and integrated circuits by a back-end-of-line (BEOL) compatible atomic layer deposition (ALD) process. High performance planar In_2O_3 transistors with high mobility of 113 cm^2/Vs and record high maximum drain current of 2.5 mA/um are achieved by channel thickness engineering and post-deposition annealing. High-performance ALD In_2O_3 based zero-V_{GS}-load inverter is demonstrated with maximum voltage gain of 38 V/V and minimum supply voltage (V_{DD}) down to 0.5 V. ALD In_2O_3 3D Fin transistors are also demonstrated, benefiting from the conformal deposition capability of ALD. These results suggest ALD oxide semiconductors and devices have unique advantages and are promising toward BEOL-compatible monolithic 3D integration for 3D integrated circuits.

T2-5 - 10:00
3D Stacked CIS Compatible 40nm Embedded STT-MRAM for Buffer Memory, M. Oka*, Y. Namba*, Y. Sato*, H. Uchida*, T. Doi*, T. Tatsuno*, M. Nakazawa*, A. Tamura*, R. Haga*, M. Kuroda*, M. Hosomi*, K. Suemitsu**, E. Kariyada**, T. Suzuki**, H. Tanigawa**, M. Ueki**, M. Moritoki**, Y. Takegawa**, K. Bessho* and T. Umebayashi*, *Sony Semiconductor Solutions Corp. and **Sony Semiconductor Manufacturing Corp., Japan

This paper presents the world's first demonstration of a 40nm embedded STT-MRAM for buffer memory, which is compatible with the 3D stacked CMOS image sensor (CIS) process. We optimized a CoFeB-based perpendicular magnetic tunnel junction (p-MTJ) to suppress the degradation of magnetic properties caused by the 3D stacked wafer process. With improved processes, we achieved high speed write operation below 40 ns under typical operation voltage conditions, endurance up to 1E+10 cycles and 1 s data retention required for a buffer memory. In addition, to broaden the application of embedded MRAM (eMRAM), we proposed a novel fusion technology that integrated embedded non-volatile memory (eNVM) and buffer memory type embedded MRAM in the same chip. We achieved a data retention of 1 s ~ >10 years with a sufficient write margin using the fusion technology.

SESSION 3

Future Logic Devices [Room 5]

Tuesday, June 15, 9:20-10:00

Chairpersons: A. V.-Y. Thean, National Univ. of Singapore
S. Datta, Univ. of Notre Dame

T3-1 - 9:20
Scaling Synthetic WS2 Dual-Gate MOS Devices Towards Sub-Nm CET, D. Lin, imec, Belgium

We present a gate stack scaling study on dual gate (DG) WS_2 transistors with scaled back gate (BG) and top gate (TG) stacks using an ALD physisorption-seeding approach. DG MOSFET with a 2ML WS_2 and 100nm channel reaches 310uA/um drain current, 320μS/μm max. transconductance at 1V Vd and sub-threshold swings of 69mV/dec and 116mV/dec at 0.1V and 1V Vd, respectively. With single charge centroid assumption, a 0.78nm DG and a 1.92nm TG CET are deduced. Statistics of 3000 FET demonstrates the performance trend and potential of EOT scaling on MX_2 MOSFET.

T3-2 - 9:30
Record RF Performance (f_t=180GHz and f_{max}=240GHz) of a FDSOI NMOS Processed within a Low Thermal Budget for 3D Sequential Integration, T. Mota Frutuoso*, J. Lugo-Alvarez*, G. Xavier*, L. Brunet*, C. Mikael*, C. Fenouillet-Beranger*, G. Romano*, T. Januel*, D. Cooper*, V. Loup*, J. Kanyandekwe*, P. Acosta-Alba*, S. Kerdiles*, A. Tavernier*, L. Brevard*, P. Ferrari**, F. Andrieu* and F. Gaillard*, *CEA-LETI and **RFIC-Lab, France

Record RF Figure-Of-Merits (FoM) is highlighted for a 42nm NMOS transistor fully processed at Low Thermal Budget (LTB) (<500°C) needed for 3D Sequential Integration (3DSI). *f_T=180GHz and f_{MAX}=240GHz are reported at V_{DD}=0.9V*; which is actually very similar to performance of reference Si MOSfets processed with a Hot Thermal Budget (HTB) (Fig. 15). This excellent result was possible thanks to a careful optimization of the LTB process after an advanced characterization and modeling of Key technological parameters such as mobility, Gate-Capacitance and Gate resistance

T3-3 - 9:40
Analog Monolayer MoS_2 Transistor with Record-High Intrinsic Gain (>100 dB) and Ultra-Low Saturation Voltage (<0.1 V) by Source Engineering, M. Liu*, C. Lu*, G. Yang*, W. Gan*, S. Peng*, Z. Wu*, J. Niu*, J. Wang*, L. Wang*, M. Li*, D. Geng*, N. Lu*, W. Cao**, L. Li*, D. Akinwande*** and M. Liu*, *Chinese Academy of Sciences, China, **Univ. of California, Santa Barbara and ***The Univ. of Texas at Austin, USA

For the first time, we demonstrate an excellent analog performance monolayer MoS_2 transistor with Schottky-source gated structure operating in sub-threshold region. By precise controlling the source injection barrier, this device exhibits pinchoff regions occurring both at source and drain with ultra-high output impedance (>10^{13} Ω) and ultra-low saturation voltage (<0.1 V). By this design, record-high intrinsic gain (>100 dB) is achieved in this device among all the transistors reported so far. This work provides a universal high-performance transistor solution for low power analog application.

T3-4 - 9:50
Impact of Asymmetric Strain on Performance of Extremely-Thin Body (100) GOI and (110) SGOI pMOSFETs, C.-T. Chen*, R. Yokogawa**, K. Toprasertpong*, A. Ogura**, M. Takenaka* and S. Takagi*, *The Univ. of Tokyo and **Meiji Univ., Japan

We demonstrate high performance asymmetric (quasi-uniaxial) strain (100) Ge-on-insulator (GOI) pFETs with body thickness ranging from 10.6 to 3.8 nm and (110) SiGe-on-insulator (SGOI) pFETs with body thickness down to 6.4 nm, for the first time, by combining Ge condensation with channel width narrowing. Thanks to quasi-uniaxial strain, effective hole mobility of 1117 and 312 cm^2/Vs is achieved on 10.6- and 3.8-nm-thick GOI, leading to 2.4x enhancement in 3.8 nm against biaxial strain. For (110) SGOI, record high effective hole mobility of 807 cm^2/Vs is observed on 6.4-nm-thick quasi-uniaxial $Si_{0.45}Ge_{0.55}OI$ pFETs, corresponding to 2x enhancement against biaxial strain and 1.7x against quasi-uniaxial (100) GOI with similar thickness. These values of effective hole mobility in body thickness thinner than 10 nm are record-high ones, indicating the effectiveness of uniaxial strain on extremely-thin channels.

T3-5 Live Q&A Session : June 17, 15:10 (JST)
Record Performance of 500°C Low-Temperature nMOSFETs for 3D Sequential Integration Using a Smart Cut™ Layer Transfer Module, L. Brunet*, S. Reboh*, T. Januel*, X. Garros*, T. Mota Frutuoso*, M. Casse*, B. Sklenard*, L. Brevard*, M. Ribotta*, A. Magalhaes-Lucas*, J. Kanyandekwe*, J.-M. Hartmann*, F. Milesi*, F. Mazen*, P. Acosta-Alba*, S. Kerdiles*, A. Tavernier*, V. Loup*, C. Morales*, V. Larrey*, F. Fournel*, L. Le Van-Jodin*, S. Leforestier*, E. Rolland*, G. Romano*, G. Gaudin**, J. Lugo*, J. Lacord*, S. Maitrejean*, J. Arcamone*, P. Batude*, I. Radu**, C. Fenouillet-Beranger* and F. Andrieu*, *CEA-LETI and **Soitec, France

We report record performances in Top-tier nMOSFETs fabricated by 3D sequential integration, as well as junction optimization guidelines to further optimize the performances within a maximum thermal budget of 500°C. We reached I_{ON}=870µA/µm at I_{OFF}=100nA/µm at V_{DD}=1V together with A_{VT}=1.35mV.µm and a decent PBTI lifetime.Moreover, a first generation of Low Temperature LT layer transfer module from Si substrates based on Smart Cut™ was developed to obtain LTSOI quality compatible with 500C FDSOI process integration: RMS=0.083nm roughness and 0.4nm SOI uniformity. The nMOSFETs fabricated on these LTSOI wafers reached I_{ON}-I_{OFF} performances at 88% of reference transistors integrated on regular SOI wafers.

T3-6 Live Q&A Session : June 17, 15:20 (JST)
High Yield and Process Uniformity for 300 mm Integrated WS_2 FETs, T. Schram, Q. Smets, D. Radisic, B. Groven, D. Cott, A. Thiam, W. Li, E. Dupuy, K. Vandersmissen, T. Maurice, I. Asselberghs and I. Radu, imec, Belgium

We demonstrate an integrated process flow on full 300mm wafers with monolayer WS_2 channel. WS_2 is a 2D semiconductor from the transition metal dichalcogenide family and holds promise for extreme gate length scaling. We report here on integration challenges and optimize process uniformity for a single-device yield higher than 90% across wafer. These transistors and integration flow are shown to be compatible with H2 anneal of at least 400 C and hence in principle suitable for hybrid integration in the BEOL.

SESSION 4

Award and Plenary Session 2 [Room 1]

Wednesday, June 16, 6:40-8:50

6:40-
Award etc.

Chairpersons: B. Nikolić, Univ. of California, Berkeley
Y. Oike, Sony Semiconductor Solutions Corp.

C4-1 - 7:20 (Plenary)
A New Era of Tailored Computing, M. Papermaster, S. Kosonocky, G. H. Loh and S. Naffziger, Advanced Micro Devices, Inc. (AMD)

The worldwide computing market grew tremendously over the past decades, and looking toward the future, these trends do not appear to be slowing down. Moore's Law coupled with incredible innovation in hardware and software are engines driving this growth. However, the entire industry faces a barrage of challenges including the slowing of Moore's Law, stringent power and energy constraints, an always-connected society, and disruptions from the on-going artificial intelligence revolution. To continue delivering ever higher-performance computing solutions amid these difficulties, the industry needs to pivot to a new mindset of "Tailored Computing." The need for and opportunities to tailor our technologies in all aspects of future compute will propel the industry toward heterogeneity in everything it does.

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

T4-1 - 8:05 (Plenary)
Pandemic Challenges, Technology Answers, S. Choi, Samsung Electronics Co., Ltd., Korea

As the global community was caught off guard with the pandemic, semiconductor industry dealt with unexpected swings in applications and demands. Health crisis created an immediate need for social distancing that disconnected and disrupted human interactions, and technology had to step up on short notice to mend and reconnect communities. In this paper, we share the insights we gained as the semiconductor technologists who were called to provide solutions in a nimble and yet comprehensive manner to deal with the unexpected, and offer our vision and new model for the foundries, not just as the manufacturers, but as solution providers. The new market reality dominated by “untact” and “connect” demand differentiated strategies in providing foundry solutions, which include close engagement with customers in earlier stages of technology R&D, as well as design infrastructure tailored to customers’ specific requirements. We present our vision to drive such change in foundry technology directions.

Technology Focus Session 1

Advanced Memory Technology [Room 4]

Wednesday, June 16, 9:20-10:00

Chairpersons: D. Kil, SK hynix Semiconductor Ltd.
G. Bronner, Rambus, Inc.

TFS1-1 - 9:20 (Invited)
3-Dimensional Integration of Epitaxial Magnetic Tunnel Junctions with New Materials for Future MRAM, K. Yakushiji, H. Takagi, N. Watanabe, A. Fukushima, K. Kikuchi, Y. Kurashima, A. Sugihara, H. Kubota, and S. Yuasa, AIST, Japan

We fabricated novel epitaxial magnetic tunnel junction (MTJ) films with single-crystal spinel $MgAl_2O_4$(001) tunnel barrier on Φ300 mm Si(001) wafers by sputtering deposition and demonstrated 3-dimensional (3D) integration of the epitaxial MTJ nano-pillars in STT-MRAM by using direct bonding of the epitaxial MTJ wafer and CMOS wafer. The epitaxial MTJs exhibit smaller bit-to-bit variations of magnetic and transport properties compared with poly-crystalline CoFeB/MgO/CoFeB MTJs due to the absence of crystal grains. The 3D integration of epitaxial MTJs allows us to replace conventional MTJ materials (MgO and CoFeB) with new dielectric and ferromagnetic materials, giving great flexibilities and scalability to future MRAM technology. The developed technologies will be applicable also to other kinds of tunnel junction devices such as the Josephson junctions in superconducting qubit for quantum computing.

TFS1-2 - 9:30
Mg-Te OTS Selector with Low I_{off} (< 100 pA), Fast Switching Speed (td = 7 ns), and High Thermal Stability (400 °C / 30min) for X-Point Memory Applications, S. Lee, J. Lee, S. Kim, D. Lee, D. Lee and H. Hwang, POSTECH, Korea

We report on nanoscale (d = 45 nm), binary Mg-Te based ovonic threshold switching (OTS) selector with low leakage current (I_{off} = 88 pA), high threshold voltage (V_{th} = 2.4V/10 nm), fast switching speed (t_d = 7 ns) and high thermal stability (400 °C/30 min). We found that OTS selector parameters (I_{off} and V_{th}) are closely related to the activation energy (E_a) of the Poole-Frenkel conduction model which can be controlled by varying the composition ratio of Mg and Te. The best OTS device characteristics such as large E_a (~ 0.7 eV), lowest I_{off} and highest V_{th} can be obtained by adopting the optimum composition of $Mg_{0.5}Te_{0.5}$.

TFS1-3 - 9:40
Cold MRAM as a Density Booster for Embedded NVM in Advanced Technology, H. L. Chiang*,**, J. F. Wang*,***, T. C. Chen*, T. W. Chiang*, C. Bair*, C. Y. Tan*, L. J. Huang*, H. W. Yang*, J. H. Chuang*, H. Y. Lee*, K. Chiang*, S. K. H.*, Y. J. Lee*, R. Wang*, C. W. Liu***, T. Wang**, X. Bao*, E. Wang*, J. Cai*, C. T. Lin*, H. Chuang*, H.-S. P. Wong* and M. F. Chang*, *TSMC, **National Chiao Tung Univ. and ***National Taiwan Univ., Taiwan

With the co-optimization of magnetic tunnel junctions and transistors, high-density 16nm-CMOS-compatible STT-MRAM designed at 77K (Cold MRAM) is proposed in this work to increase the cell density to 5.3x of a 16nm SRAM cell. A higher thermal stability factor is leveraged to scale MTJs' size for write current and further reduce the size of the access transistors dominating the cell size. An improved tunneling magnetoresistance ratio at 77K also enables a higher read margin for Cold MRAM.

TFS1-4 - 9:50
The First Embedded 14nm FeFinFET NVM: $2T1C_{FE}$ Array as Electrical Synapses and Activations for High-Performance and Low-Power Inference Accelerators, E. R. Hsieh*, T. W. L.***, Y. L. Lin**, C. H. Liu***, S. S. Chung**, Y. T. Tang****, J. R. Su****, T. P. Chen*****, S. A. Huang*****, T. J. Chen***** and O. Cheng*****, *National Central Univ., **National Chiao Tung Univ., ***National Taiwan Normal Univ., ****Taiwan Semiconductor Research Institute (TSRI) and *****United Microelectronics Corp. (UMC), Taiwan

We present the first embedded FeFinFET NVMs, 2T1C. 2T1C array shows better performance than that of the conventional 1T1C array. The gradual tuning of conductance for 2T1C has been tuned linearly and symmetrically with 30000x of the window, where 8 conductance states can be stored with crystal gaps in between. Continual endurance cycles for 8 states are more than 10 million times. The retention of 8 states has been baked in 85 Celsius for more than one month. The 8 states are also passed the PGM disturbance test. Finally, this 2T1C array is as electrical synapses in a very deep neural network (DNN) with 23.8 million parameters for practice, and most importantly, the 2T1C is also as an activation, Rectified Linear Unit (ReLU). Compared to the 5T-CMOS ReLU, the 2T1C ReLU shows 70% of accuracy. 2T1C shows strong potential as key components in high-performance low-power inference-accelerators.

