Bonding of Multi-Functional Fiber to a Micro-fabricated Robotic Appendage for Use in Biologic Environments

A. Parrott[2]*, J. Pelster[1]*, Y. Liu[2], Y. Zhang[2], H. Huang[2], J. Kim[2], Q. Liu[1], X. Jia [2], I. Cohen [1]
[1]Cornell University, Ithaca, NY, USA
[2]Virginia Tech, Blacksburg, VA, USA

***Abstract*— A bonding method combining two distinct technologies, flexible multifunctional, multi-electrode fibers [1] and robotic appendages that make use of electrochemical actuators [2], is demonstrated, showing successful operation in phosphate buffer solution (PBS). The resulting robotic fibers were ∼300 µm in diameter allowing us to shrink state of the art surgical tools by nearly an order of magnitude. This technique was demonstrated for two robotic prototypes, one in a gripper formation, and one in a sensor formation. These two distinct designs, when connected to a fiber, open the door to many new techniques for neural tissue biopsy and chronic neural recording with minimal damage to surrounding tissue.**

I. Introduction

Thermally drawn, polymer based fibers have become an important technology for applications requiring sensing and stimulation via optical, chemical, and electrical signals. For example, they have been shown to be very effective as biocompatible multi-functional neural devices capable of optical stimulation, drug delivery, and electric recording. In contrast to many rigid neuroprobes [3]–[5], fibers are flexible and enable chronic recording over long time periods. Despite their utility, however, once inserted, fibers remain rather limited in their ability to mechanically manipulate their environments beyond rudimentary operations such as spatial expansion [6], [7]. Here, we show that recent developments in microscopic robotics can be harnessed to vastly increase the mechanical manipulation capabilities of such fibers. Microscopic robots are lithographically fabricated in 2D and folded via electrochemical µ-actuators [8] to adopt their final 3D shapes and enable actuations. This technology has been used to demonstrate a variety of self-folding origami shapes ranging from Miura Ori to birds [2], autonomous walking robots [9], and artificial cilia [10], all at the 100 µm scale. Our idea is simple. We aim to use conducting wires embedded in fibers to drive and control simple actuations of a robot bonded to a fiber's tip. Our ultimate goal is to develop surgical tools that are an order of magnitude smaller than those currently available.

II. Device Fabrication

The robotic fiber is composed of two elements, a fiber that allows for in vivo implantation and can transmit electrical control signals and a fabricated microscopic robotic attachment with movable appendages.

A. The Fiber

The fiber portion of the device is fabricated using a three-step process: 1) fabrication of a fiber preform; 2) a macroscopic thermal drawing process [1]; 3) followed by a precise thermal tapering process [7]. The fiber preform is fabricated with channels that are filled with BiSn. This preform is drawn into thin fibers using the thermal drawing and thermal tapering processes. The final fibers have a large backend, approximately 2mm in diameter, and a micron scale tip, approximately 300 µm in diameter (Fig. 1 and 2). Three designs were used to demonstrate the bonding capabilities: five electrode BiSn with a solid core, five electrode BiSn with a hollow core, and high electrode count of BiSn with a hollow core, (Fig. 3). The hollow channel can be used for integration of additional sensors, delivery of chemicals, or for storing medical epoxy to aid in strengthening the bonding process with the robotic appendage. The electrochemical impedance spectrum (EIS) is measured for each design to ensure the impedance is optimal for control signals to be sent to the robot (Fig. 9).

B. The Robotic Appendage

The robotic appendage is composed of a rigid panel that contains bonding pads for connecting to the fiber as well as signal wires that connect to the actuating appendages. The layout for the two devices is shown in Fig. 4 and 5. The actuators are fabricated using a process similar to that described in [2]. Briefly, they consist of a nanometer thin metal layer such as Pt that under an applied voltage can absorb ions from solution and expand. This metal layer is bonded to an insulating layer such as Ti to form a bimorph. When the metal layer expands, the bimorph bends. Rigid panels composed of approximately 3 µm Si3N4 restrict the actuation so that bending occurs along fold lines. A 100 nm insulation layer also composed of Si3N4 prevents electrochemical reactions from occurring on the signal wire during actuation. Finally, we fabricate a metallic bonding layer on top of the rigid panels that connects to the wires through vias as shown in Fig. 4. The chip is then diced into individual devices using a DISCO dicing saw. The entire robot is fabricated on a fused silica wafer coated with approximately 180 nm AIN and 40 nm Al2O3. These two coatings act as a releasing layer, allowing the fabricated devices to be removed from the microfabricated robot before bonding to the fiber.

C. Connection Design

The individual connection pads were designed to allow the maximum area for bonding and increase alignment tolerance. A distance of 5 μm is kept between the individual bonding pads to prevent signal contamination from other channels and shorting. While the ring connection pad allows easier alignment during bonding at the cost of individual actuator movement. In the current formulation, the fiber is composed of 5 wires, equispaced in a circle of diameter 300 μm, with BiSn wires ∼40 μm in diameter.

III. Bonding Method

The microfabricated robotic appendage is cleaned using standard methods and then submersed in 2.38% tetramethyl ammonium hydroxide (TMAH) to etch away the release layer and remove the device from the substrate. After a final cleaning, the robotic appendage is then moved to a polyimide substrate sitting on an insulated flexible heater strip. The fiber is then aligned via a customized multidirectional manipulator. The manipulator controls the X, Y, Z, and theta alignment for the fiber and, when combined with a microscope, allows for a precise alignment. Once the alignment is complete, the fiber is lowered onto the bonding pads allowing for good contact with the heater strip (Fig. 6). The voltage source is set at 7.5 V for 45 seconds when using the solid core fiber and 35 seconds when using the hollow core fiber. In this process, the heater reaches a temperature of ∼423 K allowing the BiSn to melt onto the bonding pad and for the polycarbonate to spread onto the robotic base. The now connected devices are left to cool down for 2 minutes before lifting off the substrate. At this point, the fiber and robotic appendage are bonded and can be tested in solution (Fig. 7 and 8).

IV. Actuation in Solution

We tested two robot prototypes: a neural probe and a biopsy tool. Actuation and electrochemical impedance spectrum tests were conducted in phosphate buffer solution (PBS), shown in Fig. 10 and Fig. 11. Actuation tests were conducted using Gamry Virtual Front Panel Software with the single point function, to apply a fixed potential, or using Gamry Framework with the cyclic voltammetry function, for sweeping through a voltage range. The EIS test was conducted using the Gamry Framework software and analyzed at 1 kHz. For the neural probe actuation, we used the solid core fiber coated with a thin layer of medical epoxy. The device was submerged in a petri dish of PBS with an Ag/AgCl reference electrode and a platinum counter electrode. A cyclic voltammogram was chosen to cycle through -1.2V and 0.6V at a rate of 50 mV/s. This voltage sweep allowed for the neural probe actuator arms to sweep through their range of motion, as seen in Fig. 10. For the gripper probe actuation, the hollow core BiSn design was used since it did not need epoxy for a stronger connection due to the thin polycarbonate side walls heating faster and conforming to the microfabricated robot more reliably than the solid core fiber. The gripper actuation demonstration (Fig. 11) utilizes a high electrode count fiber with a hollow core; this configuration reduced the error associated with misalignment with the robot design containing 5 bonding pads as opposed to a bonding ring. After connection, the device is moved to a petri dish filled with PBS where it is lowered into solution. Using an Ag/AgCl reference electrode and platinum counter electrode, a single point voltage was applied and cycled manually through -1.2V and 0.6V; this method of voltage cycling allows for a longer period of time at each voltage before stepping to the next. These exciting results demonstrate our ability to bond microfabricated robotic appendages to fibers and, through electrical wires embedded in the fibers, control their actuation.

V. Conclusion

By successfully marrying state of the art fiber and microscopic robot technologies we have developed a new powerful platform for microscopic manipulation in tissues. The proof of concept actuation of multiple robotic designs demonstrated here could for example enable applications ranging from improved omnidirectional neural recording to minimally invasive neural tissue biopsy for potentially cancerous tumors for diagnosis. The different fiber designs show that successful bonding can be achieved without jeopardizing the multiple fiber functionalities, which will allow for new device development. For example, the center hollow channel design, when combined with mechanical functions such as robotic grippers, allows for device designs that, through fluid exchange, can chemically sense or stimulate as desired and then extract a desired tissue sample. As such, these demonstrations represent a critical step for miniaturization of surgical tools to the 100 μm scale.

Acknowledgment

We gratefully acknowledge funding support from the National Institute of Health (R21EY033080).

References

[1] A. Canales *et al.*, "Multifunctional fibers for simultaneous optical, electrical and chemical interrogation of neural circuits in vivo," *Nature biotechnology*, vol. 33, no. 3, pp. 277–284, 2015.

[2] Q. Liu *et al.*, "Micrometer-sized electrically programmable shape-memory actuators for low-power microrobotics," *Science Robotics*, vol. 6, no. 52, p. eabe6663, 2021.

[3] J. J. Jun *et al.*, "Fully integrated silicon probes for high-density recording of neural activity," *Nature*, vol. 551, no. 7679, pp. 232–236, 2017.

[4] N. A. Steinmetz *et al.*, "Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings," *Science*, vol. 372, no. 6539, p. eabf4588, 2021.

[5] C. M. Lopez *et al.*, "A neural probe with up to 966 electrodes and up to 384 configurable channels in 0.13 μm SOI CMOS," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 3, pp. 510–522, 2017.

[6] S. Jiang *et al.*, "Spatially expandable fiber-based probes as a multifunctional deep brain interface," *Nature communications*, vol. 11, no. 1, p. 6115, 2020.

[7] J. Kim *et al.*, "T-dope probes reveal sensitivity of hippocampal oscillations to cannabinoids in behaving mice," *Nature Communications*, vol. 15, no. 1, p. 1686, 2024.

[8] M. Z. Miskin *et al.*, "Electronically integrated, mass-manufactured, microscopic robots," *Nature*, vol. 584, no. 7822, pp. 557–561, 2020.

[9] M. F. Reynolds *et al.*, "Microscopic robots with onboard digital control," *Science Robotics*, vol. 7, no. 70, p. eabq2296, 2022.

[10] W. Wang *et al.*, "Cilia metasurfaces for electronically programmable microfluidic manipulation," *Nature*, vol. 605, no. 7911, pp. 681–686, 2022.

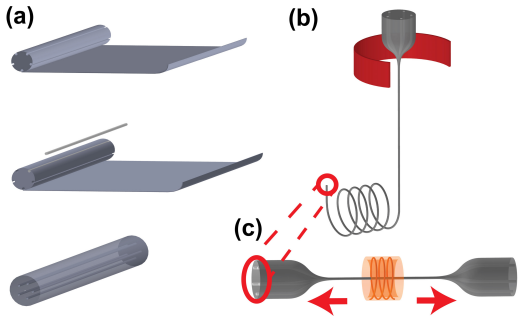


Fig. 1. (a) Fiber fabrication starts with rolling a macroscale preform with 5 channels milled into the polycarbonate and filled with BiSN (∼28 mm in diameter). (b) After thermal drawing, the preform is drawn to a minipreform (∼2 mm in diameter) and cut into 10 mm long segments. (c) Minipreforms then undergo thermal tapering, resulting in a device with a large backend for connections and a small interface for implantation.

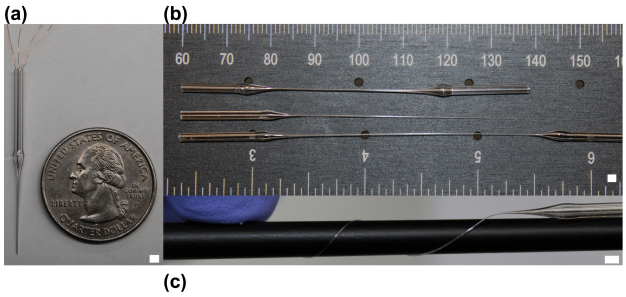


Fig. 2. (a) Copper backend connections in the tapered fiber device next to a quarter. (b) Top device is an uncut 5 electrode hollow core fiber, the middle device is a cut 5 electrode solid core fiber, and the bottom device is an uncut 5 electrode solid core fiber. (c) Tapered fiber wrapped around rod to demonstrate flexibility. Scale bars are 2 mm.