TFS1-5 (Invited) Live Q&A Session **: June 17, 15:00 (JST)**

Development Status of Gate-First FeFET Technology, S. Müller1, H. Zhou*, A. Benoist*, J. Ocker*, M. Noack*, G. Kuzmanov*, R. Iqbal*, D. Le Minh*, M. Ghazaryan*, V. Anjaneyamoorthi*, A. Daraghmah*, M. Mennenga*, F. Koushan*, F. Tassan*, S. Dünkel**, J. Mülle**, S. Beyer**, S. Soss*** and A. Pourkeramati*, *Ferroelectric Memory GmbH, **GLOBALFOUNDRIES Inc., Germany and ***GLOBALFOUNDRIES Inc., USA

Recent progress on FeFET gate-first technology development is presented. New characterization results from FeFET endurance degradation are shown and assigned to interfacial layer degradation. Two methods to overcome endurance degradation in terms of proper choice of device geometry or program / erase algorithms are highlighted. Moreover, statistical variation of FeFET memory states is characterized for single memory cells as well as mini arrays across wafer. This is complemented by 136 Kbit FeFET array results which demonstrate tail-to-tail separation of ~3µA which represents the basis for read-out operations below 25 ns. Latest results from FeFET variability from 180nm x 180nm as well as 72nm x 72nm memory cells is presented and a 32 Mbit macro incorporating 180nm x 180nm cells has been designed for future characterization.

Technology Focus Session 2

New Process and Material for Future Devices [Room 5]

Wednesday, June 16, 9:20-10:10

Chairpersons: H.-C. Lin, National Yang Ming Chiao Tung Univ.
B. Colombeau, Applied Materials, Inc.

TFS2-1 - 9:20
Low Temperature Atomic Hydrogen Treatment for Superior NBTI Reliability-Demonstration and Modeling Across SiO_2 IL Thicknesses from 1.8 to 0.6 nm for I/O and Core Logic, J. Franco*, J.-F. de Marneffe*, A. Vandooren*, H. Arimura*, L.-A. Ragnarsson*, D. Claes*, E. Dentoni Litta*, N. Horiguchi*, K. Croes*, D. Linten*, T. Grasser** and B. Kaczer*, *imec, Belgium and **Vienna Univ. of Technology, Austria

We demonstrate a low temperature treatment for excellent NBTI reliability, compatible with novel device architectures as nanosheets and CFETs, and integration schemes such as Sequential 3D tier stacking. Hydrogen radicals generated in a remote plasma are shown to efficiently passivate hole traps associated with the hydroxyl-E' SiO2 defects, which are particularly abundant in interfacial layers (IL) grown at reduced temperatures. We explore the process window on a 1.2nm IL, optimize the treatment for an ultra-thin 0.6nm chemical oxide IL, focusing on EOT control, and show the applicability also for 1.8nm thick IL's of relevance for I/O devices. The treatment dramatically enhances the NBTI reliability, outperforming a Foundry28 reference at thinner EOT and in a low thermal budget flow. The excellent NBTI rating of Vov>1V at 1.1nm EOT demonstrated here on Si compares well with the best-in-class reliability previously demonstrated in Si-capped Si0.45Ge0.55 pMOS.

TFS2-2 - 9:30 (Invited)
Monolithic 3D Integration of 2D Electronics Based on Two-Dimensional Solid-Phase Crystallization, C.-P. Lin*, Y.-W. Kang*, C.-P. Hsu*, H.-H. Hsu*, J.-H. Huang*, R.-F. Chen*, C.-T. Wu**, Y.-J. Lee** and T.-H. Hou*, *National Yang Ming Chiao Tung Univ. and **Taiwan Semiconductor Research Institute (TSRI), Taiwan

Monolithic 3D integration (M3D) of two-dimensional materials (2DMs) based on a device structure similar to the vertically stacked nanosheet (NS) gate-all-around (GAA) field-effect transistor (NS-GAAFET) is one of the most feasible paths for endof-roadmap logic device scaling. A novel synthesis route, 2D solid-phase crystallization (2DSPC), is presented in this paper for M3D of 2DMs. 2DSPC presents a unique opportunity for achieving wafer-level uniformity, centimeter-scale monocrystalline grain, and scalable synthesis for multiple vertical layers. We believe 2DSPC offers a promising pathway toward future cost-effective M3D-2D electronics.

TFS2-3 - 9:40 (Invited)
New Processes and Materials in Semiconductor Manufacturing

TFS2-4 - 9:50
Ultra-Low Specific Contact Resistivity (3.2×10^{-10} Ω-cm^2) of Ti_$Si_{0.5}Ge_{0.5}$ Contact Deep Insights into The Role of Interface Reaction and Ga Co-doping, H. Xu*, X. Wang*, S. Luo*, J. Zhang*, K. Han*, C. Sun*, C. Wang*, R. Khazaka**, Q. Xie**, Y. Huang***, Y. Zhou*, ***, J. He***, G. Liang* and X. Gong*, *National Univ. of Singapore, Singapore, **ASM International, Belgium and ***Southern Univ. of Science and Technology, China

We report an ultra-low specific contact resistivity down to 3.2×10^{-10} Ω-cm^2 on *in-situ* grown boron (B) and surface segregated gallium (Ga) co-doped p^+-$Si_{0.5}Ge_{0.5}$ with a high average active doping concentration (N_A) of 1.2×10^{21} cm^{-3}. Two batches of devices with 8 sets of data using ladder transmission line model (LTLM) were fabricated to confirm the accuracy. We also found, for the first time, that the co-doped Ga not only enhances N_A but also plays a vital role in achieving thermally stable Ti p^+-$Si_{0.5}Ge_{0.5}$ contacts with the thermal budget of up to 450 °C. A mechanism for the deep understanding of such phenomenon was proposed and experimentally verified.

TFS2-5 - 10:00

Enabling Next Generation CMOS by Novel EOT Scaling Module, L. Dong*, S. Hung*, M. Bevan*, P. Jamison**, J. Zhang**, H. Zhou**, B. Colombeau*, R. Curtis*, J. Devrajan*, R. Bao**, R. George*, M. s. Kim*, K. Karandikar*, M. Stolfi*, N. Loubet**, J. Swenberg*, T. Guarini*, L. Wang*, Y. Yan*, V. Banthia*, S. Kuppurao*, B. Haran** and S. Natarajan*, *Applied Materials, Inc. and **IBM Research, USA

A new integrated-process approach is introduced enabling precision control and co-optimization of advanced gate stacks delivering 1-2Å *EOT* scaling while maintaining same gate leakage level compared to a traditional flow. We demonstrated this *Tinv* oxide integrated scaling on advanced FinFET test vehicles and show > 8% I_{on}/I_{off} gain for SiGe Fin PFET / Si Fin NFET, and similar benefits are preliminarily observed on Nano-Sheet (NS) devices. This paves the way for current and next generation CMOS devices for scaling and performance improvement.

TFS2-6 Live Q&A Session : June 17, 15:30 (JST)

Enabling Logic with Backside Connectivity Via n-TSVs and Its Potential as a Scaling Booster, A. Veloso, A. Jourdain, G. Hiblot, F. Schleicher, K. Dhave, F. Sebaai, D. Radisic, R. Loo, T. Hopf, A. De Keersgieter, H. Arimura, G. Eneman, P. Favia, J. Geypen, G. Arutchelvan, A. Chasin, D. Jang, L. Nyns, E. Rosseel, A. Hikavyy, G. Mannaert, B. T. Chan, K. Devriendt, S. Demuynck, G. Van der Plas, J. Ryckaert, G. Beyer, E. Dentoni Litta, E. Beyne and N. Horiguchi, imec, Belgium

We report on scaled Si-channel finFETs (L_{gate}>=20nm, 45nm fin pitch) with backside connectivity enabled by extreme wafer thinning (several Si thicknesses under STI-oxide targeted: from ~370nm down to ~20nm) and W-filled nano-through-Si-vias (n-TSV) of various heights, after using low-temperature, wafer-to-wafer, dielectric bonding. This scheme aims at allowing decoupling signal and power networks, with reduced IR-drop predicted by moving the latter to the wafer backside. A thorough evaluation of the impact of 3D processing on device characteristics is presented, showing: 1) enhanced nmos mobility and drive currents (up to 15%); 2) for pmos, small ION loss (~3 to 10%), larger R_{ext}, with channel strain evaluation by NBD for various layouts; 3) ΔV_T~-130mV that can be recovered with an extra anneal at the end, keeping tight variability and matching control. No BTI degradation is observed, with further indication that the final anneal(s) selection can be beneficial for electrostatics and reliability improvement.

SESSION 5

Advanced Logic Technology [Room 6]

Wednesday, June 16, 9:20-10:10

Chairpersons: M. Kanda, Toshiba Electronic Devices & Storage Corp.
N. Mahalingam, Texas Instruments Inc.

T5-1 - 9:20

Super Low-Power Advanced 5nm Extended Platform Technology for Extreme Low Voltage Applications, C. Byungha, J. Jaehun, K. Sung Won, K. Young Gun, S. Kihwang, L. Sanghyun, K. Seungkwon, C. Hyungyul, C. Jeongmin, J. Sungil, C. Taehoon, Y. Masuoka, Y. Jong Mil and J. Gitae, Samsung Electronics Co., Ltd., Korea

For the first time, an extended advanced 5nm Platform is revealed for super low-power application. 5nm Low-Power Process is expected to play a key role to overcome the power management crisis due to Vdd/Vth scaling complications in FinFET node. 5nm Low-Power Process successfully achieved 5% dynamic power and 13% static power reduction, compared to 5nm baseline Platform, which is also 16% dynamic power reduction from 8nm low-power process. Simultaneously, this paper states an importance of a flexible platform adjustment to provide the best solution in each applications domain.

T5-2 - 9:30

Dual Damascene BEOL Extendibility With Cu Reflow / Selective TaN And Co/Cu Composite, P. Bhosale, IBM Corp., USA

The paper demonstrates the scalability of the dual damascene (DD) integration scheme below 28 nm pitch. We evaluate the performance of the 10 nm wide interconnects build using two process flows (i) Cu reflow with selectively deposited TaN barrier (Cu/R-TaN/SB), (ii) Cobalt/Copper composite (Co/Cu comp). These process innovations enable a significant improvement in via, signal and power line resistances. We discuss the corresponding implications towards performance in terms of signal delay, parasitic voltage, and FPG gains analysis. Our simulations show that the DD Cu interconnects formed using Cu/R-TaN/SB can enable next-generation (20-24nm pitch) low power mobile-like design solutions. The Co/Cu comp with high aspect ratio power rails provides the best performance for high-performance computing (HPC) applications.

T5-3 - 9:40

Cryogenic RF CMOS on 22nm FDSOI Platform with Record f_T=495GHz and f_{MAX}=497GHz, W. Chakraborty*, K. A. Aabrar*, J. Gomez*, R. Saligram**, A. Raychowdhury**, P. Fay* and S. Datta*, *Univ. of Notre Dame and **Georgia Institute of Technology, USA

Cryogenic DRAM at 77K is a promising high bandwidth memory option to complement cryogenic superconducting (SC) processors at 4K [1]. CMOS mixed-signal interface circuits with ultra-high gain-bandwidth product (~THz) is needed to bridge the three-order-of- magnitude voltage gap between SC logic (~25mV) and cryo-DRAM (~0.8-1V). In this work, we demonstrate high-frequency operation of 18nm gate length (L_G) FDSOI NMOS and PMOS from 300K to 5.5K operating temperature (T). We show record unity current-gain cutoff frequency (f_T) of 495/337 GHz (35%/25% gain over 300K) and maximum oscillation frequency (f_{MAX}) of 497/372 GHz (30%/30% gain) for NMOS/PMOS at 5.5 K. A small-signal equivalent circuit model is developed to identify the T-dependent and T-invariant components of the extrinsic and intrinsic FET. Cryo- RF CMOS on 22 nm FDSOI platform enables access to Terahertz analog regime, while providing Giga-scale digital density at the same time.

T5-4 - 9:50
A Unified Physical BTI Compact Model in Variability-Aware DTCO Flow: Device Characterization and Circuit Evaluation on Reliability of Scaling Technology Nodes, Y. Zhao*,****, L. Wang*, Z. Wu*, F. Schanovsky**, X. Xu*, H. Yu***, J. Lai*, D. Liu*,****, X. Chuai*,****, Y. Su*,****, Y. Hong*, X. Wang***, L. Li* and M. Liu*, *Insititute of Microelectronics Chinese Academy of Science, China, **Global TCAD Solutions GmbH, Vienna, Austria, Austria, ***Huazhong Univ. of Science and Technology and ****China Univ. of Chinese Academy of Sciences, China

We developed a unified physical and statistical compact model of Bias Temperature Instability (BTI) effects on scaling technology nodes towards robust VLSI design, with an excessive amount of complex stress/recovery pattern characterization, ultralong-term aging prediction, and technology of statistical variability (TSV) analysis, realizing cycle-to-cycle/device-to-device reliability evaluations. This model is based on a 2/4-state Defect-Centric (DC) theory and verified by TCAD simulation, providing a deep insight into the properties of the defects (e.g., energy level distribution, occupancy probability etc.). By calibration to Fin-FET experiments (of down to 14 nm node), it is successfully implemented into BSIM-CMG for analysis of dynamic time evolutionary and dynamic voltage scaling. This physics-, variablity-, and tolerance- aware model has the potential to boost the design technology co-optimization (DTCO) flow of reliability in VLSI to the next generation of technology nodes.

T5-5 - 10:00
Electromigration-Induced Bit-Error-Rate Degradation of Interconnect Signal Paths Characterized from a 16nm Test Chip, N. Pande*, C. Zhou*****, MH. Lin**, R. Fung***, R. Wong****, S.-J. Wen**** and C. Kim*, *Univ. of Minnesota, USA, **TSMC, Taiwan, ***Cisco Systems, Inc., China, ****Cisco Systems, Inc. and *****Maxim Integrated, USA

An array-based test-vehicle for tracking bit-error-rate (BER) degradation of signal interconnects subject to DC electromigration (EM) stress was implemented in a 16nm FinFET process. A unit interconnect path comprises five identical interconnect stages where each wire is driven by inverter based buffers. Accelerated EM stress testing is achieved entirely on-chip using metal heaters located directly above the devices-under-test (DUTs) and separate stress circuits driving both ends of the wire. BER measurement results from four individual interconnect paths are presented and analyzed.

Panel Session [Room 1, 4]

Thursday, June 17, 7:00-9:00

Chairpersons:

7:00-8:30 Circuits Panel [Room 1]
New Generation Chip Makers vs. the Incumbents

8:00-9:00 Technology Panel [Room 4]
3D/Heterogeneous Integration: Are We Running Towards a Thermal Crisis?

SESSION 6

Ferroelectric Devices and Memory -1 [Room 4]

Thursday, June 17, 8:40-9:10

Chairpersons: S. Fujii, KIOXIA Corp.
G. Yeric, Cerfe Labs, Inc.

T6-1 - 8:40
Ferroelectric Diodes with Sub-ns and Sub-fJ Switching and Its Programmable Network for Logic-in-Memory Applications, H. Bae*, T. Moon*, S. G. Nam*, K.-H. Lee*, S. Kim*, S. Hong**, D.-H. Choe*, S. Jo*, Y. Lee* and J. Heo*, *Samsung Advanced Institute of Technology and **Seoul National Univ., Korea

In this work, we report hafnium zirconium oxide (HZO)/indium gallium zinc oxide (IGZO)-based programmable ferroelectric (FE) diode memory array with ultra-fast sub-ns switching speed, thus extremely low sub-fJ switching energy for future in-memory and neuromorphic computing applications for the first time. The fabricated devices have electroresistance ratio of $3x10^5$ and show robust cyclic endurance up to 10^9. In particular, we demonstrated a nonvolatile logic-in-memory circuit to implement NOT gate and incremental conductance changes to mimic analog nature of synaptic weights.

T6-2 - 8:50
Mobility-Enhanced FET and Wakeup-Free Ferroelectric Capacitor Enabled by Sn-doped InGaZnO for 3D Embedded RAM Application, J. Wu*, F. Mo*, T. Saraya*, T. Hiramoto*, M. Ochi**, H. Goto*** and M. Kobayashi*, *The Univ. of Tokyo, **Kobe Steel, Ltd and ***Kobelco Research Institute, Inc, Japan

We have developed and integrated mobility-enhanced FET and wakeup-free ferroelectric capacitor by using Sn-doped InGaZnO, and demonstrated 1T1C FeRAM cell operation for 3D embedded RAM application, for the first time. We have illustrated the impact of thin-film access transistor on ferroelectric capacitor, and investigated the physics of mobility enhancement. The proposed memory technology will enable high-density and energy-efficient computing by the proximity of processor core and memory in monolithic 3D integration.

T6-3 - 9:00

Characterization of Fatigue and Its Recovery Behavior in Ferroelectric HfZrO, P.-J. Liao, Y.-K. Chang, Y.-H. Lee, Y.-M. Lin, S. H. Yeong, R.-L. Hwang, V. Hou, C.-H. Nien, R. Lu and C. T. Lin, TSMC, Taiwan

In this study, polarization fatigue of HfZrO ferroelectric is investigated with SILC (stress-induced-leakage-current) measurement under different E-field stresses. Under high-field, we observed strong correlation between polarization wake-up and SILC increase. This is attributed to oxygen vacancy redistribution and percolation path formation, especially at high frequency cycling. However, polarization fatigue at low field is found to occur without SILC increase. P-E loop measurements and material analysis by TEM/PED (precession electron diffraction) revealed that charge trapping is the main contributor under the low-bias. We demonstrated that the fatigue caused by low-field stress could be effectively recovered through an interspersed periodical, short-term cycles at high-field to manage charge trapping and oxygen vacancy redistribution, thus resulting in prolonged endurance to >1E12 cycles without SILC degradation at room temperature. We also validated that a negligible fatigue switching in HfZrO can be achieved at -40°C as low-temperature operation further reduces charge trapping.

T6-4 Live Q&A Session : June 17, 15:40 (JST)

Mechanism of Retention Degradation After Endurance Cycling of HfO_2-Based Ferroelectric Transistors, H. Zhou*, J. Ocker*, M. Pesic**, A. Padovani**, M. Trentzsch***, S. Dunkel***, J. Muller***, S. Beyer***, L. Larcher**, F. Koushan*, S. Muller* and T. Mikolajick****,*****, *Ferroelectric Memory GmbH, Germany, **Applied Materials, Inc., USA, ***GLOBALFOUNDRIES Inc., ****NaMLab gGmbH and *****Technical Univ. of Dresden, Germany

In this study, we provide insight into the mechanism of retention degradation after endurance cycling of HfO_2-based ferroelectric field-effect transistors (FeFETs). Transfer characteristics of the FeFET are compared with the current-voltage response of the ferroelectric capacitors (FeCAP) for better understanding of the retention loss mechanism after cycling. Furthermore, a multiscale simulation by using the Ginestra™ modeling platform is conducted and the results show that charge trapping stabilizes the polarization switching and improves the retention behavior. Retention after cycling experiments are shown as well as pathways to reduce this degradation mechanism.

SESSION 7

Ferroelectric Devices and Memory -2 [Room 4]

Thursday, June 17, 9:10-9:50

Chairpersons: B. H. Lee, POSTECH
P. Ye, Perdue Univ.

T7-1 - 9:10

Higher-k Zirconium Doped Hafnium Oxide (HZO) Trigate Transistors with Higher DC and RF Performance and Improved Reliability, W. Chakraborty*, M. San Jose*, J. Gomez*, A. Saha**, K. A. Aabrar*, P. Fay*, S. Gupta** and S. Datta*, *Univ. of Notre Dame and **Purdue Univ., USA

In this work we demonstrate a novel strategy to reduce the EOT of high K HfO_2 gate-stacks, by enhancing the dielectric constant (k) through optimum Zirconium (Zr) doping Hf Zr ratio of 3 7. Through comprehensive theoretical modeling and experimental characterization, we show: 1) Higher-k response in scaled Zr-doped (70%) HfO_2 (HZO) gate-stack, due to strong inter-domain electrostatic interactions; 2) a 22% EOT reduction in HZO FET over control HfO_2 without any mobility degradation, resulting in 20% and 18% boost in drive-current ($I_{D,SAT}$) and transconductance (g_m) respectively; 3) improved device reliability such as positive bias temperature instability (PBTI) resulting in 60mV higher V_{MAX} gain and 4) consistent enhancement in g_m arising from thinner EOT that persists in the gigahertz frequency domain.

T7-2 - 9:20

Ultra-Low Power Robust 3bit/cell $Hf_{0.5}Zr_{0.5}O_2$ Ferroelectric FinFET with High Endurance for Advanced Computing-In-Memory Technology, S. De*, D. D. Lu*, H.-H. Le*, S. Mazumder*, Y.-J. Lee**, W.-C. Tseng*, B.-H. Qiu*, M. A. Baig*, P.-J. Sung**, C.-J. Su**, C.-T. Wu**, W.-F. Wu**, W.-K. Yeh** and Y.-H. Wang*, *National Cheng Kung Univ. and **Taiwan Semiconductor Research Institute (TSRI), Taiwan

Scaled ferroelectric FinFET devices were fabricated with post fin formation surface engineering (SE) to remove the line-edge roughness (LER) from the silicon surface by dry etching. This facilitated 3bit/cell operations in 10 nm $Hf_{0.5}Zr_{0.5}O_2$ based ferroelectric FinFETs along with on-state current (I_{ON}) to off-state current (I_{OFF}) ratio of 10^6, extrapolated 10-year retention and endurance above 10^{11} cycles. Further, we have evaluated its performance in all ferroelectric neural network, where ferroelectric FinFETs are used as synaptic devices or neurons for weight storage. Synaptic core built with optimized devices achieve software-comparable 97.91% inference accuracy on MNIST data and multi-layer perceptron network.

T7-3 - 9:30

Strategy Toward HZO BEOL-FeRAM with Low-Voltage Operation (≤ 1.2 V), Low Process Temperature, and High Endurance by Thickness Scaling, K. Tahara*, K. Toprasertpong*, Y. Hikosaka**, K. Nakamura**, H. Saito**, M. Takenaka* and S. Takagi*, *The Univ. of Tokyo and **Fujitsu Semiconductor Memory Solution Ltd., Japan

The potential of thickness scaling in ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ (HZO) is investigated by a systematic study on MFM capacitors with HZO thickness from 9.5 nm down to 2.8 nm. We establish the thickness-temperature mapping indicating a clear tradeoff between the thickness scaling and crystallization temperature, which has to be taken into account in the implementation as back-end-of-line (BEOL) FeRAM. Utilizing the thickness scaling and high-field wake-up without reliability loss, we demonstrate 4-nm-thick HZO having low crystallization temperature (500°C), excellent ferroelectricity (P_r > 25 μC/cm^2), low operating voltage (0.7-1.2 V), and high read/write endurance (projected to 10^{14}).

T7-4 - 9:40
First Demonstration of BEOL-Compatible Ferroelectric TCAM Featuring *a*-IGZO Fe-TFTs with Large Memory Window of 2.9 V, Scaled Channel Length of 40 nm, and High Endurance of 10^8 Cycles, C. Sun*, K. Han*, S. Samanta*, Q. Kong*, J. Zhang*, H. Xu*, X. Wang*, A. Kumar*, C. Wang*, Z. Zheng*, X. Yin**, K. Ni*** and X. Gong*, *National Univ. of Singapore, Singapore, **Zhejiang Univ., China and ***Rochester Institute of Technology, USA

For the first time, we report BEOL-compatible ferroelectric ternary content-addressable memory (TCAM) based on amorphous IGZO (*a*-IGZO) channel and HZO ferroelectric (Fe) layer, achieving a much larger sensing margin as compared with other TCAM technologies. Our a-IGZO ferroelectric thin-film transistors (Fe-TFTs) were realized using a low-temperature process of 400 °C and an MFMIS structure with the flexibility to engineer the area ratio of the ferroelectric layer and the metal-oxide-semiconductor layer. The Fe-TFTs not only enjoy the largest memory window (MW) of 2.9 V for HZO-based Fe-TFTs with oxide semiconductor channels and high endurance of 10^8 cycles, but also a high conductance ratio and small cycle-to-cycle variation, leading to a high recognition accuracy (90.4%) of handwritten digits. Ultra-scaled devices with a channel length of 40 nm exhibits enhanced drive current with MW as large as 2.8 V.

Technology / Circuits Joint Focus Session 1

3D/Heterogeneous Integration [Room 5]

Thursday, June 17, 8:40-9:20

Chairpersons: T. Tanaka, Tohoku Univ.
M. Delaus, Analog Devices, Inc.

JFS1-1 - 8:40 (Invited)
Design and Technology Solutions for 3D Integrated High Performance Systems, G. Van der Plas and E. Beyne, imec, Belgium

3D system integration builds on interconnect scaling roadmaps of TSVs (5µm to 100nm CD) and fine pitch bumps/pads (to <1µm pitch) for D2W and W2W schemes. Si bridges connect chiplets at 9.5Gbp, 338fJ/b, while W2W fine pitch memory logic functional partitioning improves power/performance by 30% vs 2D. Impingement cooler, BSPDN, high density MIMCAP and integrated magnetics push the power wall to 300W/cm^2. On the other hand, 3D design flows require further development. Process optimization, DfT, KGD/S and heterogeneous technology optimization of functionally partitioned 3D-SOC make high performance systems cost-effective.

JFS1-2 - 8:50 (Invited)
Chiplet-Based Advanced Packaging Technology from 3D/TSV to FOWLP/FHE, T. Fukushima, Tohoku Univ., Japan

More recently, "chiplets" are expected for further scaling the performance of LSI systems. However, system integration with the chiplets is not a new methodology. The basic concept dates back well over a few decades. The symbolic configuration of this concept based on the chiplets is 3D integration with TSV we have worked on since 1989. This paper introduces our 3D and heterogeneous system integration research from its historical activities to the latest efforts, including capillary selfassembly of tiny dies with a size of less than 0.1 mm and advanced flexible hybrid electronics (FHE) using fan-out wafer-level packaging (FOWLP).

JFS1-3 - 9:00
A 7nm 0.46pJ/bit 20Gbps with BER 1E-25 Die-to-Die Link Using Minimum Intrinsic Auto Alignment and Noise-Immunity Encode, Y. Y. Hsu, MediaTek Inc., Taiwan

A high-density low bit error rate and low-power PHY for ultra-short-reach (USR) die-to-die communication has been fabricated in TSMC 7nm FinFET 1P15M CMOS technology. Interconnection is demonstrated through TSMC Chip-on-Wafer-on-Substrate (CoWoS) and TSMC Integrated Fan-Out (InFO) packaging technology. PHY exploits energy-efficient and high performance scheme, includes single-ended without termination, quarter rate strobe and unbalance scheme on transceiver, minimum intrinsic auto-alignment and novel noise-immunity coding methodology. Achieving 20Gbps per wire and 0.46pJ/bit under 1-mm ultra-short-reach platform target to BER 1E-25. Bandwidth density is shoreline 5.31Tbps/mm and area 2.25Tbps/mm^2.

JFS1-4 - 9:10
Ultra High Power Cooling Solution for 3D-ICs, C. J. Wu, S. T. Hsiao, J. Y. Wang, W. H. Lin, C. W. Chang, D. L. Shao, C. H. Tung and D. C. H. Yu, TSMC, Taiwan

A direct silicon water cooling solution using fusion bonded silicon lid is proposed. It is successfully demonstrated as an effective cooling solution with total power >2600 W on a single SoC, equivalent to power density of 4.8 W/mm^2. Low temperature logic chip to silicon lid fusion bonding, with trench/grid cooling structure cutting into silicon lid enables minimal thermal resistance between active device and cooling water and best cooling efficiency. Direct water cooling on logic chip silicon backside has also been demonstrated with power density better than 7 W/mm^2.

SESSION 8

3D Flash Memory [Room 5]

Thursday, June 17, 9:20-10:00

Chairpersons: K. Hamada, Micron Memory Japan, G.K.
J. Yu, Western Digital Corp.

T8-1 - 9:20
First Demonstration of Ruthenium and Molybdenum Word lines Integrated into 40nm Pitch 3D-NAND Memory Devices, A. Ajaykumar, L. Breuil, K. Katcko, F. Schleicher, F. Sebaai, Y. Oniki, S. Ramesh, A. Arreghini, L. Nyns, J.-P. Soulie, J. Stiers, G. Van den Bosch and M. Rosmeulen, imec, Belgium

We demonstrate the integration of Ruthenium (Ru) and Molybdenum (Mo) as Word Line (WL) metals in a record 40nm pitch 3D-NAND device through an optimized Replacement Metal Gate (RMG) process. The optimized RMG process minimizes oxide regrowth which affects WL fill capability in reduced pitches. Ru and Mo gates show better resistivity and memory characteristics compared to the currently used Tungsten WL. We demonstrate good channel control and program/erase (P/E) characteristics down to 20nm WL. Best P/E is obtained for Mo with 2nm HfOx liner after a post metallization anneal (PMA) at 750C for 20mins, while devices with Ru WL show better retention.

T8-2 - 9:30
A Novel Micro Wall Heater for Thermally-Assisted 3D AND-Type Flash Memory to Radically Boost the Write/Erase Speed and Endurance for the Applications of Write-Intensive Persistent Memory, H.-T. Lue, Macronix International Co., Ltd., Taiwan

We demonstrate a novel micro wall heater implemented in a 3D AND type Flash memory to radically enhance the speed and endurance. The micro wall heater is a low-resistance tungsten metal plate that is built in a slit in the 3D AND structure. The micro wall heater can heat up the entire small sector (2K cells) above 300C during read, write, and erase operations. It provides thermal assisted read (1.5-time boosted), thermal-assisted FN Programming (1usec write for 5V Vt window), and thermal-assisted FN erasing (10usec erase for 5V Vt window). The micro wall heater provides self-healing Flash to recover the cycling-induced damage. 50M endurance is achieved with regular heater annealing per 5K cycling stress, and 10M endurance is achieved by thermal-assisted erase during each cycling. This micro wall heater design for 3D AND-type Flash memory can support the write-intensive persistent memory applications.

T8-3 - 9:40
Highly-Reliable Cell Characteristics with 128-Layer Single-Stack 3D-NAND Flash Memory, S. Park, J. Lee, J. Jang, J. K. Lim, H. Kim, J. J. Shim, M.-t. Yu, J.-K. Kang, S. J. Ahn and J. Song, Samsung Electronics Co., Ltd., Korea

An industry leading 128-layer single-stack 3D-NAND Flash memory with high reliability cell characteristics is successfully developed for the first time. Single-stack etching of 128 layers brings about various deformation of hole profile, which leads to the degradation of cell characteristics. The degradation of cell characteristics has been overcome by advanced HARC (high-aspect ratio contact) etching process and ONO material engineering. The single-stack NAND Flash memory still has a potential to be developed further with the advanced single-stack technology.

T8-4 - 9:50
A Hybrid In-Memory-Searching and In-Memory-Computing Architecture for NVM Based AI Accelerator, P. H. Tseng, Macronix International Co., Ltd., Taiwan

A hybrid architecture combining in-memory-searching (IMS) and in-memory-computing (IMC) is proposed to accelerate neural network inference computation. The high speed NOR-flash IMS unit, with large unit cell discharge time ratio ($>10^6$) and fast sensing time (<1ns), provides coarse filtering function to reduce IMC loading. The 3D-NAND-flash IMC unit for multiplication-accumulation (MAC) operation calculates the likeliness scores between the inquiry and the IMS selected instances in the NAND array. With coarse filtering function of IMS, the efficiency for the IMC scoring function can be improved up to 1000X in our demonstration. Adding augmented instances to the abundant IMC unit can further improve the accuracy and the reliability of the system. This is enabled by enhancing the tolerance to inquiry variation and memory fail bits.