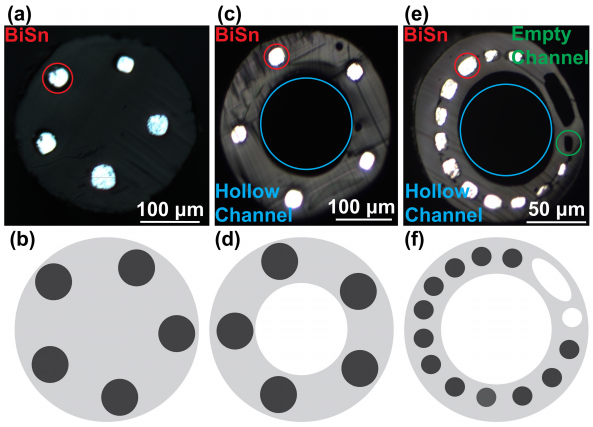


Fig. 3. (a) Optical image of 5 BiSn electrode solid core fiber cross section. (b) Ideal schematic of 5 electrode, solid core fiber cross section. (c) Optical image of 5 BiSn hollow core fiber cross section. (d) Ideal schematic of 5 electrode, hollow core fiber cross section. (e) Optical image of high electrode count fiber containing a hollow channel, 12 electrodes, and 2 empty channels for housing epoxy. (f) Ideal schematic of high electrode count fiber cross section.

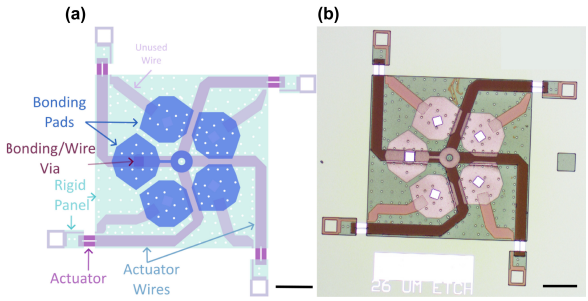


Fig. 4. (a) Schematic of neural probe chip base showing bonding pads with wire vias, actuator wires, actuators, rigid panels connecting to movable actuators creating arm like structures. (b) Optical microscope image of microrobotic neural probe. Scale bars are 100 µm.

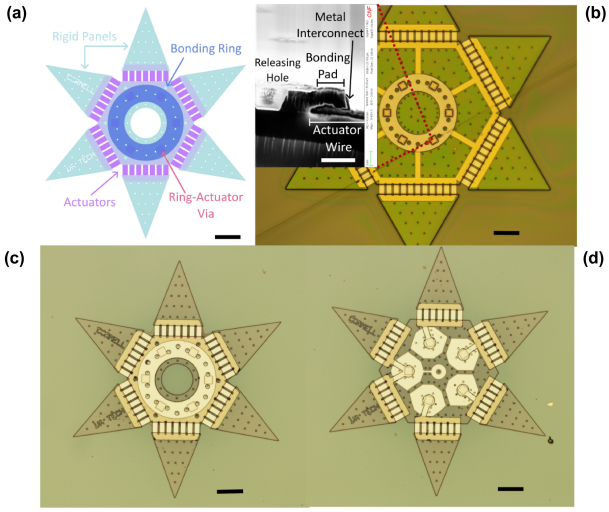


Fig. 5. (a) Schematic of robotic gripper chip base design showing ring shaped bonding pad with actuator vias, actuators, and rigid panels to form movable arms. Scale bar is 100 µm. (b) Optical image of microrobotic gripper design with wider actuator arms and SEM image showing various layers. Scale bar is 5 µm. (c) Optical image of microrobotic gripper with ring bonding pad. Scale bar is 100 µm. (d) Optical image of microrobotic gripper with 5 bonding pads. Scale bar is 100 µm.

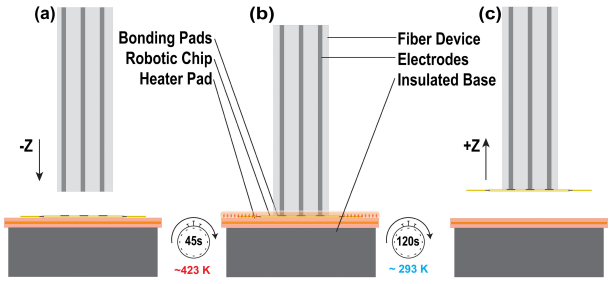


Fig. 6. Schematic of fiber to microscopic robot bonding process. (a) Fiber is lowered onto the microscopic robot base sitting on a thermal heater pad. (b) Once alignment is completed and the fiber is lowered onto the robot, the thermal pad is heated to ∼423 K for 45 seconds by utilizing a voltage source set to 7.5V. After 45 seconds, the device is left to cool for 120 seconds until it returns to room temperature. (c) Once bonded, the robotic fiber is lifted off of the thermal heating pad.

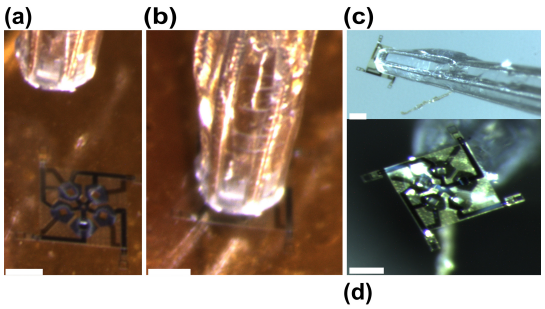


Fig. 7. Pre (a) and Post (b) bonding images for microrobotic neural probe connection to fiber device. (c) Bonded device from chip side and (d) Bonded device from fiber side. Scale bars are 200 µm.

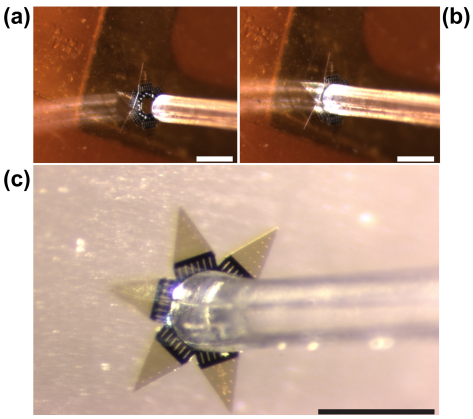


Fig. 8. Pre (a) and Post (b) bonding images for microrobotic gripper show successful bonding to fiber device. (c) Bonded device from fiber side. Scale bars are 500 µm.

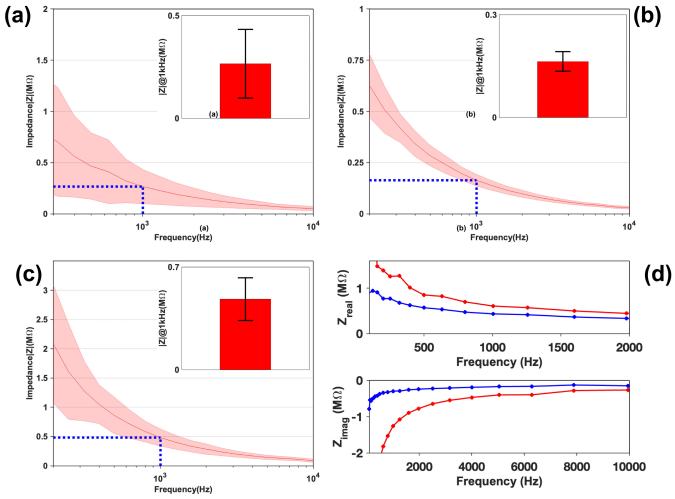


Fig. 9. (a) EIS plot for solid core fibers with shaded area and error bar representing standard deviation, n=4. (b) EIS plot for high elecrode count fibers with shaded area and error bar representing standard deviation, n=4. (c) EIS plot for hollow core fibers with shaded area and error bar representing standard deviation, n=4. For each plot, the red line indicates the mean impedance and the blue signifies the 1.004 kHz impedance. (d) EIS Pre and Post bonding high electrode count fiber to microrobotic gripper base. The red plot shows the before bonding EIS and the blue plot shows the post bonding for the magnitude (upper) and phase (bottom). The decreased impedance indicates good electrical connections post bonding.

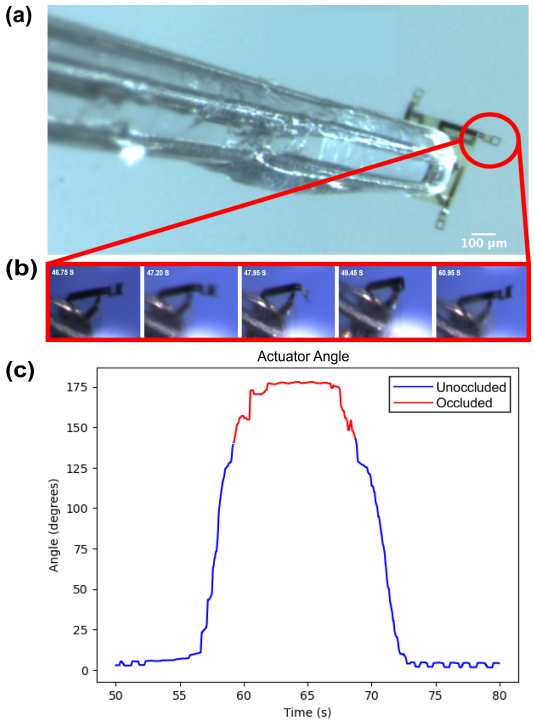


Fig. 10. (a) Bonded Neural Probe device. (b) Series of images showing actuator movement. (c) Actuation angle change versus time. Blue region represents the segment visible during actuation while the red region represents the occluded segment of actuation.

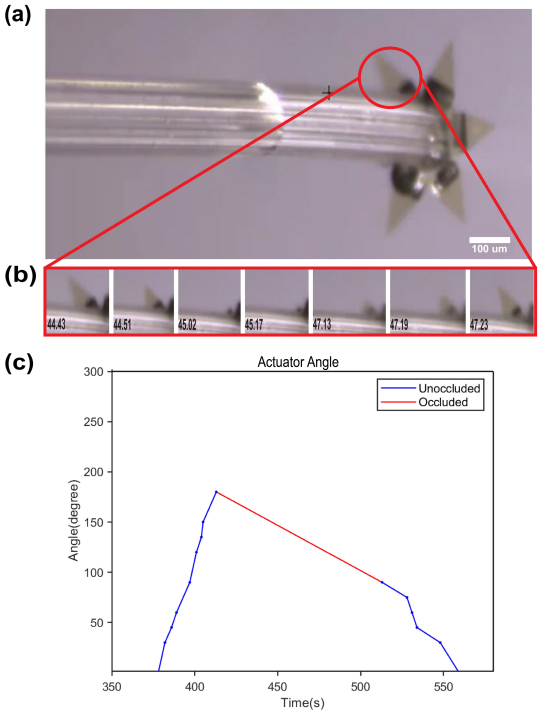


Fig. 11. (a) Bonded Gripper Probe device (b) Series of images showing actuator movement. (c) Plot of actuation angle change versus time. Blue area shows the angles that were visible during actuation while the red area represents the occluded segment of the actuation.

Implementation of double-side calibration and probing for Q-band 50GHz system application

Chia-Chu Lai[1], Sam Lin, Teny Shih, Andrew Kang, and Yu-Po Wang
Siliconware Precision Industries Co., Ltd., email: [1]chiachulai@spil.com.tw

***Abstract*—**The double-side calibration and measurement on probing system for single-end GSG (Ground-Signal-Ground) and differential pairs GSSG (Ground-Signal-Signal-Ground) type is presented in this paper. Calibration is the most important procedure before measurement due to it is used to remove the effect of cable and RF probe, and then extract the performance of DUT purely. The common calibration kit in the market is including below patterns: Open, Short, Load, and Thru for SOLT/SOLR standard calibration method, but there is still no standard thru/reciprocal pattern for double-side measurement system at present, therefore we design a reciprocal kit which is applicable for GSG-150um and GSGSG-150um RF probe and the frequency is up to 60GHz.