SESSION 9

Power Device [Room 6]

Thursday June 17, 9:10-9:40

Chairpersons: K. Tateiwa, Tower Partners Semiconductor Co., Ltd.
M. Shulaker, Massachusetts Institute of Technology (MIT)

T9-1 - 9:10
Backside Power Delivery with a Direct 14:1/19:1 High-Ratio Point-of-Load Power Converter for Servers and Datacenters, H. LIN*,**, G. Hiblot*, X. Sun*, G. Talmelli*,**, D. Velenis*, P. Bex*, C. Adelmann*, R. Lauwereins*,**, F. Catthoor*,**, G. Van der Plas* and E. Beyne*, *imec and **KU Leuven, Belgium

For computing systems, a fully integrated backside power delivery with a direct 14:1/19:1-ratio power converter is presented featuring (1) backside laterally-diffused power MOS with 1.5x lower Qon*Ron and 4.9x lower output Coss compared to the stacked IO device, and (2) in-package optimized transformer with an innovative magnetic material (Bsat=0.77 T, coercivity=0.35 Oe, and resistivity>1e8 Ω·cm). With a 14:1 conversion ratio and 166x boost in power per volume (vs discrete), the total power delivery efficiency is 72% at 1 W/mm^2.

T9-2 - 9:20
91.5%-Efficiency Fully Integrated Voltage Regulator with 86fF/μm²-High-Density 2.5D MIM Capacitor, H. Lin*,**, D. Velenis*, P. Nolmans*, X. Sun*, F. Catthoor*,**, R. Lauwereins*,**, G. Van der Plas* and E. Beyne*, *imec and**KU Leuven, Belgium

We present a 91.5% power-delivery-efficiency for high performance computing (HPC) systems including a 1/2-ratio charge pump circuit with 2.5D MIM capacitor. With high capacitance density (86 fF/μm^2, 1.36-V bias/10 year/100°C), 0.1% parasitic impact, and small form factor (~μm thick), this MIM solution enables converters with high ratio (up to 1/5 with 82% efficiency), high power density (4.8 W/mm^2 for ratio=1/3), low V_{OUT} (0.7 V), and sub-ns transient response for HPC systems.

T9-3 - 9:30
FinFET LDMOS Technology Challenges and Opportunities for Digital TV and 6GHz WiFi PA Applications, J. singh, GLOBALFOUNDRIES Inc., USA

A comprehensive 12nm FinFET process-based LDMOS DC, RF, and extended reliability study is presented for the first time. To determine the optimal intrinsic device architecture(i) extended drain(ii) dummy poly and (iii) STI LDMOS architectures were studied. The STI LDMOS shows a higher BV_{dss}, lower C_{gd}, higher F_{max} and MSG (maximum stable gain) with lower HCI degradation among these three options. Moreover, the 3.3V STI nLDMOS delivers 1.84x of I_{dsat} compared to previous node 22nm planar nLDMOS for comparable BV_{dss} values. On the other hand, driving higher current through narrow silicon fins brings HCI reliability challenges. In response, 5V nLDMOS was optimized for robust higher on-state voltage handling and reliability, while the 3.3V nLDMOS was configured for enhanced analog and RF performance. The 3.3V nLDMOS demonstrate an excellent RF F_t/F_{max} of 37.7/81GHz, MSG of 17.3dB performance with a robust $\Delta I_{dsat}/I_{dlin}$<10/20% HCI EOL performance.

SESSION 10

IGZO Transistor and III-V Device [Room 6]

Thursday, June 17, 9:40-10:20

Chairpersons: O. Cheng, United Microelectronics Corp. (UMC)
S. Choudhuri, Stanford Univ.

T10-1 - 9:40
First Demonstration of Oxide Semiconductor Nanowire Transistors: a Novel Digital Etch Technique, IGZO Channel, Nanowire Width Down to ~20 nm, and I_{on} Exceeding 1300μA/μm, K. Han, Q. Kong, Y. Kang, C. Sun, C. Wang, J. Zhang, H. Xu, S. Samanta, J. Zhou, H. Wang, V.-Y. Thean and X. Gong, National Univ. of Singapore, Singapore

We report the first realization of oxide semiconductor based nanowire field-effect transistor having ultra-scaled amorphous Indium-Gallium-Zinc-Oxide nanowire channel (width down to ~20 nm) enabled by a novel digital etch technique. A device with ~25 nm nanowire width and 100 nm channel length achieves the highest peak extrinsic transconductance of 612 μS/μm at V_{DS} = 2 V (456 μS/μm at V_{DS} = 1 V) among all IGZO-based transistors, and the highest on-state current (I_{on}) of 620 μA/μm at V_{DS}= 1 V and V_{DS} - V_{TH} = 2 V among all oxide semiconductor based transistors with top gate structure. I_{on} exceeds 1300μA/μm at higher bias voltages. In addition, a peak intrinsic transconductance of 915 μS/μm at V_{DS} = 2 V (609 μS/μm) were realized.

T10-2 - 9:50
2-kbit Array of 3-D Monolithically-Stacked IGZO FETs with Low SS-64mV/dec, Ultra-Low-Leakage, Competitive Mobility-57 cm²/V-s Performance and Novel nMOS-Only Circuit Demonstration, U. Chand, National Univ. of Singapore, Singapore

We demonstrate 3D monolithically-stacked IGZO FETs with competitive performance for BEOL circuits. The FETs show excellent switching electrostatics with near-deal subthreshold swing, ultra-low leakage density (10^{-6} A/cm^2) and superior electron mobility (57 cm^2/V-s). The device fabrication are achieved at low-temperatures (T-350°C), making the process highly compatible with low-thermal budget Cu interconnect. We have enabled body-contacted IGZO FETs, capable of dynamic body biasing and implemented these devices into functional novel circuits. We show the feasibility of stacking these FETs with little impact.

T10-3 - 10:00
NbN-Gated GaN Transistor Technology for Applications in Quantum Computing Systems, Q. Xie*, N. Chowdhury*, A. Zubair*, M. Sanchez Lozano*, J. Lemettinen*,**, M. Colangelo*, O. Medeiros*, I. Charaev*, K. K. Berggren*, P. Gumann***, D. Pfeiffer*** and T. Palacios*, *Massachusetts Institute of Technology (MIT), USA, **Aalto Univ., Finland and ***IBM T. J. Watson Research Center, USA

A NbN-gated AlGaN/GaN high electron mobility transistor (HEMT) technology for applications in quantum computing systems is demonstrated for the first time. Transistors with gate lengths scaled to 250 nm were characterized at 4.2 K, with excellent gate modulation ($I_{D,ON}/I_{D,OFF}$ ~ 10^8) and current saturation. The potential of these devices for low noise amplifiers was evaluated, revealing a low DC power dissipation of 25 μW/μm when biased for expected minimum noise. The RF performance was also characterized at 4.2 K. This work highlights the potential of NbN-gated GaN transistor technology for applications in low-noise cryogenic amplifiers in future quantum computing systems.

T10-4 - 10:10

High-Performance InGaAs-On-Insulator HEMTs on Si CMOS for Substrate Coupling Noise-Free Monolithic 3D Mixed-Signal IC, J. Jeong*, S. K. Kim*, J. Kim**, D.-M. Geum*, J. Park***, J.-H. Jang*** and S. H. Kim*, *KAIST, **Korea Advanced NanoFab Center (KANC) and ***Gwangju Institute of Science and Technology, Korea

For the substrate coupling digital noise-free monolithic 3-dimensional (M3D) mixed-signal integrated circuit (IC), we have demonstrated InGaAs HEMTs on Si CMOS. Bottom digital circuits were fabricated by 180 nm standard Si CMOS and top RF InGaAs HEMTs were fabricated on Si CMOS by wafer bonding-based layer transfer and low-processing temperature of 300 °C or less. As a result, top InGaAs HEMTs exhibit f_T and f_{MAX} of 448 and 213 GHz without any degradation in the bottom Si CMOS. It is the record-high f_T ever reported by M3D RF transistors. Furthermore, for the first time, we have successfully demonstrated that the influence of substrate digital noise can be eliminated by integrating the analog and digital circuits in the M3D platform.

T10-5 **Live Q&A Session** **: June 17, 15:50 (JST)**

First Demonstration of Sub-12 nm L_g Gate Last IGZO-TFTs with Oxygen Tunnel Architecture for Front Gate Devices, S. Subhechha, N. Rassoul, A. Belmonte, R. Delhougne, K. Banerjee, G. L. Donadio, H. Dekkers, M. van Setten, H. Puliyalil, M. Mao, S. Kundu, M. Pak, L. Teugels, D. Tsvetanova, N. Bazzazian, L. Klijs, H. Hody, A. V. Chasin, J. Heijlen, L. Goux and G. S. Kar, imec, Belgium

The first amorphous IGZO-based transistors integrated in a BEOL compatible gate-last integration scheme with buried oxygen tunnel under the channel and self-aligned contacts are demonstrated. This architecture reduces the number of critical process steps containing hydrogen and mitigates the R_{series} increase during oxygen anneal for defect passivation. Higher I_{on} is achieved without compromising on I_{off}. We also report the shortest ever IGZO-TFTs with L_g < 12 nm. Further, normally-off devices with I_{on} > 6 μA/μm (without defect passivation anneal) were achieved through gate-oxide and IGZO thickness scaling.

Additional Q&A Session

Thursday, June 17 15:00-16:00

Chairpersons: F. Amaud, ST Microelectronics
K. Maekawa, Renesas Electronics Corp.

TFS1-5 - 15:00 (Invited)

Development Status of Gate-First FeFET Technology, S. Müller1, H. Zhou*, A. Benoist*, J. Ocker*, M. Noack*, G. Kuzmanov*, R. Iqbal*, D. Le Minh*, M. Ghazaryan*, V. Anjaneyamoorthi*, A. Daraghmah*, M. Mennenga*, F. Koushan*, F. Tassan*, S. Dünkel**, J. Mülle**, S. Beyer**, S. Soss*** and A. Pourkeramati*, *Ferroelectric Memory GmbH, **GLOBALFOUNDRIES Inc., Germany and ***GLOBALFOUNDRIES Inc., USA

Recent progress on FeFET gate-first technology development is presented. New characterization results from FeFET endurance degradation are shown and assigned to interfacial layer degradation. Two methods to overcome endurance degradation in terms of proper choice of device geometry or program / erase algorithms are highlighted. Moreover, statistical variation of FeFET memory states is characterized for single memory cells as well as mini arrays across wafer. This is complemented by 136 Kbit FeFET array results which demonstrate tail-to-tail separation of ~3μA which represents the basis for read-out operations below 25 ns. Latest results from FeFET variability from 180nm x 180nm as well as 72nm x 72nm memory cells is presented and a 32 Mbit macro incorporating 180nm x 180nm cells has been designed for future characterization.

T3-5 - 15:10

Record Performance of 500°C Low-Temperature nMOSFETs for 3D Sequential Integration Using a Smart Cut™ Layer Transfer Module, L. Brunet*, S. Reboh*, T. Januel*, X. Garros*, T. Mota Frutuoso*, M. Casse*, B. Sklenard*, L. Brevard*, M. Ribotta*, A. Magalhaes-Lucas*, J. Kanyandekwe*, J.-M. Hartmann*, F. Milesi*, F. Mazen*, P. Acosta-Alba*, S. Kerdiles*, A. Tavernier*, V. Loup*, C. Morales*, V. Larrey*, F. Fournel*, L. Le Van-Jodin*, S. Leforestier*, E. Rolland*, G. Romano*, G. Gaudin**, J. Lugo*, J. Lacord*, S. Maitrejean*, J. Arcamone*, P. Batude*, I. Radu**, C. Fenouillet-Beranger* and F. Andrieu*, *CEA-LETI and **Soitec, France

We report record performances in Top-tier nMOSFETs fabricated by 3D sequential integration, as well as junction optimization guidelines to further optimize the performances within a maximum thermal budget of 500°C. We reached I_{ON}=870μA/μm at I_{OFF}=100nA/μm at V_{DD}=1V together with A_{VT}=1.35mV.μm and a decent PBTI lifetime.Moreover, a first generation of Low Temperature LT layer transfer module from Si substrates based on Smart Cut™ was developed to obtain LTSOI quality compatible with 500C FDSOI process integration: RMS=0.083nm roughness and 0.4nm SOI uniformity. The nMOSFETs fabricated on these LTSOI wafers reached I_{ON}-I_{OFF} performances at 88% of reference transistors integrated on regular SOI wafers.

T3-6 - 15:20

High Yield and Process Uniformity for 300 mm Integrated WS_2 FETs, T. Schram, Q. Smets, D. Radisic, B. Groven, D. Cott, A. Thiam, W. Li, E. Dupuy, K. Vandersmissen, T. Maurice, I. Asselberghs and I. Radu, imec, Belgium

We demonstrate an integrated process flow on full 300mm wafers with monolayer WS_2 channel. WS_2 is a 2D semiconductor from the transition metal dichalcogenide family and holds promise for extreme gate length scaling. We report here on integration challenges and optimize process uniformity for a single-device yield higher than 90% across wafer. These transistors and integration flow are shown to be compatible with H2 anneal of at least 400 C and hence in principle suitable for hybrid integration in the BEOL.

TFS2-6 - 15:30
Enabling Logic with Backside Connectivity Via n-TSVs and Its Potential as a Scaling Booster, A. Veloso, A. Jourdain, G. Hiblot, F. Schleicher, K. Dhave, F. Sebaai, D. Radisic, R. Loo, T. Hopf, A. De Keersgieter, H. Arimura, G. Eneman, P. Favia, J. Geypen, G. Arutchelvan, A. Chasin, D. Jang, L. Nyns, E. Rosseel, A. Hikavyy, G. Mannaert, B. T. Chan, K. Devriendt, S. Demuynck, G. Van der Plas, J. Ryckaert, G. Beyer, E. Dentoni Litta, E. Beyne and N. Horiguchi, imec, Belgium

We report on scaled Si-channel finFETs (L_{gate}>=20nm, 45nm fin pitch) with backside connectivity enabled by extreme wafer thinning (several Si thicknesses under STI-oxide targeted: from ~370nm down to ~20nm) and W-filled nano-through-Si-vias (n-TSV) of various heights, after using low-temperature, wafer-to-wafer, dielectric bonding. This scheme aims at allowing decoupling signal and power networks, with reduced IR-drop predicted by moving the latter to the wafer backside. A thorough evaluation of the impact of 3D processing on device characteristics is presented, showing: 1) enhanced nmos mobility and drive currents (up to 15%); 2) for pmos, small ION loss (~3 to 10%), larger R_{ext}, with channel strain evaluation by NBD for various layouts; 3) ΔV_T~-130mV that can be recovered with an extra anneal at the end, keeping tight variability and matching control. No BTI degradation is observed, with further indication that the final anneal(s) selection can be beneficial for electrostatics and reliability improvement.

T6-4 - 15:40
Mechanism of Retention Degradation After Endurance Cycling of HfO_2-Based Ferroelectric Transistors, H. Zhou*, J. Ocker*, M. Pesic**, A. Padovani**, M. Trentzsch***, S. Dunkel***, J. Muller***, S. Beyer***, L. Larcher**, F. Koushan*, S. Muller* and T. Mikolajick****,*****, *Ferroelectric Memory GmbH, Germany, **Applied Materials, Inc., USA, ***GLOBALFOUNDRIES Inc., ****NaMLab gGmbH and *****Technical Univ. of Dresden, Germany

In this study, we provide insight into the mechanism of retention degradation after endurance cycling of HfO_2-based ferroelectric field-effect transistors (FeFETs). Transfer characteristics of the FeFET are compared with the current-voltage response of the ferroelectric capacitors (FeCAP) for better understanding of the retention loss mechanism after cycling. Furthermore, a multiscale simulation by using the Ginestra™ modeling platform is conducted and the results show that charge trapping stabilizes the polarization switching and improves the retention behavior. Retention after cycling experiments are shown as well as pathways to reduce this degradation mechanism.