The reciprocal pattern is designed in the 4-layer stack up and simulated in ANSYS HFSS environment, the return loss is under -10dB for all 60GHz frequency range, and the insertion loss is -0.6dB at 60GHz. The actual measurement results after calibration with SOLR showed -10dB of the return loss under 60GHz, and -1dB of insertion loss at 60GHz. Finally, we successfully calibrated the double-side differential system by the reciprocal kit in Q-band 50GHz and also measured the two transmission lines with total 36mm length. First transmission line is all built in one layer, and the second one is divided in two layers. The measurement result showed the trace in one layer is better about 0.56dB@30GHz.

I. INTRODUCTION

Recently, there are a lot of discussions about advanced packaging technology below 7nm process node. Especially driven by the application for artificial intelligence (AI) application, the demand for the GPU die is continuing to rise up those results in demand exceeds provision. Especially the advanced packaging CoWoS (Chip on Wafer on Substrate) is mainly suitable for the GPU. Fig.1 is the schematic diagram of the CoWoS package. Connect the logic chip and the High Band Memory (HBM) chip on the interposer by the u-bumps, and integrate the signal between logic chip and HBM chip by the RDL, and connect down to the substrate through TSVs and u-bumps, and finally go to the position of the bottom ball through the metal layers, vias, and core vias. The interfaces of the vertical interconnect are including the u-bumps, TSVs, vias, and the PTHs in a package substrate. It is interesting to understand the effect of these interface, but the double-side measurement is the challenging over a wide range of frequencies.

The common configuration of frequency measurement system is composed of a probe station and a VNA (Vector Network Analyzer) instrument which is shown in Fig.2. The probe station is to load the wafer or DUT board and the VNA is specialized for high frequency power sweep. Before starting the DUT test, we must do the calibration step to eliminate the noise and effect of the cable and probe, and then extract the pure characterization of DUT. And the calibration kit in the market is ready for single-side probe but not for double-side probe. The main key point is there is no standard Thru kit for double-side calibration. Therefore, some literatures descript different structures of the Thru kit, for example, the first is Thru kit with non-contact technology [1], the second is developing a whole double-side probe system and design a via in 2-layer PCB as a Thru kit [2], the third is design a paralleled and a diagonal Thru kit in only a layer PCB structure to implement differential double-side calibration [3].

In this paper, we organized as follows. Section II presents the different calibration methods and the design conception of the calibration Thru kit. Section III gives the measurement results for the calibration Thru Kit and the differential transmission lines which are built in 22-layer substrate. Finally, a conclusion is given in Section IV.

II. Double-Side Calibration

A. Calibration Method

There are many calibration choices. The common probing calibration is SOLT and SOLR. For SOLT (Short-Open-Load-Thru), it is very simple and not band- limited, but requires very well-defined standards. For SOLR (Short-Open-Load-Reciprocal), the advantages are the same as SOLT but it doesn't need very good for the thru line with symmetrical type. For SSST (Short-Short-Short-Thru), the advantages are the same as SOLT but it has better accuracy at higher frequency. For LRL/TRL (Line-Reflect-Line), the advantages are the highest accuracy and minimal standard definition, but it requires very good transmission lines. The total comparison with different calibration is arranged in Table I. In this paper, we chose the SOLR to calibrate the measurement system due to the thru line is no need very well performance. The short, open, and load kits are using the standard calibration kit and the design conception is descripted in next portion.

B. Design Conception of the Thru Kit

The Thru kit is designed in 4-layer substrate with dielectric constant 3.4, and each thickness of the dielectric layer is 60um

and metal thickness is 20um. The width and spacing of the Thru kit are optimized to comply with 50ohm impedance. The signal traces are layout in layer 2 and 3 and these two traces are connected by multiple vias in order to get symmetrical structure as like Fig. 3. In the top side and bottom side, we design a 150um pitch pad for GSG RF probing.

The Thru kit is simulated by the software ANSYS HFSS. The frequency range is up to 70GHz including Q-band 33GHz~50GHz. And the simulation results are shown in Fig4. The return loss is under -12.5dB for all frequency range and the insertion loss is about -0.6dB at 60GHz. It meets our target the return loss is under -10dB and the insertion loss is under -1dB for all frequency range.

III. Measurement Result

A. GSG Transmission Line & Thru Kit Measurement

Before measuring the DUT, we did the double-side calibration. The parameters of the standard calibration kit need be set in the instrument Anritsu Vector Star MS4647B in advance. Firstly, we probed on the short, open, and load of the standard calibration kit on the top side with the 150um GSG RF probe, and on the back side also repeat the same step. The calibration kit could be attached on the back side of the probe station by the vacuum. Secondly, use the clamps and insulation screws to fix the thru kit, and then the RF probes could touch on the top and back side of the thru kit simultaneously. Finally, calculate the error terms by the VNA instrument and completed the calibration steps. The reciprocal thru kit and transmission line with 2mm length are measured after the calibration and the results are displayed in Fig.5 which's curves are all reasonable. The insertion loss is about -1dB and return loss is under -12.5dB for DC-60GHz frequency range. We could find it has -0.4dB difference of insertion loss between the simulation in section II and measurement, it maybe that we didn't consider the roughness effect of the copper in the simulation. The Fig.5(b) shows the insertion loss with the 2mm double-side line, the light blue curve is the measurement and the black curve is the simulation, the correlation is over 90% and the difference is about 0.5dB between simulation and measurement.

B. GSSG Differential Pair Measurement

Fig. 6 displayed the structure for GSSG differential pairs layout. Two differential pairs are built in multi-layer substrate, and both pairs meet the 85Ω impedance by calculating the space, gap, and width based on the dielectric thickness 35um and dielectric constant 3.4. The difference of the two differential pairs is one pair is all signal trace placed in one layer, and the other pair is divided in two layers. The probing pad on the top side and bottom side are designed in the form of GSGSG arrangement for 150um differential RF probes due to GSGSG RF probe has better isolation and calibrated performance. Before measuring the DUT, we also did the calibration step by using the same design thru kit. Initially we define port1 and port3 is one set, port2 and port4 is another set. Different from the GSG calibration is the thru measured: we need to measure the design thru kit twice with two paths of the port1 to port2 and port3 to port4, and measure the thru lines of the standard calibration kit for port1 to port3 and port2 to port4. The open, short, load measured are the same as the steps of the GSG calibration for each port. Complete the above steps, we finished the all steps of the calibration and let instrument VNA calculate and eliminate the error terms. The calibration results are shown in Fig. 7. The load response with each port is good that return loss is all under -20dB with DC-50GHz frequency range. The open smith chart of each port in air is no abnormality with DC-50GHz frequency range.

The two differential pairs were measured after calibration and shown in Fig.8. The pair routed in one layer with 36mm length is blue curve and the pair routed in two layers which the ratio is 2:1 with total 36mm length is the green curve. We could see the return loss of green curve is slightly worse compared to blue curve, but both are well design since the return losses are less than -10dB. Another comparison is the insertion loss, the pair routed in one layer is better 0.56dB at 30GHz. This is because the conversion interface is added once for the pair routed in two layers.

C. Measurement Environment

Fig. 9 shows the testing environment of double-side probing and the probe positioners are moved manually to the right position and observe the testing pad touching on the top side and bottom side under the respective microscope. Measure the signal performance in the substrate with the Anritsu MS4647B vector star.

IV. Conclusion

This paper successfully demonstrated the double-side measurement. Both GSG single-end signal and GSSG differential pairs have been tested the reasonable curves. The frequency range of the calibration in Double-side system is DC-60GHz for GSG Probe and DC-50GHz for GSGSG Probe. The correlation is over 90% and the difference is about 0.5dB between simulation and measurement with GSG single-end signal. For the measurement results of the GSSG differential pairs, it is recommended to route the differential pairs in one layer. Finally, the hole double-side test system could be used for Q-band applications.

Acknowledgment

The authors would like to acknowledge the assistance and support from RD, Substrate, Material Section, and Leaders in SPIL.

References

[1] H. -J. Hsu, M. -H. Tu, S. -M. Wu and C. -C. Chen, *"Study of double-side Thru calibration kit by non-contact coupling structure,"* 2015 Asia-Pacific Microwave Conference (APMC), Nanjing, China, 2015, pp. 1-3.

[2] K. -C. Lu et al., *"Vertical interconnect measurement techniques based on double-sided probing system and short-open-load-reciprocal calibration,"* 2011 IEEE 61st Electronic Components and Technology Conference (ECTC), Lake Buena Vista, FL, USA, 2011, pp. 2130-2133.

[3] H. -H. Shih, M. -H. Tu and S. -M. Wu, "Non-exchanging structure calibration kit for double-side direct contact measurement," 2015 Asia-Pacific Microwave Conference (APMC), Nanjing, China, 2015, pp. 1-3.

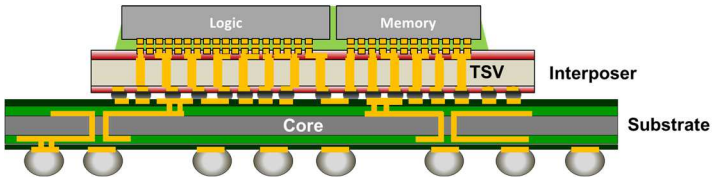


Fig 1. 2.5D Advanced package

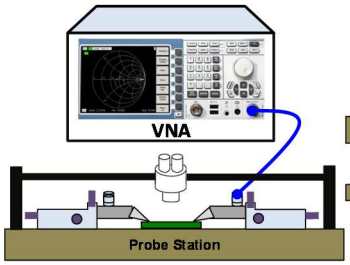


Fig. 2(a) Single-Side System

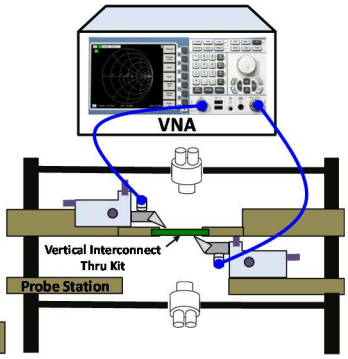


Fig. 2(b) Double-Side System

Table I. The Comparison of Different Calibration

Calibration Method	Advantages	Disadvantages
SOLT (Short-Open-Load-Thru)	- Simple - Not band-limited	- Requires very well-defined standards
SSST(Short-Short-Short-Thru) (shorts w/ different offset lengths)	- Same as SOLT - better accuracy at high frequency	- Requires very well-defined standards - Band-limited
SOLR(Short-Short-Short-Reciprocal)	- Same as SOLT - Does not require well-defined thru	- Some accuracy degradation due to less defined thru.
LRL/TRL (Line-Reflect-Line)	- Highest accuracy - Minimal standard definition	- Requires very good transmission lines - Band-limited.