T10-5 - 15:50
First Demonstration of Sub-12 nm L_g Gate Last IGZO-TFTs with Oxygen Tunnel Architecture for Front Gate Devices, S. Subhechha, N. Rassoul, A. Belmonte, R. Delhougne, K. Banerjee, G. L. Donadio, H. Dekkers, M. van Setten, H. Puliyalil, M. Mao, S. Kundu, M. Pak, L. Teugels, D. Tsvetanova, N. Bazzazian, L. Klijs, H. Hody, A. V. Chasin, J. Heijlen, L. Goux and G. S. Kar, imec, Belgium

The first amorphous IGZO-based transistors integrated in a BEOL compatible gate-last integration scheme with buried oxygen tunnel under the channel and self-aligned contacts are demonstrated. This architecture reduces the number of critical process steps containing hydrogen and mitigates the R_{series} increase during oxygen anneal for defect passivation. Higher I_{on} is achieved without compromising on I_{off}. We also report the shortest ever IGZO-TFTs with L_g < 12 nm. Further, normally-off devices with I_{on} > 6 μA/μm (without defect passivation anneal) were achieved through gate-oxide and IGZO thickness scaling.

Joint Panel Session

The New Normal… How will it Change Work, Life and Education? [Room 1]

Friday, June 18, 7:30-8:30

Technology / Circuits Joint Focus Session 2

Computing-in-Memory [Room 1]

Friday, June 18, 8:40-10:00

Chairpersons: K. Sohn, Samsung Electronics Co., Ltd.
F. Sheikh, Intel Corp.

JFS2-1 - 8:40
A 13.7 TFLOPS/W Floating-Point DNN Processor Using Heterogeneous Computing Architecture with Exponent-Computing-in-Memory, J. Lee, J. Kim, W. Jo, S. Kim, S. Kim, J. Lee and H.-J. Yoo, KAIST, Korea

An energy-efficient floating-point DNN training processor is proposed with heterogenous bfloat16 computing architecture using exponent computing-in-memory (CIM) and mantissa processing engine. Mantissa free exponent calculation enables pipelining of exponent and mantissa operation for heterogenous bfloat16 computing while reducing MAC power by 14.4 %. 6T SRAM exponent computing-in-memory with bitline charge reusing reduces memory access power by 46.4 %. The processor fabricated in 28 nm CMOS technology and occupies 1.62x3.6 mm^2 die area. It achieves 13.7 TFLOPS/W energy efficiency which is 274 times higher than the previous floating-point CIM processor.

JFS2-2 - 8:50
PIMCA: A 3.4-Mb Programmable In-Memory Computing Accelerator in 28nm for On-Chip DNN Inference, S. Yin*, B. Zhang**, M. Kim*, J. Saikia*, S. Kwon***, S. Myung***, H. Kim***, S. J. Kim***, M. Seok** and J.-s. Seo*, *Arizona State Univ., **Columbia Univ., USA and ***Samsung Advanced Institute of Technology, Korea

We present a programmable in-memory computing (IMC) accelerator integrating 108 capacitive-coupling-based IMC SRAM macros of a total size of 3.4 Mb, demonstrating one of the largest IMC hardware to date. We developed a custom ISA featuring IMC and SIMD functional units with hardware loop to support a range of deep neural network (DNN) layer types. The 28nm prototype chip achieves system-level peak energy-efficiency of 437 TOPS/W and peak throughput of 4.9 TOPS at 40MHz, 1V supply.

JFS2-3 - 9:00
A 6.54-to-26.03 TOPS/W Computing-In-Memory RNN Processor Using Input Similarity Optimization and Attention-Based Context-Breaking with Output Speculation, R. Guo*, H. Li*, R. Liu**, Z. Zhang**, L. Tang*, H. Sun*, L. Liu*, M.-F. Chang**, S. Wei* and S. Yin*, *Tsinghua Univ., China and **National Tsing Hua Univ., Taiwan

This work presents a 65nm RNN processor with computing-in-memory (CIM) macros. The main contributions include: 1) A similarity analyzer (SimAyz) to fully leverage the temporal stability of input sequences with 1.52 times performance speedup; 2) An attention-based context-breaking (AttenBrk) method with output speculation to reduce off-chip data accesses up to 30.3%; 3) A double-buffering scheme for CIM macros to hide writing latency and a pipeline processing element (PE) array to increase the system throughput. Measured results show that this chip achieves 6.54-to-26.03 TOPS/W energy efficiency vary from various LSTM benchmarks.

JFS2-4 - 9:10
Fully Row/Column-Parallel In-Memory Computing SRAM Macro Employing Capacitor-Based Mixed-Signal Computation with 5-b Inputs, J. Lee, H. Valavi, Y. Tang and N. Verma, Princeton Univ., USA

This paper presents an in-memory computing (IMC) macro in 28nm for fully row/column-parallel matrix-vector multiplication (MVM), exploiting precise capacitor-based analog computation to extend from binary input-vector elements to 5-b input-vector elements, for 16x increase in energy efficiency and 5x increase in throughput. The 1152(row)x256(col.) macro employs multi-level input drivers based on a digital-switch DAC implementation, which preserve compute accuracy well beyond the 8-b resolution of the output ADCs, and whose area is halved via a dynamic-range doubling (DRD) technique. The macro achieves the highest reported IMC energy efficiency of 5796 TOPS/W and compute density of 12 TOPS/mm^2 (both normalized to 1-b ops). CIFAR-10 image classification is demonstrated with accuracy of 91%, equal to the level of ideal SW implementation.

JFS2-5 - 9:20
HERMES Core - A 14nm CMOS and PCM-Based In-Memory Compute Core Using an Array of 300ps/LSB Linearized CCO-Based ADCs and Local Digital Processing, R. Khaddam Aljameh*, M. Stanisavljevic*, J. Fornt Mas*, G. Karunaratne*, M. Braendli*, F. F. Liu**, A. Singh*, S. Mueller***, U. Egger*, A. Petropoulos****, T. Antonakopoulos****, K. Brew**, S. Choi**, I. Ok**, F. L. Lie**, N. Saulnier**, V. Chan**, I. Ahsan**, V. Narayanan*****, N. Sasidharan Rajaleksh*, M. Le Gallo*, P. A. Francese*, A. Sebastian* and E. Eleftheriou*, *IBM Research Europe, Switzerland, **IBM Research, USA, ***IBM Systems and technology, Germany, ****Univ. of Patras, Greece and *****IBM T. J. Watson Research Center, USA

We present a 256x256 in-memory compute (IMC) core designed and fabricated in 14nm CMOS with backend-integrated multi-level phase-change memory (PCM). It comprises 256 linearized current controlled oscillator (CCO)-based ADCs at a compact 4um pitch and a local digital processing unit performing affine scaling and ReLU operations. A novel frequency-linearization technique for CCOs is introduced, leading to accurate on-chip matrix-vector-multiply (MVM) when operating over 1 GHz. Measured classification accuracies on MNIST and CIFAR-10 datasets are presented when two cores are employed for deep learning (DL) inference. The measured energy efficiency is 10.5 TOPS/W at a performance density of 1.59 TOPS/mm^2.

JFS2-6 - 9:30
A 20x28 Spins Hybrid In-Memory Annealing Computer Featuring Voltage-Mode Analog Spin Operator for Solving Combinatorial Optimization Problems, J. Mu*, Y. su* and B. Kim**, *Nanyang Technological Univ., Singapore and **Univ. of California, Santa Barbara, USA

This work proposes a hybrid analog-digital implementation of an annealing computer that achieves major improvements in both area and programmability. A compact hybrid spin circuit adopts eight voltage-mode analog spin operators that accumulate the spin interactions from neighboring spins using segmented voltage-mode drivers, two additional pairs of spin operator units for a magnetic coefficient with offset calibration, and external binary random bits for simulated annealing. A sense amplifier converts the analog spin operation result to a binary spin state, and a register stores and transmits the spin state to the neighboring spins. The test-chip is fabricated using the 65nm process, and the measured power consumption is 9.9mW at 0.8V and 320MHZ. It can achieve 1.58x improvement in the area and >3x reduction in annealing time compared with recent works.

JFS2-7 - 9:40
Analog In-Memory Computing in FeFET-Based 1T1R Array for Edge AI Applications, D. Saito*, T. Kobayashi*, H. Koga*, N. Ronchi***, K. Banerjee***, Y. Shuto**, J. Okuno**, K. Konishi**, L. D. Piazza**, A. Mallik***, J. V. Houdt***,****, M. Tsukamoto*, K. Ohkuri*, T. Umebayashi** and E. Takayuki*, *Sony Corp., **Sony Semiconductor Solutions Corp., Japan, ***imec and ****KU Leuven, Belgium

Deep neural network (DNN) inference for edge AI requires low-power operation, which can be achieved by implementing massively parallel matrix-vector multiplications (MVM) in the analog domain on a highly resistive memory array. We propose a 1T1R compute cell (1T1R-cell) using a ferroelectric hafnium oxide-based FET (FeFET) and TiN/SiO_2 tunneling junction of mega-ohm-resistor (MOR) for analog in-memory computing (AiMC). The MOR exhibited a tunneling current behavior and mega-ohm resistance. A 1T1R-cell array-level evaluation was also performed. A random access for writing with low write disturbance scheme was confirmed from the DC current summation output, and binaries were successfully classified into "T" and "L". Based on the experimental results of our proposed 1T1R-cell, we obtained a state-of-the-art energy efficiency of 13700 TOPS/W including the periphery. Furthermore, we confirmed that a high inference accuracy can be obtained with our low-resistance-variability 1T1R-cell with a properly trained model.

JFS2-8 - 9:50
Energy-Efficient Reliable HZO FeFET Computation-in-Memory with Local Multiply & Global Accumulate Array for Source-Follower & Charge-Sharing Voltage Sensing, C. Matsui, K. Toprasertpong, S. Takagi and K. Takeuchi, The Univ. of Tokyo, Japan

Energy efficient, high throughput, noise immune, high density HZO FeFET Computation-in-Memory (CiM) is proposed. Local Multiply & Global Accumulate Array is realized by source-follower read, which multiplies neural network inputs and weights (FeFET V_{TH}), and charge-sharing, which accumulates multiplied values. Proposed FeFET CiM operates 32 Multiply-lines (MLs) and 1024 Accumulate-lines (ALs) in parallel. Source-follower voltage sensing achieves 3 bit/cell for weight storage. Proposed CiM is immune to read-disturb. After 10-year data-retention, 3 bit/cell FeFET is feasible. Assuming FeFET read time of 100 ns, 66 TOPS/W is achieved. Conventional pre-charge/discharge voltage-sensing operates only bit-line (BL)-parallel (Not word-line (WL)-parallel) and increases read time. Conventional current-sensing CiM operates with 32 WL- and 16 BL-parallel, which is restricted by DC current of memory cells. Thus, proposed CiM provides 64 times higher TOPS/W than conventional current-sensing CiM.

SESSION 11

SOT-MRAM [Room 5]

Friday, June 18, 8:40-9:10

Chairpersons: H. Wu, Tsinghua Univ.
A. Agraval, Intel Corp.

T11-1 - 8:40
BEOL Compatible High Retention Perpendicular SOT-MRAM Device for SRAM Replacement and Machine Learning, S. Couet*, S. Rao*, S. Van Beek*, V. D. Nguyen*, K. Garello*,**, V. Kateel*,***, G. Jayakumar*, J. D. Costa*, K. Cai*, F. Yasin*, D. Crotti* and G. S. Kar*, *imec, Belgium, **Spintec, France and ***KU Leuven, Belgium

Spin orbit torque-MRAM devices are promising both for high performance cache replacement and for low power neural network requiring non-volatile weights, such as analog in-memory compute (AiMC). Using a new free layer design, we demonstrate a BEOL compatible perpendicular SOT device with high retention and excellent switching efficiency. $\Delta > 75\ k_BT$ is reported for 50nm devices at 125°C operating temperatures with critical current < 400μA at 1ns. This concept offers a flexible way to adjust Δ and to exploit advanced SOT material both for high performance and machine learning applications.

T11-2 - 8:50
Multi-Pillar SOT-MRAM for Accurate Analog In-Memory DNN Inference, J. Doevenspeck*, K. Garello*,**, S. Rao*, F. Yasin*, S. Couet*, G. Jayakumar*, A. Mallik*, S. Cosemans*, P. Debacker*, D. Verkest*, R. Lauwereins*,***, W. Dehaene*,*** and G. S. Kar*, *imec, Belgium, **SPINTEC, France and ***KU Leuven, Belgium

Deep neural network (DNN) inference can be performed efficiently with analog in memory computing (AiMC). MRAM is an attractive solution to implement the DNN weights due to its non-volatility and scalability. However, accurate inference requires memories with multi-level conductance values. while MRAM is binary. In this work, we propose and demonstrate a multi-level SOT-MRAM device concept by placing multiple MTJ pillars between a single SOT track and common top electrode. Selective level programming is achieved by smartly using a pillar-position dependent VCMA-assist effect. Three DNN algorithm-driven technology requirements are derived: number of conductance levels, bit-error rate and conductance variation. This work demonstrates that multi-pillar SOT-MRAM meets all derived specifications, making it a promising candidate as weight memory device for accurate analog in-memory DNN inference.

T11-3 - 9:00
First Demonstration of Interface-Enhanced SAF Enabling 400°C-Robust 42 nm p-SOT-MTJ Cells with STT-Assisted Field-Free Switching and Composite Channels, Y.-J. Tsou*,**, K.-S. Li*, J.-M. Shieh*, W.-J. Chen**, H.-C. Chen*, Y.-J. Chen*, C.-L. Hsu*, Y.-M. Huang*, F.-K. Hsueh*, W.-H. Huang*, W.-K. Yeh*, H.-C. Shih**, P.-C. Liu**, C. W. Liu**, Y.-S. Yen***, C.-H. Lai***, J.-H. Wei****, D. D. Tang**** and J. Y.-C. Sun*, *Taiwan Semiconductor Research Institute (TSRI), **National Taiwan Univ., ***National Tsing Hua Univ. and ****Industrial Technology Research Institute of Taiwan, Taiwan

CMOS compatible 400°C-robust 42 nm perpendicular spin-orbit torque magnetic tunnel junction (p-SOT-MTJ) devices with the tunnel magnetoresistance ratio of 130% is demonstrated for the first time by the interface-enhanced synthetic anti-ferromagnet (SAF) and the improved ion-beam etching. The record high 440oC thermal robustness of SAF is achieved. The SAF field and the magnetic coupling between Co/Pt multilayer and reference layer are enhanced by the magnet-coupling fcc-texture multilayer (MCFTM) buffer. The Pt-Fe inter-diffusion during thermal stress is effectively reduced by the W(3Å)-based texture-decoupling diffusion multi-barrier (TDDMB) for magnetic field immunity. The composite SOT channel of TaN/W and Ta/W breaks the thickness limitation of beta-W (< ~5 nm) and enlarges the MTJ etching window. The TaN/W channel exhibits the large effective spin Hall angle of ~-0.27. The deterministic field-free SOT writing with spin-transfer torque (STT) assist is achieved. The size dependence on STT-assisted SOT switching is investigated using micromagnetic simulation.

SESSION 12

STT-MRAM [Room 5]

Friday, June 18, 9:10-10:00

Chairpersons: Y.-C. Yeo, TSMC
S.-C. Song, Qualcomm, Inc.

T12-1 - 9:10
Reliability and Magnetic Immunity of Reflow-Capable Embedded STT-MRAM in 16nm FinFET CMOS Process, C.-H. Chen, C.-Y. Chang, C.-H. Weng, T.-H. Kuo, C.-Y. Wang, M.-C. Shih, T.-W. Chiang, Y.-J. Lee, R. Wang, K.-H. Shen, A. Hung and H. Chuang, TSMC, Taiwan

We demonstrate the reliability and magnetic immunity of STT-MRAM embedded in 16nm FinFET CMOS process. The technology supports endurance cycles up to 10^5 for wide temperature range from -40°C to 125°C with low bit error rate and passes 10^6 cycles at the worst temperature case of -40°C. Data retention sustains three solder-reflow cycles and up to 10 years with less than 1ppm error rate at 234°C. Read disturb error rate is less than 10^{-20} per read. Magnetic immunity of standby and active mode can reach 550Oe for 10 years 1ppm error rate and 800Oe for 0.1ppm error rate per write at 125 °C, respectively.

T12-2 - 9:20
STT-MRAM: A Robust Embedded Non-Volatile Memory with Superior Reliability, and Immunity to External Magnetic Field and RF Sources, V. B. Naik, K. Yamane, J. H. Kwon, S. K, J. H. Lim, Z. Ali Syed Mohammed, B. Behin-Aein, N. L. Chung, L. Y. Hau, R. Chao, C. S. Chiang, Y. Huang, L. Pu, Y. Otani, H. Dixit, S. H. Jang, N. Balasankaran, F. Tan, W. P. Neo, L. C. Goh, E. H. Toh, T. Ling, J. W. Ting, H. Yoon, G. Congedo, J. Mueller, B. Pfefferling, O. Kallensee, A. Vogel, T. Merbeth, C. S. Seet, J. Wong, J. Bordelon Jr, Y. S. You, S. Soss, T. H. Chan, E. Quek and S. Y. Siah, GLOBALFOUNDRIES Singapore Pte. Ltd., Singapore

We report STT-MRAM's robustness, superior reliability and immunity performance to external magnetic field and RF sources for next-generation embedded-MRAM (eMRAM) technology based on 22FDX+RF+MRAM. Using 40Mb eMRAM integrated in 22FDX, we demonstrate Write/Read repeatability performance with BER variation of <0.2 PPM with ECC-OFF and zero failure with ECC-ON at -40~125°C. We show angle dependence of stand-by magnetic immunity and ways to improve it with optimized shielding designs. Using a specially designed RF-MRAM EMI chip, we demonstrate no impact of RF interference on MRAM and vice versa. Furthermore, using optimized MTJ process, design and testing schemes, we show <2 PPM post 5x solder reflow performance and zero mean endurance failure after 100K cycles at -40°C.

T12-3 - 9:30
12.5 Mb/mm^2 Embedded MRAM for High Density Non-Volatile RAM Applications, K. Suh*, J. Lee**, H. Shin**, J. Lee*, K.-M. Lee**, Y. Hong**, S. Han*, Y.-J. Kim*, C. Kim*, S. Pyo*, H. Jung*, Y. Ji*, B. Seo*, S. Hwang*, K. Lee**, D. Kim**, Y. Kim**, S. Oh**, D.-e. Jeong**, K. Nam*, B. Kwon*, M. Cho*, Y. Song**, K. Lee*, Y. Hong* and G. Jeong*, *Samsung Electronics Co., Ltd. and **Samsung Electronics Co., Ltd., Korea

We demonstrate 28-nm embedded MRAM (eMRAM) macro for non-volatile RAM (nvRAM) applications, featuring macro density of 12.5 Mb/mm2, write speed < 100 ns, 10-yrs retention at 125 °C, and endurance > 1E9 cycles.

T12-4 - 9:40
Advanced 18 nm Quad-MTJ Technology Overcomes Dilemma of Retention and Endurance Under Scaling Beyond 2X nm, H. Naganuma, S. Miura, H. Honjo, K. Nishioka, T. Watanabe, T. Nasuno, H. Inoue, A. Nguyen, Y. Endo, Y. Noguchi, M. Yasuhira, S. Ikeda and T. Endoh, Tohoku Univ., Japan

Advanced quad-interfaces perpendicular magnetic tunnel junction (Quad-MTJ) was developed by engineering a low effective damping constant material in free layer with high perpendicular magnetic anisotropy (PMA), and low resistance area product (*RA*) in MgO layers, and stable reference layer. The advanced 18 nm Quad-MTJ fabricated by the developed low-damage 300 mm fabrication process exhibited following performances over those of Double-MTJ; (a) 1.77 times larger thermal stability factor (Δ), (b) 0.83 times smaller writing current (I_c) at 10 ns, (c) 2.1 times higher write efficiency (Δ/I_c) at 10 ns. Thanks to the above excellent MTJ stack design, it is the first time beyond 2X nm generation that the advanced 18 nm Quad-MTJ achieves at least 6×10^{11} endurance with 10 years retention. Consequently, the advanced Quad-MTJ technologies have broken out the dilemma issue of retention and endurance even under scaling of 2X nm.

T12-5 - 9:50
22 nm Embedded STT-MRAM Macro with 10 ns Switching and >10^{14} Endurance for Last Level Cache Applications, Z. Wang, X. Hao, L. Hu, D. Jung, W. Kim, Z. Wei, L. Wang, K. Satoh, J. Zhang and Y. Huai, Avalanche Technology, USA

We demonstrate high performance 22 nm embedded STT-MRAM with a distinct combination of 10 ns write speed and >10^{14} endurance at chip level. This is achieved by developing a unique MTJ free layer that exhibits high H_k, low moment and low damping, which dramatically reduced switching current at short pulse widths. We further show that this MTJ design meets data retention requirement of 10 years at 105°C.

Technology / Circuits Joint Focus Session 3

Photonics Interconnect and Compute [Room 6]

Friday, June 18, 8:40-9:50

Chairpersons: Y. Shiratori, Nippon Telegraph and Telephone Corp. (NTT) / M. Takenaka, The Univ. of Tokyo
T. Letavic, GLOBALFOUNDRIES Inc.

JFS3-1 - 8:40 (Invited)
On-Silicon Photonic Integrated Circuit toward On-Chip Interconnection and Distributed Computing, N. Nishiyama and T. Amemiya, Tokyo Institute of Technology, Japan

Heterogeneous material integration technology gives us freedom of material choices in both electronic and photonic devices. In this presentation, status, technology and characteristics of photonic devices in photonic integrated circuits (PICs) on Si (SOI) will be reviewed. Membrane (thin III-V film) PICs can realize low power consumption data transmission on Si substrate. This PICs can be applicable to on-chip interconnection to reduce power dissipation under higher speed transmission. 93 fJ/bit transmission with 20 Gbps has been demonstrated. Hybrid PICs were also demonstrated to realize 10-Tbps-class transceiver with low energy cost for distributed computing. This structure can integrate multiple function and many array devices in one chip. Also, by dense integration, some function of electronics can be moved to photonics part. This enables power consumption reduction.

JFS3-2 - 8:50 (Invited)
Monolithic Microring-Based WDM Optical I/O for Heterogeneous Computing, M. Wade, D. Jeong, B. Kim, M. Zhang, W. Bae, C. Zhang, P. Bhargava, D. Van Orden, S. Ardalan, C. Ramamurthy, E. Anderson, A. Katzin, H. Lu, S. Buchbinder, B. Beheshtian, A. Khilo, M. Rust, C. Li, F. Sedgwick, J. Fini, R. Meade, V. Stojanović and C. Sun, Ayar Labs Inc, USA

For the first time, we demonstrate an error-free, 128Gbps (8x16Gbps) optical transceiver using a microring-based wavelength-division multiplexed (WDM) architecture. The optical transceiver ran for 12 hours with zero errors, resulting in a measured bit-error rate of <1.45e-15 per optical lane. The total number of bits sent during this time was ~691 terabits per lane and ~5.5 petabits aggregate across all lanes.

JFS3-3 - 9:00
Graphene Electro-Absorption Modulators Integrated at Wafer-Scale in a CMOS Fab, C. H. Wu*,**, S. Brems*, D. Yudistira*, D. Cott*, A. Milenin*, K. Vandersmissen*, A. Maestre***, A. Centeno***, J. Van Campenhout*, C. Huyghebaert* and M. Pantouvaki*, *imec, **Ghent Univ., Belgium and ***Graphenea Semiconductor SLU, Spain

We demonstrate graphene electro-absorption modulators (EAM) integrated on 300mm wafers. The integration is based on imec's 300mm silicon photonics platform and the full integration sequence is using standard CMOS production tools expect for the 6-inch CVD graphene growth and transfer, transferred by Graphenea. 164x TE EAMs were measured per wafer and demonstrate 90% yield with modulation efficiency (ME) of 41dB/mm for 8V voltage swing, after process optimization. The 3dB bandwidth of the EAMs is 14.9GHz for the device with 50um active length. Both parameters show comparable performance with lab-based devices, obtained on coupons using similar CVD graphene. This work paves the way to enable high-volume manufacturing of 2D-material-based photonics devices.

JFS3-4 - 9:10
Silicon Photonic Micro-Ring Modulator-Based 4 x 112 Gb/s O-Band WDM Transmitter with Ring Photocurrent-Based Thermal Control in 28nm CMOS, J. Sharma, H. Li, Z. Xuan, R. Kumar, C.-M. Hsu, M. N. Sakib, P. Liao, H. Rong, J. Jaussi and G. Balamurugan, Intel Corp., USA

We present a 4λ x 112 Gb/s/Lambda hybrid-integrated silicon photonic TX suitable for 400G Ethernet modules and co-packaged optics. The photonic IC (PIC) uses cascaded micro-ring modulators (MRMs) with integrated heaters for efficient wavelength division multiplexing (WDM). The 28nm CMOS electronic IC includes PAM4 MRM drivers with nonlinear FFE and control circuits to stabilize MRM performance against process and temperature variations. A thermal control scheme based on sensing MRM photocurrents is used to minimize monitoring hardware in the PIC. Measured results demonstrate 112 Gb/s PAM4 operation with <0.7 dB TDECQ from each of the 4 channels. To our best knowledge, this is the highest per-λ data rate reported for an O-band ring-based WDM transmitter.

JFS3-5 - 9:20
First InGaAs/InAlAs Single-Photon Avalanche Diodes (SPADs) Heterogeneously Integrated with Si Photonics on SOI Platform for 1550 nm Detection, J. Zhang, H. Xu, G. Zhang, Y. Chen, H. Wang, K. H. Tan, S. Wicaksono, C. Wang, C. Sun, Q. Kong, C. C. W. Lim, S. F. Yoon and X. Gong, National Univ. of Singapore, Singapore

For the first time, heterogeneous integration of InGaAs/InAlAs singlephoton avalanche diodes (SPADs) with Si photonics was realized and demonstrated through a temperature die-to-die bonding technique. Together with the adoption of a triple-mesa structure in SPADs which not only avoids the surface exposure to the high electric field but also alleviate the electric field crowding at mesa edges, our integrated SPADs exhibit high single-photon detection efficiency (SPDE) of 22% and low dark count rate (DCR) of 8.6 $\times 10^{5}$ Hz, which are among the best performance reported for InGaAs/InAlAs SPADs, and are approaching that of InGaAs/InP SPADs. High device yield and performance uniformity were also achieved.

JFS3-6 - 9:30
Bandgap-Tunable III-V-OI Photonics Platform with Quantum Well Intermixing for Versatile Active-Passive Integration of Chip-Scale Photonic Integrated Circuits, N. Sekine, K. Toprasertpong, S. Takagi and M. Takenaka, The Univ. of Tokyo, Japan

We investigated monolithic active-passive integration on III-V-OI photonics platform using quantum well intermixing for photonic integrated circuits. We developed the void-free wafer bonding technique and established large bandgap modulation through quantum well intermixing using low-energy hot P_2^+ molecular ion implantation. We monolithically integrated passive waveguides, waveguide PDs, electro-absorption (EA) modulators, optical switches with carrier-injection optical phase shifters on III-V-OI operating at 1.55 um wavelength.

JFS3-7 - 9:40
First Demonstration of Waveguide-Coupled Ge0.92Sn0.08 Ge Multiple-Quantum-Well Photodetector on the SOI Platform for 2000 nm Wavelength Optoelectronic Integrated Circuit, H. Wang*, Y. Chen*, J. Zhang*, G. Zhang*, Y.-C. Huang** and X. Gong*, *National Univ. of Singapore, Singapore and **Applied Materials, Inc., USA

We report the first demonstration of a silicon-on-insulator (SOI) waveguide-coupled $Ge_{0.92}Sn_{0.08}$ Ge multiple-quantum-well (MQW) photodiode (PD) for 2 um wavelength using a flip-chip bonding technology. The light in the waveguide couples to the PD for detection via a grating coupler. The grating coupler and waveguide were designed and fabricated on the standard SOI wafer for 2 um and bonded with the GeSn Ge PDs. On the same wafer, back illuminated GeSn/Ge PDs were also integrated using the same technology for free space optical detection. Our waveguide-coupled PD exhibits responsivity of 10.3 mA/W at 2 um wavelength and one of the lowest dark current densities of 38.4 mA cm^2 for Ge1-xSnx PDs. In addition, no degradation of the dark current was found after the bonding.

Technology / Circuits Joint Focus Session 4

Image Sensors [Room 1]

Saturday, June 19, 8:40-10:00

Chairpersons: T. Takahashi, Sony Semiconductor Solutions Corp.
A. Thomsen, Cirrus Logic, Inc.

JFS4-1 - 8:40 (Invited)
A CMOS Image Sensor and an AI Accelerator for Realizing Edge-Computing-Based Surveillance Camera Systems, F. Morishita, N. Kato, S. Okubo, T. Toi, M. Hiraki, S. Otani, H. Abe, Y. Shinohara and H. Kondo, Renesas Electronics Corp., Japan

This paper presents a CMOS image sensor and an AI accelerator to realize surveillance camera systems based on edge computing. For CMOS image sensors to be used for surveillance, it is desirable that they are highly sensitive even in low illuminance. We propose a new timing shift ADC used in CMOS image sensors for improving high sensitivity performance. Our proposed ADC improves non-linearity characteristics under low illuminance by 63%. Achieving power-efficient edge computing is a challenge for the systems to be used widely in the surveillance camera market. We demonstrate that our proposed AI accelerator performs inference processing for object recognition with 1 TOPS/W.

JFS4-2 - 8:50
All-Directional Dual Pixel Auto Focus Technology in CMOS Image Sensors, E. S. Shim, Samsung Electronics Co., Ltd., Korea

We developed a dual pixel with accurate and all-directional auto focus (AF) performance in CMOS image sensor (CIS). The optimized in-pixel deep trench isolation (DTI) provided accurate AF data and good image quality in the entire image area and over whole visible wavelength range. Furthermore, the horizontal-vertical (HV) dual pixel with the slanted in-pixel DTI enabled the acquisition of all-directional AF information by the conventional dual pixel readout method. These technologies were demonstrated in 1.4um dual pixel and will be applied to the further shrunken pixels.

JFS4-3 - 9:00
Development of Advanced Inter-Color-Filter Grid on Sub-Micron-Pixel CMOS Image Sensor for Mobile Cameras with High Sensitivity and High Resolution, J. In-Sung, Y. Lee, H. Y. Park, J. U. Kim, D. Kang, T. Kim, M. Kim, K. Lee, M. Heo, I. Ro, J. Kim, I. Park, S. Kwon, K. Yoon, D. Park, C. Lee, E. Jo, M. Jeon, C. Park, K. R. Byun, C. K. Chang, J. S. Hur, K. Yoon, T. Jeon, J. Lee, J. Park, B. Kim, J. Ahn, H. Kim, C.-R. Moon and H.-S. Kim, Samsung Electronics Co., Ltd., Korea

Sub-micron pixels have been widely adopted in recent CMOS image sensors to implement high resolution cameras in small form factors, i.e. slim mobile-phones. Even with shrinking pixels, customers demand higher image quality, and the pixel performance must remain comparable to that of the previous generations. Conventionally, to suppress the optical crosstalk between pixels, a metal grid has been used as an isolation structure between adjacent color filters. However, as the pixel size continues to shrink to the sub-micron regime, an optical loss increases because the focal spot size of the pixel's microlens does not downscale accordingly with the decreasing pixel size due to the diffraction limit: the light absorption inevitably occurs in the metal grid. For the first time, we have demonstrated a new lossless, dielectric-only grid scheme. The result shows 29 % increase in sensitivity and +1.2-dB enhancement in Y-SNR when compared to the previous hybrid metal-and-dielectric grid.