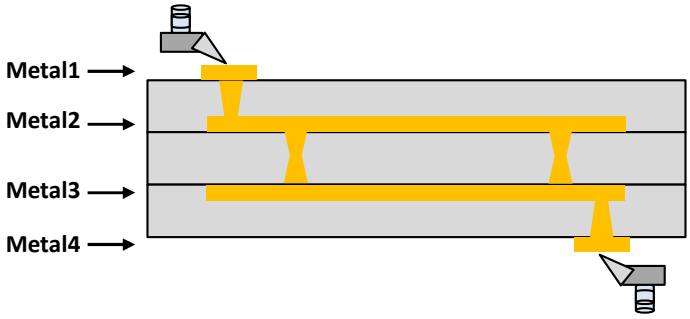


Fig. 3 The Cross-section Diagram of the Thru kit

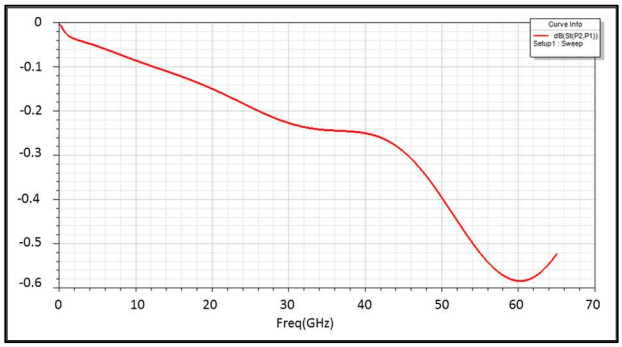


Fig. 4 (a) Insertion loss

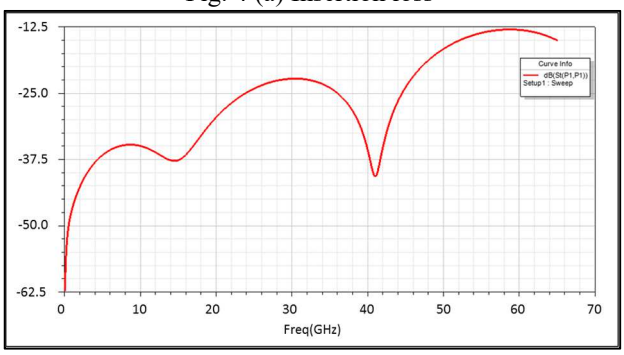


Fig. 4(b) Return Loss of Thru kit

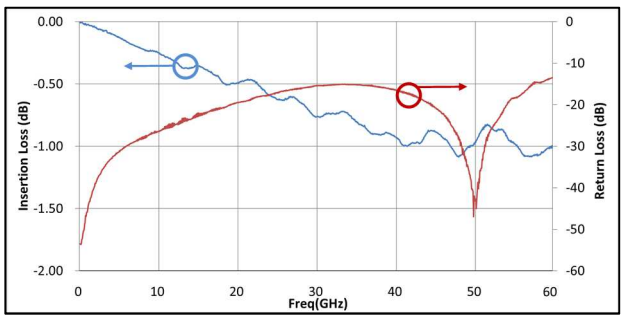


Fig. 5 (a) The insertion loss and return loss with Thru kit

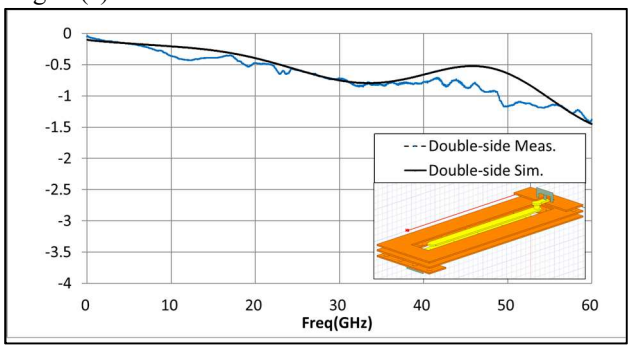


Fig. 5(b)The insertion loss with 2mm length line

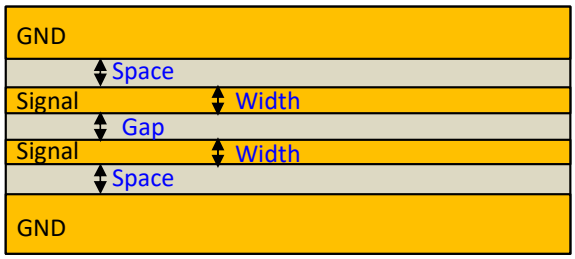


Fig. 6(a) The top view of GSSG layout in the substrate

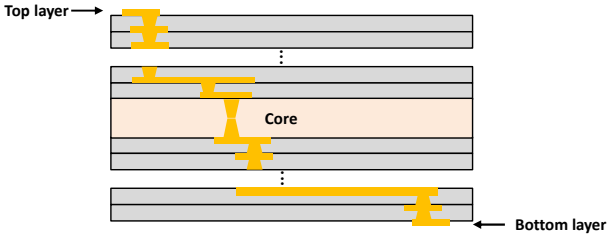


Fig. 6(b) The cross-section connected in the substrate

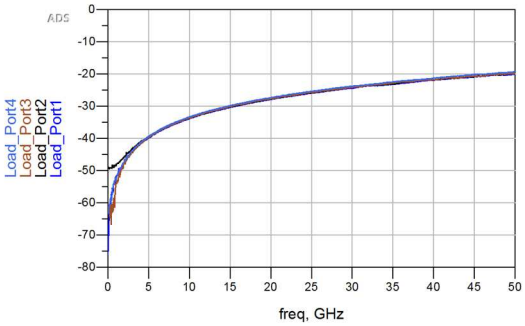


Fig. 7(a) The load response for each port

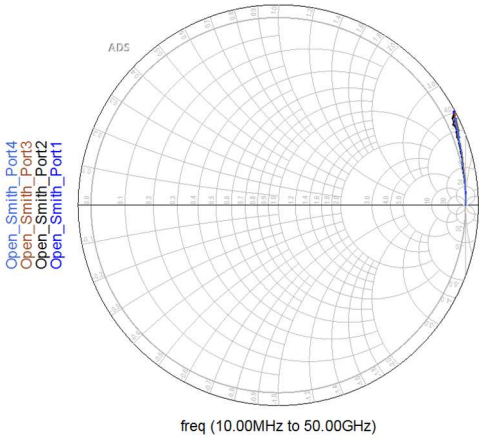


Fig. 7(b) The open smith chart for each port

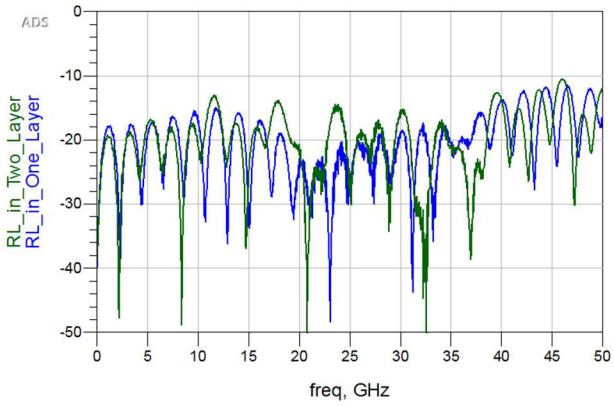


Fig. 8(a) The comparison of return loss

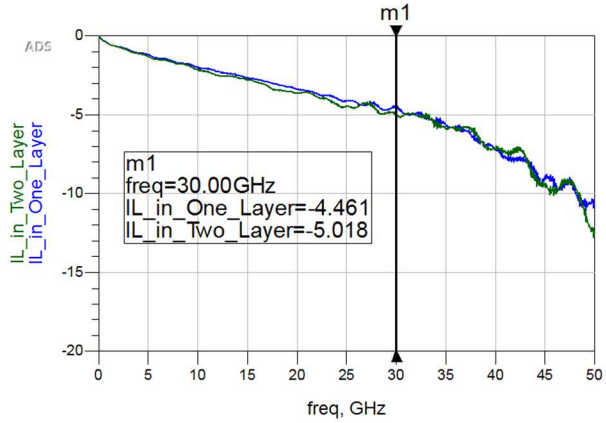


Fig. 8(b) The comparison of insertion loss

Fig. 9 The Double-side Probe station and Probing status

Antiferroelectric Multilayers: a Multifunctional Platform for Energy and Information Storage

Hung-Wei Li, Sadegh Kamaei, Cyrille Masserey, Pascal Morel, Niccolò Martinolli, Tom Schlatter, Océane Mauroux, Igor Stolichnov, and Adrian M. Ionescu.
EPFL, Nanoelectronic Devices Laboratory (NanoLab), 1015, Lausanne, Switzerland

***Abstract*— This work demonstrates the integration of energy storage and multi-bit memory functionalities in a single platform device using multilayer dielectric (DE) with ferroelectric (FE) or antiferroelectric (AFE) materials. Employing a hafnium-based oxide, we achieve an energy storage density of approximately 50 J/cm^3 and advanced multi-bit storage capabilities. We investigate the effects of adjusting silicon doping concentration in multilayer Si:HfO_2 during the atomic layer deposition (ALD) process to form and compare both FE and AFE layers. Inserting an Al_2O_3 dielectric (DE) layer between the FE or AFE layers, we create a built-in electric field, enhancing both energy storage efficiency and multi-bit memory storage beyond the current state of the art. Our innovative approach harnesses the intrinsic properties of these materials to boost energy efficiency and data storage capacity, significantly improving device functionality and advancing electronic component design. Our results reveal that strategically layering DE with AFE materials enhances energy and multi-bit storage, opening new avenues for future multifunctional electronic applications.**

I. Introduction

Hafnium oxide-based ferroelectric (FE) and antiferroelectric (AFE) thin films are emerging as key materials in the evolution of advanced memory and energy storage technologies. Integrating these materials into commercially viable devices demands not only comprehensive material characterization but also innovative engineering to enhance both endurance and performance. This study introduces a groundbreaking device that seamlessly integrates high-density energy storage with robust multi-bit memory functionality within a techmological framework. Utilizing multilayer configurations of dielectric (DE), FE, and AFE materials, we have achieved an unprecedented energy density of approximately 50 J/cm^3, coupled with enhanced multi-bit storage capabilities. This advancement is made possible through meticulous control of silicon doping in Si:HfO_2 during atomic layer deposition (ALD), allowing for the precise formation of distinct FE and AFE phases. A key innovation in our device is the incorporation of an Al_2O_3 dielectric layer between the FE and AFE layers, which generates a built-in electric field. This field significantly boosts energy storage efficiency while simultaneously stabilizing data retention in multi-bit memory configurations (Fig. 1). As the demand for ultracompact electronic devices increases, particularly in wearable and implantable technologies, our research addresses the *multi-functional need for devices that not only store energy efficiently but also retain data reliably*. The miniaturized energy autonomous systems enabled by this technology will integrate energy harvesting and storage with memory, offering a comprehensive solution for long-term, reliable device operation. Our findings not only demonstrate substantial improvements in energy and memory performance but also pave the way for the development of next-generation electronic components.

II. Device Fabrication Process

A. Tailoring silicon doped hafnium oxide properties

In the ALD process for Si:HfO_2 thin films, adjusting the SiO_2 and HfO_2 cycle ratios (Fig. 2a) yields either ferroelectric (FE) or antiferroelectric (AFE) behavior [1]. The metal-ferroelectric-metal (MFM) capacitor exhibits ferroelectric (Fig. 2(b)) or antiferroelectric (Fig. 2(c)) behavior depending on the SiO_2 doping concentration.

B. FeCap fabrication

The fabrication process for the multilayer capacitors illustrated in Fig. 3(a). We begin by preparing a silicon substrate with a 200-nm thermally grown SiO_2 layer on both sides. After the standard cleaning process, the bottom electrodes composed of titanium (Ti) and platinum (Pt) were deposited by sputtering. Without breaking the vacuum, we sputter a titanium nitride (TiN) layer. The multi-layer structure is then created through ALD. Table 1 provides details on the various sample configurations. For consistency in comparison, we maintained the overall thickness of FE/AFE layers at 40 nm across all samples. After the multi-layer structure is complete, we deposit a second TiN and the entire stack then undergoes a rapid thermal annealing (RTA) in a nitrogen atmosphere. This step is crucial in achieving the desired orthorhombic crystalline phase, which is essential for tuning the FE or AFE properties of our device. The final stage of fabrication involves defining individual capacitors, by sputtering Ti and Pt layers, followed by a ion beam etching (IBE) process. This results in capacitors with a well-defined area of in the range of 100 μm^2 for characterization and performance evaluation.

III. Results and Discussion

A. Multi-layer Capacitor Characterization

Fig 5 illustrates the experimental hysteresis curves of dielectric and ferroelectric behaviors in fabricated multilayer

capacitors, highlighting the influence of interfacial engineering and material composition on electrical properties. Fig. 5(a) and (b) show the polarization-electric field (P-E) curves for DE1FE10×4 and DE1AFE10×4, respectively. The P-E curve for DE1FE10×4 displays characteristic ferroelectric behavior with suppressed remanent polarization (P_r), attributed to the dielectric layer's effect on ferroelectric switching. While the MFM capacitor exhibits a $2P_r$ of 40 μC/cm^2, DE1FE10×4 shows a reduced $2P_r$ of 20 μC/cm^2. The P-E curve of DE1AFE10×4 demonstrates antiferroelectric characteristics and the versatility of layering strategies in tailoring material properties. The observed crossover of capacitance-voltage (C-V) branches (Figs 5(c) and (d)) at a non-zero electric field indicates the presence of a built-in electric field, likely resulting from asymmetric charge distributions or interfacial dipole moments within the multilayer structure. Fig. 6(a) and (b), representing DE1FE5×8 and DE1AFE5×8 respectively, demonstrate how a thicker dielectric layer alters the P-E response. This is evident from the linear capacitor behavior observed at lower electric fields, which also contributes to an increase in the electric breakdown voltage at high electric fields.