JFS4-4 - 9:10

A 2.6 E-Rms Low-Random-Noise, 116.2 mW Low-Power 2-Mp Global Shutter CMOS Image Sensor with Pixel-Level ADC and In-Pixel Memory, M.-W. Seo, M. Chu, H.-Y. Jung, S. Kim, J. Song, J. Lee, S.-Y. Kim, J. Lee, S.-J. Byun, D. Bae, M. Kim, G.-D. Lee, H. Shim, C. Um, C. Kim, I.-G. Baek, D. Kwon, H. Kim, H. Choi, J. Go, J. Ahn, J.-k. Lee, C. Moon, K. Lee and H.-S. Kim, Samsung Electronics Co., Ltd., Korea

This paper presents a low-random noise of 2.6 e-rms, a low-power of 116.2 mW at video rate, and a high-speed up to 960 fps 2-mega pixels global-shutter type CMOS image sensor (CIS) using an advanced DRAM technology. To achieve a high performance global-shutter CIS, we proposed a novel architecture for the digital pixel sensor which is a remarkable global-shutter operation CIS with a pixel-wise ADC and an in-pixel digital memory. Each pixel has two small-pitch Cu-to-Cu interconnectors for the wafer-level stacking, and the pitch of each unit pixel is less than 5 um which is the world's smallest pixel embedding both pixel-level ADC and 22-bit memories.

JFS4-5 - 9:20

A Photon-Counting 4Mpixel Stacked BSI Quanta Image Sensor with 0.3e- Read Noise and 100dB Single-Exposure Dynamic Range, J. Ma, D. Zhang, O. Elgendy and S. Masoodian, Gigajot Technology Inc., USA

This paper reports a 4Mpixel, 3D-stacked backside illuminated Quanta Image Sensor (QIS) with 2.2um pixels that can operate simultaneously in photon-counting mode with deep sub-electron read noise (0.3e- rms) and linear integration mode with large full-well capacity (30k e-). A single-exposure dynamic range of 100dB is realized with this dual-mode readout under room temperature. This QIS device uses a cluster-parallel readout architecture to achieve up to 120fps frame rate at 550mW power consumption.

JFS4-6 - 9:30

A 5.1ms Low-Latency Face Detection Imager with In-Memory Charge-Domain Computing of Machine-Learning Classifiers, H. Song*, S. Oh*, J. Salinas*, S.-Y. Park** and E. Yoon*, *Univ. of Michigan, USA and **Pusan National Univ., Korea

We present a CMOS imager for low-latency face detection empowered by parallel imaging and computing of machine-learning (ML) classifiers. The energy-efficient parallel operation and multi-scale detection eliminate image capture delay and significantly alleviate backend computational loads. The proposed pixel architecture, composed of dynamic samplers in a global shutter (GS) pixel array, allows for energy-efficient in-memory charge-domain computing of feature extraction and classification. The illumination-invariant detection was realized by using log-Haar features. A prototype 240x240 imager achieved an on-chip face detection latency of 5.1ms with a 97.9% true positive rate and 2% false positive rate at 120fps. Moreover, a dynamic nature of in-memory computing allows an energy efficiency of 419pJ_pixel for feature extraction and classification, leading to the smallest latency-energy product of 3.66ms.nJ_pixel with digital backend processing.

JFS4-7 - 9:40

A CMOS LiDAR Sensor with Pre-Post Weighted-Histogramming for Sunlight Immunity Over 105 klx and SPAD-Based Infinite Interference Canceling, S. Hyeongseok, Sungkyunkwan Univ., Korea

This paper presents a CMOS LiDAR sensor with high background noise (BGN) immunity. The sensor has on-chip pre-post weighted histogramming to detect only time-correlated time-of-flight (TOF) out of BGN from both sunlight and exponentially increased dark noise while enhancing sensitivity through higher excess voltage (V_{ex}) of SPADs. The sensor also employs a SPAD-based random number generator (SRNG) for canceling interference (IF) from an infinite number of LiDARs. The sensor shows 8.08 cm accuracy for the range of 32 m under high BGN (105 klx sunlight and 48.72 kcps dark-count rate with increased V_{ex}).

JFS4-8 - 9:50

Advanced Multi-NIR Spectral Image Sensor with Optimized Vision Sensing System and Its Impact on Innovative Applications, H. Sumi*,**, H. Takehara**, J. Ohta** and M. Ishikawa*, *The Univ. of Tokyo and **Nara Institute of Science and Technology, Japan

Innovative applications with multiple near-infrared (multi-NIR) spectral CMOS image sensors (CIS) and camera systems have recently been developed. The multi-NIR filter is an indispensably key technology in practical of using the multi-NIR camera system in consumer camera. Advanced processing technology for multi-NIR signals has been developed using a Fabry-Perot structure. Three types of NIR wavelength filters are formed as a Bayer pattern with 2-x-2um^2 pixel size on a 5-M pixel BSI-CIS. The thickness differences of the three types of bandpass filters are suppressed to less than 75 nm. To enable applications in surveillance, automobiles, and fundus cameras for health management, signal processing technology has also been developed that processes and mixes each signal of a multi-NIR signal with low-intensity visible light images. This provides good image SNR (Signal-to-Noise Ratio) under low lighting conditions of 0.1 lux or less allowing changes of state to be easily identified.

SESSION 13

PCM and ReRAM [Room 4]

Saturday, June 19, 8:40-9:30

Chairpersons: H.-T. Lue, Macronix International Co., Ltd.
V. Narayanan, IBM Corp.

T13-1 - 8:40
A New BJT Selector for Sub 0.02µm² High Density Embedded PCM Memory in FDSOI CMOS Technology, O. Weber*, C. Pigot*, R. Berthelon*, A. Gandolfo**, P. Mattavelli**, J. Jasse**, G. Samanni**, E. Gomiero**, E. Richard*, J.-C. Grenier*, R. Ranica*, S. Chouteau*, R. Beneyton*, B. Duclaux*, C. Beylier*, D. Pelissier*, C. Gallon*, C. Toulouse*, C. Jenny*, M. Croisy*, C. Borowiak*, S. Haendler*, J.-L. Ogier*, E. Batail*, F. Gilibert*, P. Boivin*, D. Turgis*, A. Conte**, F. Disegni**, A. Redaelli**, R. Annunziata**, P. Cappelletti**, F. Piazza*, P. Ferreira* and F. Arnaud*, *STMicroelectronics N.V., France and **STMicroelectronics N.V., Italy

Fabrication and electrical characteristics of a new BJT selector enabling a 1T1R embedded phase change material (ePCM) memory cell of 0.019µm² are extensively reported in this paper. A smart process, leveraging the specific feature of the FDSOI substrates with its thin buried oxide (BOX) has been developed to create an innovative isolation wall between bitlines (BL) and to totally suppress the parasitic BL-to-BL leakage. Compared to its MOS selector counterpart, -48% cell area reduction is obtained at same driving current, demonstrating the high density and the cost competitiveness of this FDSOI BJT selector solution. Finally, a shrunk BJT-ePCM cell of 0.015µm², which is the smallest 1T1R eNVM reported to date, is demonstrated for the first time.

T13-2 - 8:50
MLC PCM Techniques to Improve Nerual Network Inference Retention Time by 10^5X and Reduce Accuracy Degradation by 10.8X, W.-S. Khwa*, K. Akarvardar**, Y.-S. Chen*, Y.-C. Chiu*, J.-C. Liu*, J.-J. Wu*, H.-Y. Lee*, S.-M. Yu*, C.-H. Lee*, T.-C. Chen*, Y.-C. Lin*, C.-F. Hsu*, T.-Y. Lee*, T.-K. Ku*, C.-H. Kuo*, J.-Y. Wu**, X.-Y. Bao*, C.-S. Chang*, Y.-D. Chih**, H.-S. P. Wong** and M.-F. Chang*, *TSMC, Taiwan and **TSMC, USA

We present three novel MLC PCM techniques - (1) device requirement balancing, (2) prediction-based MSB-biased referencing, and (3) bit-prioritized placement to address the MLC device challenges in neural network applications. Using measured MLC bit error rates, the proposed techniques can improve the MLC PCM retention time by 10^5 times while keeping the ResNet-20 inference accuracy degradation within 3% and reduce the accuracy degradation by 91% (10.8X) for CIFAR-100 dataset in the presence of temporal resistance drift.

T13-3 - 9:00
Fully On-Chip MAC at 14nm Enabled by Accurate Row-Wise Programming of PCM-Based Weights and Parallel Vector-Transport in Duration-Format, P. Narayanan*, S. Ambrogio*, A. Okazaki**, K. Hosokawa**, H. Tsai*, A. Nomura**, T. Yasuda**, C. Mackin*, S. C. Lewis****, A. Friz*, M. Ishii**, Y. Kohda**, H. Mori**, K. Spoon*, R. Khaddam-Aljameh***, N. Saulnier*****, M. Bergendahl*****, J. Demarest*****, K. Brew*****, V. Chan*****, S. Choi*****, I. Ok*****, I. Ahsan*****, F. L. Lie*****, W. Haensch****, V. Narayanan**** and G. W. Burr*, *IBM Research - Almaden, USA, **IBM Research - Tokyo, Japan, ***IBM Research - Zurich, Switzerland, ****IBM T. J. Watson Research Center and *****IBM Albany NanoTech, USA

We report on *ARES* - a 14nm Phase Change Memory (PCM) based test chip comprising multiple crossbar tiles, each capable of parallel Multiply ACCumulate (MAC) inference on 512x512 unique weights. A massively parallel 2D mesh transports Deep Neural Network (DNN) excitations in duration format across the chip, between tiles and integrated Landing Pads (LPs) where digital data enters and leaves the chip. For accurate weight programming (<3% weight error), we employ a row-wise programming scheme that efficiently programs the 4 PCM devices in each analog weight with minimal overshoot. We implement two DNNs at near software-equivalent accuracy, demonstrating tile-to-tile transport with a fully on-chip 2-layer network, and testing resilience to error propagation with a recurrent LSTM network, using off-chip activation functions before looping back to the next on-chip MAC.

T13-4 - 9:10
One-Shot Learning with Memory-Augmented Neural Networks Using a 64-Kbit, 118 GOPS/W RRAM-Based Non-Volatile Associative Memory, H. Li, W.-C. Chen, A. Levy, C.-H. Wang, H. Wang, P.-H. Chen, W. Wan, H.-S. P. Wong and P. Raina, Stanford Univ., USA

Learning from a few examples (one/few-shot learning) on the fly is a key challenge for on-device machine intelligence. We present the first chip-level demonstration of one-shot learning using a 2T-2R resistive RAM (RRAM) non-volatile associative memory (AM) as the backend of memory-augmented neural networks (MANNs). The 64-kbit fully integrated RRAM-CMOS AM core (0.2 mm² at 40 nm node) enables long-term feature embedding and retrieval, demonstrated in a challenging 32-way one-shot learning task using Omniglot dataset. Using only one example per class for 32 unseen classes during on-chip learning, our AM chip achieves ~72% measured inference accuracy on Omniglot as the first chip accuracy report compared to software accuracy (~82%), while reaching 118 GOPS/W for in-memory L1 distance computation and prediction.

T13-5 - 9:20
A 14nm 100Kb 2T1R Transpose RRAM with >150X Resistance Ratio Enhancement and 27.95% Reduction on Energy-Latency Product Using Low-Power Near Threshold Read Operation and Fast Data-Line Current Stabling Scheme, L. Wang*,******, W. Ye*,******, J. Lai*,******, J. Liu*, J. Yang*,***, X. Si****, C. Huo*, C. Dou*, X. Xu*, Q. Liu*,**, D. Shang*, F. Zhang*, H. Lv*, M.-F. Chang****, H. Iwai***** and M. Liu*,**, *Institute of Microelectronics of the Chinese Academy of Sciences, **Frontier Institute of Chip and System, Fudan Univ., ***Zhejiang Laboratory, China, ****National Tsing Hua Univ., Taiwan, *****Tokyo Institute of Technology, Japan and ******Univ. of Chinese Academy of Sciences, China

This study proposes an 2T1R Transpose RRAM (T-RRAM) macro supports highly efficient transpose accessibility featuring (1) a 2T1R cell with low-power near-threshold-voltage (NTV) read operations for resistance ratio (R-ratio) enhancement (>150X) and read disturb suppression, and (2) a customized macro structure with fast data-line current stabling scheme (FDCS) to reduce the energy-latency product (27.95%). A 100Kb 2T1R T-RRAM macro is silicon verified using 14nm CMOS process with TaOx-based RRAM. This paper firstly demonstrates a 14nm T-RRAM with large R-ratio, small area overhead and improved energy-latency product.

SESSION 14

New Memory [Room 4]

Saturday, June 19, 9:30-10:00

Chairpersons: Y. Sasago, Hitachi, Ltd.
S. Yu, Georgia Institute of Technology

T14-1 - 9:30
Excellent Synapse Characteristics of 50 nm Vertical Transistor with WO_x Channel for High Density Neuromorphic System, C. Lee, W. Choi, M. Kwak, S. Kim and H. Hwang, POSTECH, Korea

We report a WO_x based vertical synapse transistor (ST) with excellent device characteristics such as small device area ($4F^2$), large conductance range (> 10^5), good retention (~ 8 hours) and excellent weight update property. With scaling the device dimension down to 50 nm, significant improvement (>10^3) of switching speed was obtained. To further scale down the device area ($2F^2$) and store G^+/G^- weights, two ST were vertically stacked. We have confirmed that weight update can be maintained without disturbing neighbor cell.

T14-2 - 9:40
OTS-Based Analog-to-Stochastic Converter for Fully-Parallel Weight Update in Cross-Point Array Neural Networks, M. Kwak, S. Lee, W. Choi, C. Lee, S. Kim and H. Hwang, POSTECH, Korea

In this study, we experimentally demonstrate a highly linear (R^2=0.995) and area-efficient 1S1R analog-to-stochastic converter (ASC) which is a core enabler for fully-parallel weight update operation in cross-point synaptic array-based neural networks. We confirm that the previously reported, sigmoid-like ASCs cannot be used for stochastic updates owing to the limitation of nonlinear characteristics. We analyze the characteristics of the device-based ASC for stochastic update operation to show that our ASC consumes 1,000 times less power than a CMOS-based ASC and does not require a power-hungry digital-to-analog converter (DAC). A software-level image recognition accuracy (97.96%) is achieved when performing neural network training with our ovonic threshold switching (OTS) device-based ASC.

T14-3 - 9:50
Low-Voltage (~1.3V), Arsenic Free Threshold Type Selector with Ultra High Endurance (> 10^{11}) for High Density 1S1R Memory Array, C. H. Wu, C. M. Lee, Y. S. Chen, H. Y. Lee, E. Ambrosi, C. F. Hsu, S. Vaziri, Y. Y. Chen, C. H. Nien, R. L. Hwang, P. J. Liao, D. H. Hou, Y.-H. Lee, T. Y. Lee, T. C. Chen, M. F. Chang, H.-S. P. Wong and X. Y. Bao, TSMC, Taiwan

Low voltage selectors are critical for low power operation of high density non-volatile memories. In this work, selectors based on arsenic free chalcogenide materials are demonstrated with record high endurance over 1011 cycles together with threshold voltage ~1.3V and leakage current ~5nA. The enhanced endurance is attributed to suppression of phase separation with more stable amorphous network by proper dopants.

SESSION 15

Nanosheet and DTCO [Room 5]

Saturday, June 19, 8:40-9:30

Chairpersons: Y. Masuoka, Samsung Electronics Co., Ltd.
Y. Liang, NVIDIA Corp.

T15-1 - 8:40
External I/O Interfaces in Sub-5nm GAA NS Technology and STCO Scaling Options, W.-C. Chen*,**, S.-H. Chen*, G. Hellings*, E. Bury*, M. Simicic*, Z. Wu*,**, G. Van der Plas*, G. Groeseneken*,** and E. Beyne*, *imec and **KU Leuven, Belgium

In this paper, the challenges of the I/O development roadmap are discussed. The impact of I/O application in FEOL scaling and 3D integration are evaluated. A cost-efficient circuit solution to the I/O implementation in a gate-all-around (GAA) nanosheet (NS) technology is proposed. The functionality verification and the relevant reliability concerns are assessed in a mature industrial CMOS process. Finally, to foresee the design strategies in future STCO scaling era, the impact of the fully back-side (BS) connections on I/O performance is investigated. Capacitance is doubled compared to front-side (FS) I/O due to the contributions from BS connection layers. The layout techniques can mitigate ~30% of the extra capacitance and the technology option of the deep trench isolation (DTI) is considered to reduce the extra capacitance to only ~8%.

T15-2 - 8:50
First Highly Stacked $Ge_{0.95}Si_{0.05}$ nGAAFETs with Record I_{ON} = 110 µA (4100 µA/µm) at V_{OV}=V_{DS}=0.5V and High $G_{m,max}$ = 340 µS (13000 µS/µm) at V_{DS}=0.5V by Wet Etching, Y.-C. Liu, C.-T. Tu, C.-E. Tsai, Y.-R. Chen, J.-Y. Chen, S.-R. Jan, B.-W. Huang, S.-J. Chueh, C.-J. Tsen and C. W. Liu, National Taiwan Univ., Taiwan

The 8-stacked $Ge_{0.75}Si_{0.25}$ nanosheets and the 7-stacked $Ge_{0.95}Si_{0.05}$ nanowires are realized by H_2O_2 wet etching. High inter-channel uniformity of the 8-stacked $Ge_{0.75}Si_{0.25}$ is demonstrated. Thanks to small transport effective mass (m_t) and large DOS effective mass (m_{DOS}) in L4 valley, and low $R_{S/D}/R_{total}$, high performance of the 7-stacked $Ge_{0.95}Si_{0.05}$ is demonstrated. The record I_{ON}=110µA per stack (4100µA/µm per channel footprint) at V_{OV}=V_{DS}=0.5V and high $G_{m,max}$=340µS (13000µS/µm) at V_{DS}=0.5V are achieved among reported Ge/GeSi 3D nFETs.

T15-3 - 9:00
First Demonstration of Multi-V_T Stacked $Ge_{0.87}Sn_{0.13}$ Nanosheets by Dipole-Controlled ALD WN_xC_y Work Function Metal with Low Resistivity and Thermal Budget ≤ 400 °C, C.-E. Tsai, Y.-R. Chen, C.-T. Tu, Y.-C. Liu, J.-Y. Chen and C. W. Liu, National Taiwan Univ., Taiwan

Dipole-controlled WN_xC_y work function metal (WFM) by plasma-enhanced atomic layer deposition (PEALD) is integrated with stacked GeSn nanosheet transistors to enable multi-VT with a wide tunability of 510 mV at the low process temperature ≤ 400 °C. The effective work function (EWF) of WN_xC_y is modulated by the relative thicknesses of pWFM and nWFM while maintaining the constant total thickness. EWF tuning of 500 meV around Si midgap is obtained, and the best achievable resistivity of ALD WN_xC_y is 990 µΩ-cm, much lower than industrial ALD TiAl value (2000 µΩ-cm [1]). Multi-V_T stacked $Ge_{0.87}Sn_{0.13}$ nanosheets are demonstrated by controlling the different group electronegativity and dipole strength at nWFM/WN_xC_y:O interface.

T15-4 - 9:10
Challenges and Opportunities for Stacked Transistor: DTCO and Device, J. Wang, S. D. Suk, A. Chu, T. Hook, A. Young, R. Krishnan, R. Bao, I. Seshadri, B. Senapati, V. Zalani, T. Philip, H. Zhou, K. Zhao, D. Dechene, B. Haran, D. Guo and H. Bu, IBM Research, USA

We report our investigation of challenges and opportunities for vertically stacked transistors, with a focus on block-level scaling and device performance consideration. At the standard cell level, our DTCO innovation of splitting the power rails can free up a metal track for signal routing. Implementing two circuit rows for complex cells also increases available tracks. At the block level, we performed the routing study through PnR and overcome the shortage of pin access by DTCO innovations, achieving 0.55x area scaling vs non-stack technology. For device design, choices of device structures and materials are evaluated. SiGe FinFET(p) on Nanosheet(n) is identified as a strong candidate for vertically stacked device architecture. MOL resistance with one-sided power rail is found to be a bottleneck limiting the device performance. Double-sided power rail is introduced for the first time in this work to effectively address the MOL resistance for stacked transistors.

T15-5 - 9:20
A Novel Framework for DTCO: Fast and Automatic Routability Assessment with Machine Learning for Sub-3nm Technology Options, C. Chidambaram**, A. B. Kahng*, M. Kim*, G. Nallapati**, S. C. Song** and M. Woo*, *Univ. of California, San Diego and **Qualcomm Technologies Inc., USA

We report, for the first time, block-level area scaling that is reversed from cell-level scaling in cell height <150nm regime, using a machine learning (ML)-assisted design-technology co-optimization (DTCO) methodology to optimize block-level area accounting for routability. >400 unique standard-cell architectures are studied, combining cell height (CH, 80~150nm), contacted poly pitch (CPP, 39~57nm), metal pitch (MP, 16~30nm) and use/non-use of buried power rails (BPR). Cell- and design-level routability assessments are performed to obtain routability index (a.k.a. K_{th}) and utilization of designs. CH <120nm with four available tracks increases block-level area due to routing difficulty, showing diminishing return from further pushing of ground rules. BPR improves block-level scaling with CH <120nm, but benefit slows down with CH <100nm. A 1:2 gear ratio (MP:CPP) improves block-level area compared to a 2:3 ratio. This DTCO flow is automated, and the ML-based prediction expedites the process.

SESSION 16

Ferroelectric Devices and Memory -3 [Room 5]

Saturday, June 19, 9:30-10:10

Chairpersons: M. Kobayashi, The Univ. of Tokyo
P. Grudowski, NXP Semiconductors N.V.

T16-1 - 9:30
10^5 Endurance Improvement of FE-HZO by an Innovative Rejuvenation Method for 1z Node NV-DRAM Applications, T. Gong*, L. Tao*, J. Li**, Y. Cheng***, Y. Xu*, W. Wei*, P. Jiang*, P. Yuan*, Y. Wang*, Y. Chen*, Y. Ding*, Y. Yang*, Y. Wang*, B. Chen**, Q. Luo*, S. S. Chung****, S. Du* and M. Liu*, *Chinese Academy of Sciences, **Zhejiang Univ., ***East China Normal Univ., China and ****National Chiao Tung Univ., Taiwan

High operating voltage and low endurance are challenges for FE HZO to be a viable candidate for NV-DRAM technology. In this work, we provide a breakthrough solution for HZO towards 1z node NV-DRAM application. Firstly, the endurance failure mechanism of HZO film under low-electric field (<1.5MV/cm) is systematically investigated by electrical characterizations, DFT calculations and STEM-ABF technique. It is found that fatigue under low-electric field is relevant to the electron de-trapping rather than defect generation. Furthermore, based on the new insight on the failure mechanism, a novel rejuvenation method is proposed. Five orders of endurance enhancement can be achieved. The excellent properties including low operating voltage (1.1V), non-volatile and fairly high endurance (>10^{14}) are quite promising towards 1z node NV-DRAM applications.

T16-2 - 9:40
Critical Role of GIDL Current for Erase Operation in 3D Vertical FeFET and Compact Long-Term FeFET Retention Model, F. Mo*, J. Xiang*, X. Mei*, Y. Sawabe*, T. Saraya*, T. Hiramoto*, C.-J. Su**, V. P-H. Hu*** and M. Kobayashi*, *The Univ. of Tokyo, Japan, **Taiwan Semiconductor Research Institute (TSRI) and ***National Taiwan Univ., Taiwan

We have investigated and revealed the critical role of GIDL current for efficient erase operation in 3D vertical FeFET by developing proper test structures and demonstrated a compact long-term FeFET retention model based on nucleation-limited switching, for the first time. We also proposed novel FeFET process for low voltage operation by controlling oxygen intrusion into the gate stack. This work contributes to the realization of high-density and low-power 3D vertical FeFET.

T16-3 - 9:50
In-Situ* Atomic Visualization of Structural Transformation in $Hf_{0.5}Zr_{0.5}O_2$ Ferroelectric Thin Film: from Nonpolar Tetragonal Phase to Polar Orthorhombic Phase,** Y. Zheng*, C. Zhong*, Y. Zheng*, Z. Gao**,, Y. Cheng*, Q. Zhong*, C. Liu*, Y. Wang*, R. Qi*, R. Huang* and H. Lyu**,***, *East China Normal Univ., **Chinese Academy of Sciences and ***Univ. of Chinese Academy of Sciences, China

For the first time, we *directly visualized the dynamic process of phase transformation* in polycrystalline ferroelectric (FE) $Hf_{0.5}Zr_{0.5}O_2$ (HZO) thin film though *in-situ* spherical aberration (Cs)-corrected transmission electron microscopy (TEM) technique. The main observations are: (1) the dynamic atomic scale structural evolution from centrosymmetric tetragonal (t-) phase to FE orthorhombic (o-) phase under electric field, and (2) the deformation of atomic arrangements in lattice caused by stress is helpful to make the transition happen. These observations provide solid evidence on understanding the fundamental mechanism of the root cause of ferroelectricity in fluorite-type FE materials.

T16-4 - 10:00
Energy Efficient and Robust Reservoir Computing System Using Ultrathin (3.5 nm) Ferroelectric Tunneling Junctions for Temporal Data Learning, J. Yu*,**, L. Yi*,**, W. Sun*,**, W. Zhang*,**, Q. Zuo***, Y. Zhao***, H. Chen****, X. Xu*,**, D. Shang*,** and M. Liu*,**, *Chinese Academy of Sciences, **Other, ***Shanghai IC R&D Center and ****Shanghai Huali Microelectronics Corporation, China

Reservoir computing (RC) can compute temporal data with low training cost. To enhance data processing capability, high dimensionality of reservoir is required, which poses a significant challenge on RC hardware implementation using Si-friendly devices. In this work, for the first time, we use ultra-thin (3.5 nm) ferroelectric tunneling junctions (FTJs) with transient depolarization property as physical nonlinear virtual nodes to address this challenge. By constructing an FTJ-based dynamic reservoir and combining it with RRAM-based binarized readout layer, the high energy efficiency (35 pJ), processing speed (500 ns), and recognition accuracy (92.3%) have been demonstrated for digital sequence classification.

Technology / Circuits Joint Focus Session 5

Circuit and Technology for Quantum Computing [Room 6]

Saturday, June 19, 8:40-9:40

Chairpersons: M. Tada, NanoBridge Semiconductor, Inc.
B. En, Advanced Micro Devices, Inc. (AMD)

JFS5-1 - 8:40 (Invited)
Solid State Qubits: How Learning from CMOS Fabrication Can Speed-Up Progress in Quantum Computing

JFS5-2 - 8:50 (Invited)
Superconducting Quantum Computer: a Hint for Building Architectures, Y. Tabuchi*, S. Tamate** and S. Yorozu*, *Riken and The Univ. of Tokyo, Japan

We discuss the scalability of superconducting quantum computers, especially in a wiring problem. The number of wiring inside a cryostat is almost proportional to the number of qubits in the current wiring architecture. We introduce "The three-Y's": regularity, modularity, and hierarchy to an architecture design of superconducting quantum computers. The key to the wiring elimination is found in the quantum error correction codes having thresholds and spatial translational symmetry, i.e., the surface code. We show a superconducting-digital-logic-based architecture and introduce a stacked heterogeneous structure of the quantum module.

JFS5-3 - 9:00
Uniform Spin Qubit Devices with Tunable Coupling in an All-Silicon 300 mm Integrated Process, N. Dumoulin Stuyck*,**, R. Li*, C. Godfrin*, A. Elsayed*,**, S. Kubicek*, J. Jussot*, B. T. Chan*, F. A. Mohiyaddin*, M. Shehata*,**, G. Simion*, Y. Canvel*, L. Goux*, M. Heyns*,**, B. Govoreanu* and I. P. Radu*, *imec and **KU Leuven, Belgium

Larger arrays of electron spin qubits require radical improvements in fabrication and device uniformity. Here we demonstrate excellent qubit device uniformity and tunability from 300K down to mK temperatures. This is achieved, for the first time, by integrating an overlapping polycrystalline silicon-based gate stack in an 'all-Silicon' and lithographically flexible 300mm flow. Low-disorder Si/SiO_2 is proved by a 10K Hall mobility of $1.5x10^4$ cm^2/Vs. Well-controlled sensors with low charge noise (3.6 ueV/Hz^0.5 at 1 Hz) are used for charge sensing down to the last electron. We demonstrate excellent and reproducible interdot coupling control over nearly 2 decades (2-100 GHz). We show spin manipulation and single-shot spin readout, extracting a valley splitting energy of around 150 ueV. These low-disorder, uniform qubit devices and 300mm fab integration pave the way for fast scale-up to large quantum processors.

JFS5-4 - 9:10

A 5.5mW/Channel 2-to-7 GHz Frequency Synthesizable Qubit-Controlling Cryogenic Pulse Modulator for Scalable Quantum Computers, K. Kang*, B. Kim*, G. Choi**, S.-K. Lee**, J. Choi**,***, J. Lee*, S. Kang*, M. Lee*, H.-J. Song*, Y. Chong**** and J.-Y. Sim*, *POSTECH, **Korea Research Institute of Standards and Science (KRISS), ***Univ. of Science and Technology (UST) and ****Sungkyunkwan Univ., Korea

This work presents a qubit controller IC based on the direct synthesis. The IC consists of six independently-working pulse modulators utilizing the same LO frequency. We propose a sinusoid-shaping nonlinear DAC followed by a linear interpolating DAC to improve both of energy and hardware efficiencies. The implemented IC in 40nm CMOS is verified by superconducting qubit operations with Rabi and Ramsey oscillations while consuming power of < 1/60 compared with the previous state-of-the-art.

JFS5-5 - 9:20

Buried Nanomagnet Realizing High-Speed/Low-Variability Silicon Spin Qubits: Implementable in Error-Correctable Large-Scale Quantum Computers, S. Iizuka, K. Kato, A. Yagishita, H. Asai, T. Ueda, H. Oka, J. Hattori, T. Ikegami, K. Fukuda and T. Mori, AIST, Japan

We propose a buried nanomagnet (BNM) realizing high-speed/low-variability silicon spin qubit operation, inspired by buried wiring technology, for the first time. High-speed quantum-gate operation results from large slanting magnetic- field generated by the BNM disposed quite close to a spin qubit, and low-variation of fidelity thanks to the self-aligned fabrication process. Employing TCAD-based simulation, we demonstrate that the BNM realizes 10 times faster Rabi oscillation (faster spin-flip) than previous works and >99% fidelity under certain process variations. Also, the proposed BNM arrangement is implementable for error-correctable large-scale quantum computers employing a 2D-latticed qubit layout. This technology paves the way to practical large-scale quantum computers with silicon.

JFS5-6 - 9:30

Reconfigurable Germanium Quantum-Dot Arrays for CMOS Integratable Quantum Electronic Devices, I.-H. Wang*, T. Tsai*, R.-C. Pan*, P.-Y. Hong*, M.-T. Kuo**, T. George*, H.-C. Lin* and P.-W. Li*, *National Chiao Tung Univ. and **National Central Univ., Taiwan

We report the first-of-kind scalability and tunability of Ge QDs that are controllably sized, closely coupled, and self-aligned with control gates, using a combination of lithographic patterning, spacer technology, and self-assembled growth. The core experimental design is based on the thermal oxidation of poly-SiGe spacer islands designated at each included-angle location of designed Si_3N_4/c-Si ridge structures. Multiple Ge QDs with good size tunability of 7-20 nm were controllably achieved by adjusting the process times for deposition, etch back and thermal oxidation of poly-SiGe spacer islands. Our Ge QDs array provides a common platform for engineering diverse QD electronic devices with desired reconfigurability and optimizing their performance.