B. Energy storage enhancement with multilayered AFE/FE

Fig. 7(b) and (c) compares energy storage capabilities of ferroelectric (FE) and antiferroelectric (AFE) multilayers. Our work demonstrated that AFE materials exhibit superior performance due to their unique phase transition properties. Energy storage is calculated from the area under the P-E curve, revealing total stored energy and hysteresis losses (Fig 7(a)). This method highlights fundamental differences in dipole behavior between FE and AFE materials under electric fields. FE multilayers like DE10FE12×2 achieve energy densities up to 30 – 40 J/cm^3 but are limited by breakdown electric field. In contrast, AFE multilayers, particularly DE1AFE10×4, exceed 60 J/cm^3 with efficiencies up to 80% at 3 MV/cm. This performance stems from AFE materials' ability to *switch between non-polar and polar states*, allowing for controlled energy discharge. To further increase the breakdown voltage, we compared more stacks of the multilayer. However, while increasing the number of layers achieves a higher breakdown voltage, it also results in lower energy density. Both phenomena can be observed in FE and AFE devices, which is due to the more pronounced DE contribution suppressing the dipole from FE and AFE layers. Unlike the rapid discharge in ferroelectrics, AFE materials demonstrate *controlled energy release*. This driven by reversible phase transitions, enhancing energy storage capacity and efficiency, and making AFE multilayers suitable for applications requiring efficient, high-density energy storage. Charge-discharge characteristics of FE and AFE multilayer capacitors, providing insights into their practical performance. The setup uses a DC power supply and controlled discharge through a resistor to simulate real-world conditions (Fig. 8(a)). The voltage-time curves in Fig. 8(b) and 8(c) compare different multilayer configurations during charge-discharge cycles, using a 2.2 nF capacitor as a reference which is close to the maximum capacitance (Fig.5 (d)) we can obtain from AFE device. FE structures display rapid voltage decay, typical of quick energy release due to polarization switching, with DE1FE10×4 showing a slightly slower but still significant drop from 15V to near 0V within about 100 μs. In contrast, AFE structures exhibit more controlled voltage decay, with DE1AFE10×4 maintaining higher voltage for longer. This confirms the superior performance of AFE materials in controlled energy release, making them ideal for applications requiring energy efficiency and longevity. Fig. 9 demonstrates the long-term stability of AFE materials under extensive field cycling, crucial for durable energy storage applications.

C. Multibit memory storage enabled by multilayered AFE/FE

FORC measurements (Fig. 9(a-f)) offer insights into the switching behavior of these materials. By inserting the DE layer, the current peaks can be finely tuned due to the built-in electric field (Fig. 9(b)). The DE1FE10×4 structure shows typical ferroelectric hysteresis with distinct current peaks at coercive fields. The presence of the DE layer subtly shifts these coercive fields closer to ±2 MV/cm. Integrating DE layers making DE1FE10×4 suitable for memory and energy storage technologies with fast response times and improved endurance. For DE1AFE10×4 (Fig. 9(e)), the FORC diagram reveals a shift in current peaks towards the same field polarity, a phenomenon not observed without the DE layer. This shift, caused by the DE layer's influence on the internal electric field, indicates multibit nonvolatile properties. By stabilizing specific states, this shift ensures reliable quaternary multibit memory operations. The read-out transient current data further supports these findings (Fig. 11). A ±12V pulse switches one peak, while a ±20V pulse switches both, with the current in the mA range. This demonstrates precise control, enabling accurate multibit memory storage and retrieval. The C-V characteristics of AFE materials (Fig. 12) display a double-peak butterfly shape with significant frequency dispersion, indicating rapid switching speeds in DE+AFE structures.

IV. Conclusion

This study comprehensively analyzes for the first time the energy storage capabilities, switching behavior, and memory performance of ferroelectric (FE) and antiferroelectric (AFE) multilayer capacitors. Through detailed charge-discharge measurements, C-V characteristics, and FORC analysis, it is demonstrated that *AFE multilayers exhibit superior energy storage capacity, controlled energy release, and enhanced nonvolatile properties compared to FE structures*. These advantages position AFE materials as a promising platform candidate for next-generation advanced electronic devices with multifunctional energy storage and memory capability.

References

[1] T. Boscke, et al, *Applied Physics Letters*, vol. 99, 2011. [2] F. Ali, et al, . *J Appl Phys* , vol. 122, 2017. [3] J. P. B. Silva, et al, *J Mater Chem A*, vol. 8, 2020. [4] Cheema, S. S., et al, *Nature*, 2024.

Multilayer Structure Fabrication Process

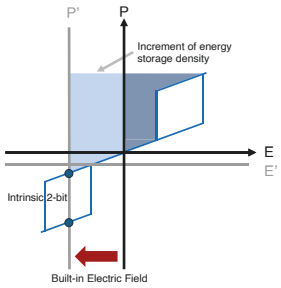


Fig 1. Schematic illustration of the qualitative P-E loop for the multilayer metal-insulator-ferro-electric-metal (MIFM) capacitor structure. The integration of a dielectric (DE) layer with antiferroelectric (AFE) or ferroelectric (FE) materials introduces a built-in electric field. This field significantly enhances the energy storage density and enables multi-bit memory functionality within the same MIFM capacitor structure, optimizing both memory capacity

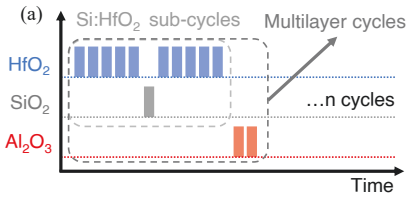
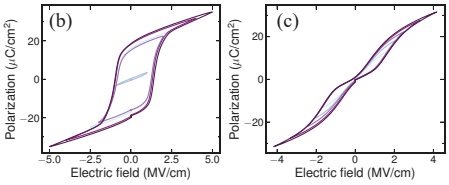


Fig 2. (a) ALD process schematic for Si:HfO_2 thin films. Si:HfO_2 sub-cycle ratio adjustments yield FE or AFE behavior.



P-E hysteresis loops of MFM capacitors: (b) 3.12% Si concentration (ferroelectric), (c) 5% Si concentration (antiferroelectric).

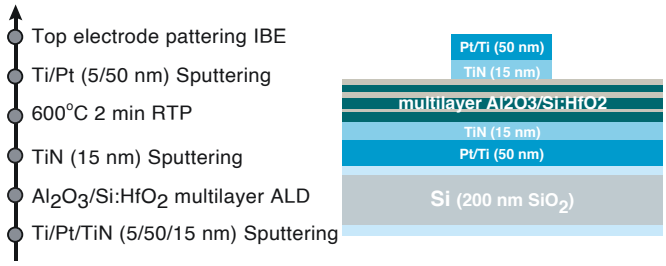


Fig 3. Key fabrication steps for the multilayer capacitor structure. The ALD multilayer cycles are adjusted to achieve different material combinations in the multilayer structure.

Sample ID	T_{AlO} (nm)	T_{SiHfO} (nm)	Layer numbers	Si:HfO_2 properties
DE10FE10x2	10	10	2	Ferroelectric
DE1FE10x4	1	10	4	Ferroelectric
DE1FE8x5	1	8	5	Ferroelectric
DE1FE5x8	1	5	8	Ferroelectric
DE1AFE10x4	1	10	4	Antiferroelectric
DE1AFE8x5	1	8	5	Antiferroelectric
DE1AFE5x8	1	5	8	Antiferroelectric

Fig 4. The tables summarize key samples, providing its thickness of each layer and the ferroelectric/antiferroelectric properties.

Experimental Device Characterisitics

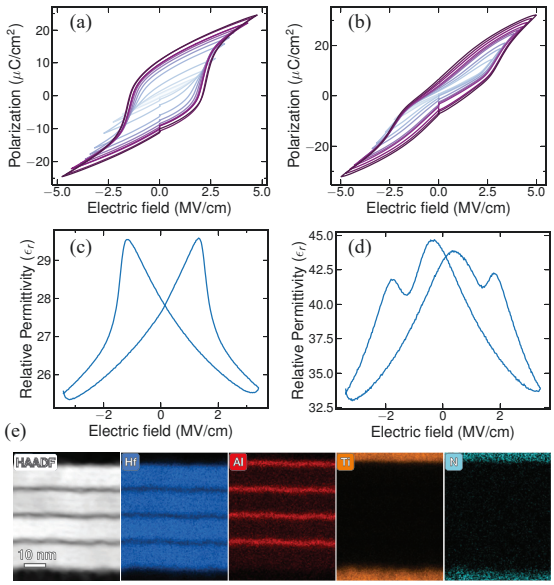


Fig 5. (a) P-E curve of DE1FE10x4 showing typical ferroelectric behavior, with suppressed remanent polarization due to the dielectric layer's contribution. (b) P-V curve of DE1AFE10x4 demonstrating antiferroelectric characteristics. (c) Butterfly-like dielectric response of the FE capacitor after interfacial engineering of interlayers, with a crossover of C-V branches at a non-zero electric field indicating a built-in field. (d) Butterfly-like dielectric response of the AFE capacitor under similar conditions, also showing a crossover due to a built-in field. (e) TEM image and EDS results for the multilayer structure, providing a detailed view of the material composition and interfaces.

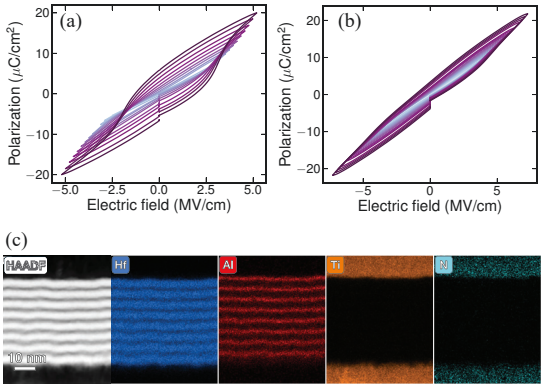
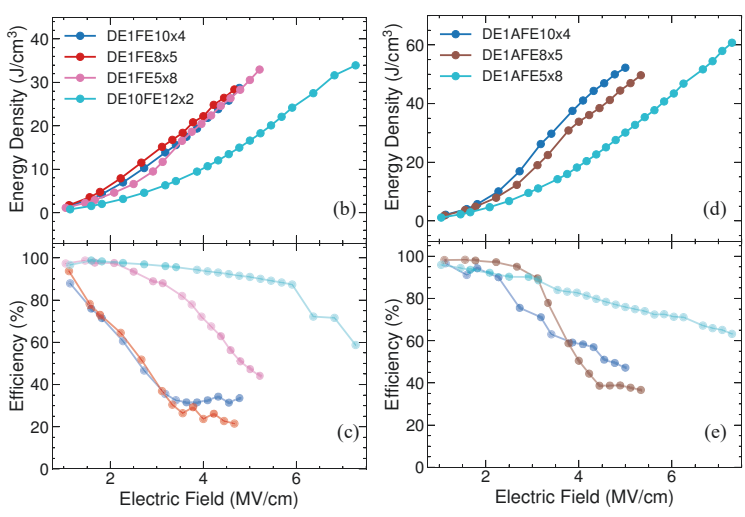


Fig 6. (a) P-E curve of DE1FE5x8, showing ferroelectric properties with a noticeable linear capacitor response at low electric fields due to the increased dielectric layer contribution. (b) P-V curve of DE1AFE5x8, also exhibiting a linear capacitor response at low electric fields, influenced by the dielectric layer. (c) TEM image and EDS results, providing detailed insights into the material structure and composition.

Experimental Device Characterisitics: Energy Storage Density



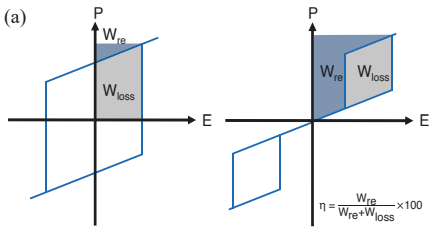


Fig 7. (a) Calculation of energy density and efficiency derived from the P-E loop, involving the integration of the area under the polarization-electric field curve to assess the energy stored and the corresponding efficiency based on energy loss. (b, c) Energy density versus electric field plots for ferroelectric multilayers, showing the material's capacity to store more energy when adjusting the inserting DE layers. (d, e) Similar plots for antiferroelectric multilayers, highlighting the distinct energy storage characteristics compared to their ferroelectric counter-parts, including differences in energy density and efficiency.

Charge - Discharge Measurement

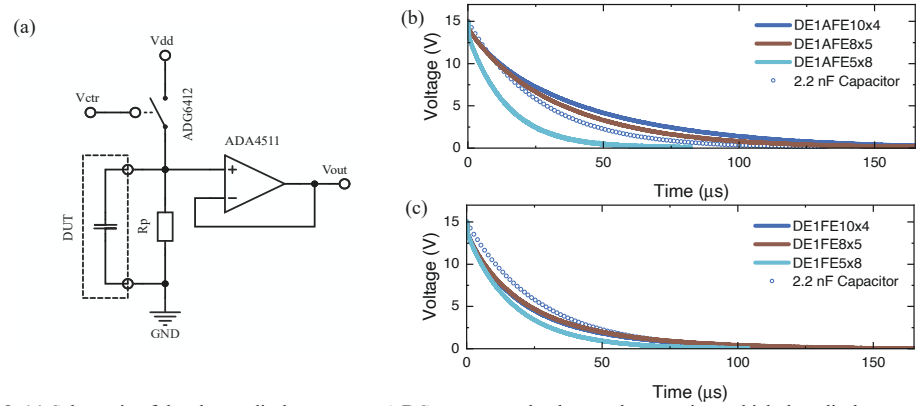


Fig 8. (a) Schematic of the charge-discharge setup: A DC power supply charges the capacitor, which then discharges through a 12 kΩ resistor. The pulse duration (10 µs) is controlled by a switch (ADG6412), and the voltage is monitored using an oscilloscope. A follower (ADA4511) reduces the parastics from the oscilloscope. (b, c) Comparison of discharge curves for ferroelectric and antiferroelectric multilayer structures with those of an ideal 2.2 nF capacitor. These simulated values are used as a reference to assess the performance of the multilayer structures.

Endurance

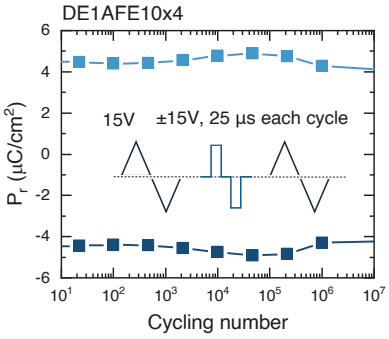


Fig 9. The device demonstrates fatigue-free behavior during field cycling, benefiting from the intrinsic properties of the antiferroelectric (AFE) material.

First-order Reversal Curve (FORC)

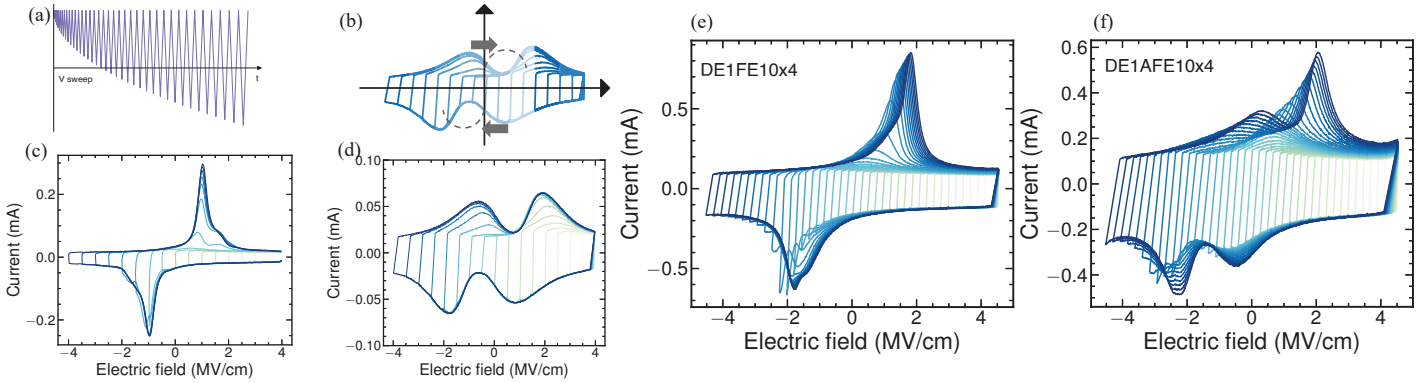


Fig 10 (a) Schematic outlining the First-order Reversal Curve (FORC) measurement approach. The applied field is swept from positive saturation to a reversal point, then back to positive saturation. This process is repeated for multiple reversal points to generate a series of curves that together form a comprehensive map of the material's electric behavior. (b) Schematic of AFE with DE multilayer, the current peaks will shift. FORC result of the MFM capacitors exhibiting (c) FE properties and (d) AFE properties. (e) FORC results showing the effect of the DE layer on the FE multilayer capacitor, illustrating changes in the coercive field and interaction fields due to the DE layer. (f) FORC result of the AFE multilayer capacitor with a DE interlayer, where the addition causes the two peaks at opposite field polarities to shift toward the same polarity, indicating multi-bit nonvolatile properties. This shift suggests a modification in the energy landscape of the AFE material due to the presence of DE layers.

Multi-bit Cell Write and Read

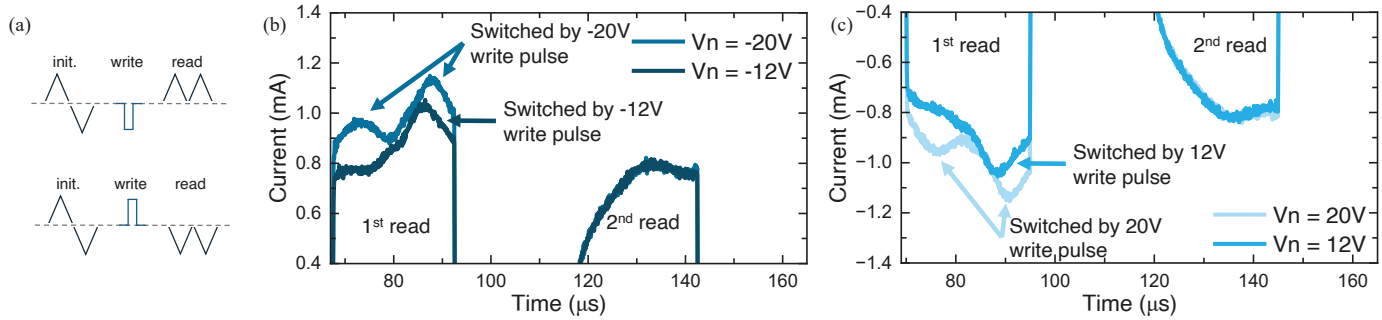


Fig. 11 (a) Schematic of the pulse sequence used to verify the write operation. After the write operation, two pulses (18V) with opposite polarity are applied to read out the switching polarization achieved during the write process. (b, c) Read-out transient current for different write pulse levels. The results show that a ± 12 V pulse can switch only one peak, while a ±20V pulse can switch both peaks, demonstrating the control over the switching behavior with varying pulse levels.

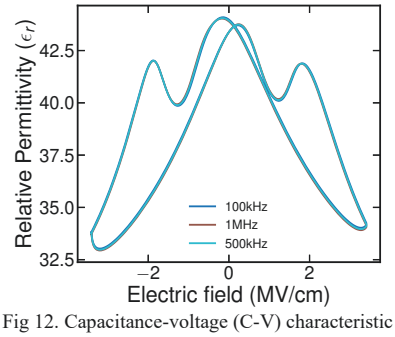


Fig 12. Capacitance-voltage (C-V) characteristic showing a double-peak antiferroelectric (AFE) butterfly shape. The robust frequency dispersion observed indicates a fast switching speed in the multilayer DE AFE structure.

	Material	ESD (J/cm³)	Efficiency	Memory and Energy Storage Integration
F. Ali, et al [2]	Si-HfO_2	61	65%	-
J. P. B. Silva, et al [3]	Al_2O_3: DIL-$Hf_{0.5}Zr_{0.5}O_2$	54	51%	-
Cheema, S. S. et al [4]	HfO_2/ZrO_2	115	90%	-
This work	Al_2O_3/Si:HfO_2	50	80%	Yes

Table 2. Benchmarking of various on-chip CMOS compatible energy storage capacitors.

Thermal Management for In-Vehicle Emergency Call Systems

Boyoun Park and Chungwoo Park
Samsung Electronics, Hwaseong, South Korea, email: boyoun.park@samsung.com

***Abstract*—**Given the life-saving significance of eCall, it is imperative that it operates reliably under extreme environmental conditions such as high external temperatures. To address this, this paper proposes an integrated thermal management solution from a product perspective, transcending the traditional SoC approach. This solution holds potential for widespread application across various embedded systems where external cooling methods are not available.

I. Introduction

eCall is an emergency call system for vehicles to assist in road traffic accidents, which is legislated by the European Union Parliament (EU-2015/758). eCall is triggered automatically when an accident is detected or manually when the eCall SOS button is pressed, providing the vehicle's precise location (Fig. 1). Since eCall is essential for saving lives, it needs to operate even in extreme conditions such as high ambient temperature ranging from -40 ℃ to 85 ℃. However, a TCU (Telematics Control Unit) that is an embedded device supporting eCall is mainly positioned in a place that is not suitable for cooling (e.g., between a vehicle's roof and roof liner). In addition, the increased power consumption of MPSoC (Multi-Processor System-On-Chip) due to high power density also results in a significantly higher temperature. Under the circumstances, the existing temperature solutions have been insufficient to solve those problems. Therefore, this paper suggests an integrated temperature control solution at the product level surpassing the SoC level and from the total power perspective, not being limited to the perspective of the dynamic power.

II. Power and Thermal Characteristics of SoC

In modern processors, to improve performance, both the number of transistors per die and clock frequencies have been increased. The increase in power density and performance has led to heat dissipation. As the transistors consume more power, the temperature of the transistors, which is also known as the junction temperature, increases ((1) in Table 1). Therefore, by taking into account the ambient temperature, the thermal resistance value, and the maximum temperature that the chip can withstand, the maximum allowed power dissipation, which is also called power budget, can be determined.

Another one of the important features of SoC is that the power increases as the ambient temperature increases. The total power can be divided into dynamic and static as (2) in Table 1 and static power or leakage power, which is primarily composed of sub-threshold current, is highly sensitive to temperature as (3) in Table 1 [1] and increases exponentially as the temperature rises [2].

Based on the two main characteristics, an increase in the ambient temperature can increase power, and an increase in power leads to greater heat dissipation. This circular relationship could result in a thermal runaway [3]. Therefore, it is important to ensure the appropriate cooling of the SoC.

Various cooling systems have been designed including heat pipe cooling, liquid cooling, fan cooling, and others. However, these methods are commonly used for CPUs in high end server platforms [4]. Since it is hard to add external cooling methods in embedded systems with limited size and allowed amounts of power consumption, SoC-internal approaches have been generally applied.

Dynamic Thermal Management (DTM) refers to "a range of possible software and hardware strategies while work dynamically to control a chip's operating temperature at runtime" [3]. Most widely used method is CPU Throttling which is constraining voltage and frequency (Dynamic Voltage and Frequency Scaling; DVFS) [5]. It slows down the CPUs whenever it reached trip points. If a critical trip point is reached (e.g. 105 ℃), then thermal shutdown will be triggered. However, as (4) in Table 1, DVFS can only reduce dynamic power. Since leakage power has the property of increasing with the external temperature, it is also important to reduce leakage power. There are several methods that can control leakage power. For example, unused CPUs could be disabled (CPU hotplug out) [6] and unrelated power domains could be turned off [7]. Although the methods reducing leakage power are not widely used to control the temperature of the chip, this paper considers both dynamic and leakage power to minimize consumed power.

Another limitation of traditional methods is that the temperature and power are only considered from the perspective of SoC. However, according to the results in this paper, 34% of the total power consumed by the board was consumed by SoC and the remaining 66% of the power was consumed by other components on the board. Therefore, to reduce the total power, all components on the board should be considered to lower the overall temperature inside the product.

III. Proposed Thermal Management System

The system configured in this paper with Exynos Auto T5123 was same as the virtualized system from [8]. To support eCall, operating systems for AP (Application Processor) and MP (Modem Processor; Communication Processor) were virtualized based on Xen Hypervisor (Fig. 2). The sequence of this system is shown in Fig. 3. When the chip reaches a high junction temperature threshold (e.g. 65 ℃) in a normal situation, it triggers default thermal management such as DVFS. However, if it is in eCall mode and reaches a high temperature, then it goes into the restricted mode to reduce the power consumption as much as possible using all possible methods.

All the methods applicable in the restricted mode can be categorized into the following four categories in Table 2. To reduce the dynamic power of SoC, we minimized the voltage and frequency (CPU Throttling). Also, by setting the logging level to the minimum, CPU loads were reduced. On the other hand, to reduce the static power of SoC, CPUs or power domains could be turned off. However, in this paper, there was

no isolated power domain that could be turned off during call operation and only one unnecessary CPU could be turned off.

At the board level, external devices can be dynamically or statically controlled. For example, the throughput for an Ethernet device could be reduced from 1000 Mbps to 10 Mbps. In this paper, to maximize the effect of optimization, the link connected to the Ethernet device was disconnected.

IV. Measurement and Results

A. Experimental Setup

Temperature and power samples were collected at both ambient room temperature (25 ℃) and high temperature (85 ℃). The high temperature was artificially created using a high-temperature chamber (ESPEC, SH-662), while the room temperature data was collected outside the chamber for comparison. The junction temperature of the chip was measured using a Samsung temperature sensor located near the CPUs which is the hottest spot. On the other hand, the surrounding temperature of the chip was measured by a thermistor (Murata Electronics, NCU15WF104F6SRC) attached to the board. The resolution of the both temperature sensors was 1 ℃ and the data was read for every 10 seconds. Meanwhile, the power consumption data for the chip and its CPUs were measured by a DAQ system (National Instruments, USB-6289) with a sampling rate of 5 Hz and the data for the board was measured by another power measurement system (Monsoon Solutions, AAA10F power monitor) with the sampling rate of 5 Hz.

Test steps are as follows. First, for high temperature, the ambient temperature was gradually increased to 85 ℃ by the high temperature chamber, which takes about 20 minutes. Second, after turning on the board power using the power monitor, we waited 5 minutes for saturation. Third, we began collecting data on the temperature and power consumption. Forth, we triggered a voice call using 3G (UMTS) networks. After maintaining the call for 20 minutes, the measurements were stopped. After completing these test steps without hanging up, we calculated the CPU utilization based on the unscheduled idle time over a period of 30 seconds to compare the CPU loads under different conditions.

B. Results and Discussion

For analysis, the temperature and power consumption data from last 5 minutes were averaged to eliminate the influence of unsaturated data. As shown in Fig. 4, under the high ambient temperature, the temperature and power consumption were remarkably higher than under the room temperature. The junction temperature of the SoC increased by 56 ℃ under the high ambient temperature compared to the low ambient temperature and that of the board increased by 50.9 ℃. Also, the power consumption of the SoC increased by 604.91 mW and that of the board increased by 720.55 mW when comparing the high ambient temperature to the low ambient temperature condition.

The dynamic power and static power of CPUs of the SoC could be estimated based on (5) and (2) in Table 1. The CPU utilization was calculated to be 0.42 at the low ambient temperature (25 ℃) and the high ambient temperature (85 ℃) (Table 3). Therefore, the dynamic power under the low and high temperature is estimated to be 40.52 mW. As the leakage power is the remainder after subtracting the dynamic power from the total power, the leakage power is estimated to be 123.53 mW and 193.58 mW at the low and high ambient temperature, respectively because the total power for CPUs were 164.05 mW and 234.1 mW. The results indicate that the leakage power highly increased at the high ambient temperature and this is a characteristic equally applied to other semiconductor components.

According to Table 4, most of the conditions reduced temperature and power consumption. The CPU throttling was the most effective in reducing the junction temperature and power of SoC. Reducing CPU loads by reducing the log level resulted in lower CPU utilization (0.33) compared to the default condition (0.42) as in Table 3. Although it lowered the power consumed by SoC, it was not obviously observed in the junction temperature results due to the low resolution of the temperature measurement. Contrary to expectations, turning off an unnecessary CPU had no effect. It seems that turning off just one CPU of ARM Cortex-A55 is not that effective, as it is a low power processor and is generally used as a 'LITTLE' CPU in big.LITTLE architectures. On the other hand, disconnecting the Ethernet links reduced the power and temperature of the board, which also resulted in a reduction in the junction temperature of the SoC. Finally, applying all the methods achieved the greatest reduction in the temperature and power, both at the SoC and board level, thereby increasing the power budget based on the maximum junction temperature that the SoC can withstand. The increased power budget would allow the SoC to operate longer before thermal shutdown even at higher temperatures.

V. Conclusion

In this paper, we proposed an integrated thermal management solution for eCall systems that not only reduces the dynamic power of SoC but also static power and a more expanded perspective than SoC. The measured results showed that disconnecting external components was effective in lowering leakage power and the temperature surrounding the SoC, which also resulted in a decrease in the junction temperature of the SoC. The implemented solution could be modified in various ways. For example, the CPU load could be reduced by shutting down applications unrelated to eCall, instead of changing the log level. Also, if it is hard to power off external components to reduce the leakage power, it is possible to restrict their maximum operating speed to reduce the dynamic power. In addition, if the chip is designed to power off unused sub-components, then the leakage power could be significantly decreased by turning them off. In conclusion, the proposed solution could not only be applied to automotive systems but also to all embedded systems which cannot use external cooling.

REFERENCES [1] D. Helms *et al.*, "Leakage models for high-level power estimation." [2] K. Roy *et al.*, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits." [3] D. Brooks *et al.*, "Dynamic thermal management for high-performance microprocessors." [4] G. Xu *et al.*, "Extension of air cooling for high power processors." [5] L. Zhou *et al.*, "Thermal management of ARM SoCs using Linux CPUFreq as cooling device." [6] S. P. Kamat, "Thermal management in embedded systems: A software approach." [7] A. K. Thakur *et al.*, "Physical Implementation of Multi Power Domain SoC Design." [8] B. Park *et al.*, "DRX mode implementation based on virtual machine."

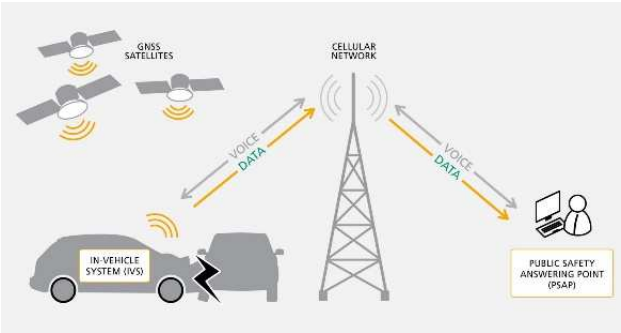


Fig. 1. The concept of eCall systems cited from IZT Labs.

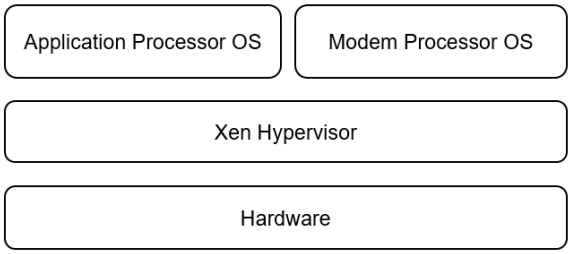


Fig. 2. Overall Architecture of the proposed solution

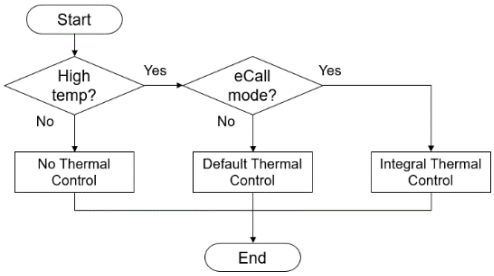


Fig. 3. Overall flow of the system under different conditions.

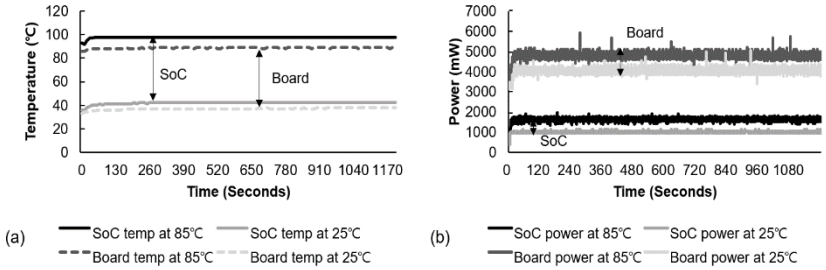


Fig. 4. Temperature (a) and power measurement data (b) under the low and high ambient

$T_j = T_a + R_{ja} \times P$ T_a: ambient temperature R_{ja}: thermal resistance between the junction and ambient temperature P: total power dissipated in the chip.	(1)
$P_{total} = P_{dynamic} + P_{static}$	(2)
$I_{sub-threshold} = \alpha \frac{W}{L} T^2 e^{-\beta V_{th}/T}$ L, W: channel length and width for transistors. V_{th}: threshold voltage, which is the minimum voltage needed for the transistor to begin conducting current. T: absolute temperature, which is the temperature measured using the Kelvin scale where zero indicates absolute zero.	(3)
$P_{dynamic} = C \times V_{dd}^2 \times \mathrm{f}$ V_{dd}: supply voltage f: operating frequency C: switching capacitance that is the ability of a circuit to collect and store energy.	(4)
$P_{dynamic} = \mathrm{Coeff} \times Volt^2 \times Freq \times U\mathrm{til}$ Coeff: coefficient value. The coefficient value for ARM Cortex-A55, taken from the Linux Kernel for Google's GS101 SoC, is 70. Volt: voltage. The voltage for CPUs in this paper was set to 0.9375 V. Freq: clock frequency. The clock frequency for CPUs in this paper was set to 1.568 GHz. Util: CPU utilization	(5)

Table 1. Equations used in this paper

Types	Examples
SoC and Dynamic	CPU Throttling (DVFS), reducing CPU loads, and etc.
SoC and Static	CPU hotplug out (CPU off) and power domain off.
Board and Dynamic	restricting performance of external device
Board and Static	external device off

Table 2. Four types of methods for lowering the temperature of the system

Ambient temp (℃)	Condition	CPU0	CPU1	CPU2	CPU3	Sum
25	Default	0.08	0.09	0.25	0	0.42
85	Default	0.11	0.07	0.24	0	0.42
	Throttle	0.4	0.35	0.67	0	1.42
	Load down	0.05	0.11	0.17	0	0.33
	CPU disable	0.15	0	0.26	0	0.41
	Eth link down	0.1	0.07	0.31	0	0.48
	All	0.57	0	0.42	0	0.99

Table 3. The ratio of idle time per CPU over a 30-second period and total sum under different conditions (with a maximum utilization value of 1 per CPU)

(a)

Ambient temp (℃)	Condition	SoC temp (℃)	Board temp (℃)
25	Default	42.00	37.97
85	Default	98.00	88.87
	Throttle	95.00	88.00
	Load down	98.00	88.63
	CPU disable	98.00	88.70
	Eth link down	97.03	88.00
	All	94.97	88.00

(b)

Ambient temp (℃)	Condition	SoC power (mW)	Board power (mW)
25	Default	1033.30	4052.26
85	Default	1638.21	4772.81
	Throttle	1321.69	4458.70
	Load down	1604.06	4745.07
	CPU disable	1632.37	4774.76
	Eth link down	1627.66	4216.56
	All	1286.63	3868.53

Table 4. Temperature (a) and power (b) under different conditions. Board power includes SoC power.

Wafer-Scale 2D MoS_2 Transistors Using Transfer-Free Location-on-Demand Selective Synthesis

Anthony Cabanillas,[1] Chu Te Chen,[1] Asma Ahmed,[1] Anthony Butler,[1] Yu Fu,[1] Ajay Yadav,[2] Gabriel Lee,[2] Keith Wong,[2]* Fei Yao,[1]* and Huamin Li[1]*
[1]University at Buffalo, The State University of New York, Buffalo, NY, USA
[2]Applied Materials, Inc., Sunnyvale, CA, USA
*Email: keith_wong@amat.com, feiyao@buffalo.edu, huaminli@buffalo.edu

Abstract — Compatible integration of emerging two-dimensional (2D) materials with mature Si-CMOS technology is promising to enable high-performance energy-efficient electron devices. In this work, we exploited wafer-scale, location-on-demand, selective growth of 2D semiconducting transition metal dichalcogenide (TMD), MoS_2, on a SiO_2/Si substrate for transfer-free electron device applications. We investigated the impact of native oxide MoO_3 dielectrics on the performance of MoS_2 field-effect transistor (FET) arrays through a comparative study with SiO_2 dielectrics, and demonstrated great potential of the selective growth of 2D semiconductors to lower down the technological requirement for practical integration with Si-CMOS technology.

I. Introduction

To continue the Moore's law and maintain device miniaturization, 2D semiconducting TMDs have been extensively explored as one of the most promising channel materials. Despite the early success in fundamental science explorations and proof-of-concept device demonstrations [1-4], TMD films with good scalability, uniformity, crystallinity, and compatibility on suitable substrates remain a significant roadblock to the realization of commercially viable TMD-based electron devices [5, 6]. To mitigate this problem, we investigate a location-on-demand selective growth methodology to realize high-quality TMD layers with consistent layer characteristics. Without any wet or dry transfer process which inevitably leads to material degradation, the TMD growth is catalyzed directly at the desired channel area with improved time and cost effectiveness. Specifically, we take 2D MoS_2 as an example, and exploit patterned MoO_3 seedings to enable wafer-scale location-on-demand selective growth of MoS_2 arrays for transfer-free FET applications. We directly observe the chemical vapor deposition (CVD) growth of MoS_2 on micrometer scale within a few seconds (a high growth rate of 0.5 μm/s), and investigate the impacts of the dielectric interfaces (MoO_3, SiO_2, and Al_2O_3) on the synthetic MoS_2 FET performance in terms of on-current density ($J_{D,on}$), field-effect mobility (μ_{FE}), on/off ratio, contact resistance (R_C), transfer length (L_T), and Schottky barrier height (SBH) etc. We find that the polycrystal MoS_2 with MoO_3 dielectric interface has comparable and even superior performance compared to other MoS_2 FETs using selective and non-selective growth, despite the presence of rich grain boundaries (GBs). Our work demonstrates the wafer-scale, transfer-free, location-on-demand selective growth of 2D TMD, which can significantly ease the integration with Si-CMOS processing and paves a feasible way for realizing 2D semiconductor implementation.

II. Device Fabrication and Measurement

MoS_2 growth and its dynamic evolution. First, the MoO_3 seedings were patterned using electron-beam lithography (EBL) and sputtered on p-Si substrates (1-10 Ω·cm) with 285 nm SiO_2 dielectric. Next, the sample was loaded in a two-zone CVD furnace for controlled sulfurization, and the MoS_2 thin films were formed at the seeding areas, which spatial morphology and crystallinity can be manipulated by the CVD synthetic parameters. Especially, the dynamic evolution of MoS_2 growth was monitored by a micro-chamber CVD with time-resolved in-situ microscopy, as shown in Fig. 1(a) and (b). A single crystal MoS_2 domain, up to 10 μm in size, can be grown within ~20 s, indicating a high growth rate of 0.5 μm/s.

Controlled sulfurization. MoO_3 is not only a Mo precursor but also a high-k dielectric (k = ~35) [7]. In principle, MoO_3 can serve as a better dielectric interface to screen the charge impurities and boost the charge transport. Compared to a full sulfurization process which leads to a single crystal "triangle" domain formation, a limited sulfurization process only at the MoO_3 surface creates polycrystal MoS_2 around the MoO_3 seedings, as shown in Fig. 1(c). This "circular" MoS_2 pattern suggests isotropic homogeneity of MoS_2 growth on SiO_2, and the remaining MoO_3 layer provides a native high-k dielectric interface, benefiting the charge transport of the synthetic MoS_2 on top of it. Confocal Raman and photoluminescence (PL) spectroscopies also confirm the signature modes of MoS_2 on both MoO_3 and SiO_2 dielectrics, suggesting excellent uniformity of localized homogeneity.

Large-scale location-on-demand selective growth of MoS_2. Scanning electron microscopy (SEM) and Raman spectroscopy were performed to confirm the MoO_3 seedings (e.g., an array of 5×5 μm^2 squares) and the corresponding MoS_2 growth (e.g., an array of circular MoS_2 with a diameter of ~14 um) across a 1×1 cm^2 area, as shown in Fig. 1(d-g). Atomic force microscopy (AFM) confirms the presence of a non-sulfurized MoO_3 layer which is less than 10 nm, as shown in Fig. 1(h). Excellent uniformity of the synthetic MoS_2 size across a 1×1 cm^2 area was evaluated, as shown in Fig. 1(i). With the increased seeding density (i.e., the miniaturized spacing distance down to 1 μm), the selective MoS_2 growth is still consistent, as shown in Fig. 1(j). All these characterizations suggest excellent reproducibility and uniformity of selective growth for the ease of integration, especially at a large scale.

Device fabrication and measurement. On top of the as-grown MoS_2, Bi/Au electrodes (20 nm/50 nm) were patterned by EBL and sputtered to form a back-gate FET configuration, and Al_2O_3 was deposited by atomic layer deposition (ALD) to form a top dielectric in a top-gate FET configuration. Electrical and photoresponse characterizations were performed using a semiconductor parameter analyzer, a temperature-variable vacuum probe station, and an integrated laser system. With the grounded source, drain current (I_D) was measured as a function of the applied drain, back-gate, and top-gate voltages (V_D, V_{BG}, and V_{TG}), and was normalized in current density ($J_D = I_D/W$ where W is the channel width) for comparison.

III. Device Results and Discussion

MoO_3 dielectric interface. A comparative investigation is performed for the synthetic MoS_2 FETs with the MoO_3 and SiO_2 dielectric interfaces, in terms of output and transfer characteristics (J_D-V_D and J_D-V_{BG}), and a statistical analysis of $J_{D,on}$, subthreshold swing (SS), μ_{FE}, threshold voltage (V_{th}), hysteresis window (ΔV_{th}), and on/off ratios, as shown in Fig. 2(a-j). The linear J_D-V_D characteristics suggest Ohmic contact for both FET types. With the identical synthetic process and device geometry, the MoO_3 interface provides comparable FET performance metrics, and more importantly, in much narrower distributions. On the wafer scale, the MoS_2 arrays formed by the controlled sulfurization at on-demand locations are still reproduceable, which gives the average $J_{D,on}$ of 2 μA/μm, μ_{FE} of 5 cm^2/Vs, and on/off ratio exceeding 10^5, as shown in Fig. 2(k-m). By optimizing the synthetic parameters, our best MoS_2 FET device possesses $J_{D,on}$ of 3 μA/μm, μ_{FE} of 20 cm^2/Vs, and on/off ratio up to 10^6, as shown in Fig. 2(n).

Compared to conventional SiO_2, MoO_3 as both the precursor and native oxide provides a much intimate contact to MoS_2 with less defects, traps, mismatch, strain, or interfacial states, as shown in Fig. 3(a) and (b). These facts are evidenced by the extracted interfacial trap density (D_{it}), as shown in Fig. 3(c). Here D_{it} was calculated from SS [8], and the MoO_3 dielectric interface provides much higher homogeneity compared to the SiO_2 dielectric interface through a wafer-scale statistical analysis. We also fabricated the top-gate MoS_2 FETs, and the synthetic MoS_2 has good interfacial states to enable ALD-produced Al_2O_3 dielectrics, as shown in Fig. 3(d).

Impacts of crystallinity and GBs. Compared to the single crystal triangle MoS_2, the polycrystal circular MoS_2 possesses abundant GBs. To understand the impact of crystallinity and GBs on synthetic MoS_2, we design one MoS_2 FET with the channel being parallel with a GB, and another device with the channel being perpendicular to the same GB, as shown in Fig. 3(e). Our results show comparable FET performance including $J_{D,on}$, on/off ratio, and V_{th}, suggesting negligible impact of GMs on the synthetic MoS_2 in this work.

On-demand geometric manipulation of MoS_2 growth and metal contact improvement. Owing to the well-controlled growth, we can define the geometry of as-grown MoS_2 in arbitrary shapes without any lithography and etching process. For example, we create a long MoO_3 ribbon (5 μm × 200 μm) for transmission line measurement (TLM). The MoS_2 growth is remarkably uniform along the seeding pattern, as shown in Fig. 4(a). It is intriguing that the more involvement of MoO_3 interface, in contrast to SiO_2 interface, can lead to greater improvement of the metal contact condition such as the lowering of R_C and T_L, but keep the MoS_2 channel resistance (R_{CH}) intact, as shown in Fig. 4(b-d). The SHB, evaluated from a temperature-variable measurement, suggests a flat-band barrier height of 12 meV with the MoO_3 interface, which is much lower than that with the SiO_2 interface (60 meV) and is consistent with the improvement of metal contact condition. Moreover, the extraction of R_C in this work was the first report of any selectively grown MoS_2, thanks to the ease of the MoS_2 geometric definition. The metrics such as R_C of ~200 kΩ·μm and L_T of ~0.1 μm in this work can be further improved by doping, phase transition, and semimetal contact which are well explored for the non-selectively grown MoS_2.

Photoresponse. The back-gate MoS_2 FET arrays can act as 2D phototransistors, and their photoresponsive performance is evaluated, as shown in Fig. 5. Both the power-dependent static photocurrent (PC) generation and time-resolved high-speed photo-switching dynamics suggest excellence photoresponse of MoS_2 FETs in the visible spectrum.

Performance benchmarking. The MoS_2 FETs using location-on-demand direct CVD synthesis in this work were benchmarked with other MoS_2 FETs using various selectively grown techniques (i.e., patterned seeding, plasma treatment, and laser annealing), as shown in Fig. 6(a-d). Our work shows superior metrics in $J_{D,on}$, μ_{FE}, and on/off ratios, owing to the excellent MoO_3 dielectric interface. Meanwhile, our devices were also compared with other MoS_2 FETs using non-selective grown strategies on special substrates (e.g., HfO_2, sapphire, Au, polymer, and glass), as shown in Fig. 6(e). Our wafer-scale location-on-demand MoS_2 FETs don't require any transfer process, and the performance metrics, such as μ_{FE} and on/off ratios, are comparable with the state-of-the-art MoS_2 FETs.

IV. Conclusion

In this work, we presented wafer-scale MoS_2 FET arrays using transfer-free location-on-demand selective growth. This technique allows time- and cost-efficient synthesis (0.5 μm/s) of 2D semiconductor channel arrays at designed locations, and the controlled sulfurization creates MoO_3 dielectric interfaces to enhance FET performance. Our devices show comparable and even superior performance, such as $J_{D,on}$, μ_{FE}, and on/off ratios, compared to other MoS_2 FETs using selective and non-selective growth, and demonstrate great potential to ease the integration of 2D semiconductors for various electron devices.

Acknowledgment

The authors acknowledge support from the SUNY Applied Materials Research Institute (SAMRI) and the National Science Foundation (NSF) under Award ECCS-1944095.

References

[1] S. Das et al., *Nat. Electron.*, 4, 786 (2021).
[2] M. Liu et al., *ACS Nano*, 15, 5762 (2021).
[3] H. N. Jaiswal et al., *Adv. Mater.*, 32, 2002716 (2020).
[4] M. Liu et al., *IEEE IEDM*, 251 (2020).
[5] N. Briggs et al., *2D materials*, 6, 022001 (2019).
[6] Y. Zhang et al., *Adv. Mater.*, 31, 1901694 (2019).
[7] B. Holler et al., *Adv. Electron. Mater.*, 6, 2000635 (2020).
[8] C. J. McClellan et al., *ACS Nano*, 15, 1587 (2021).

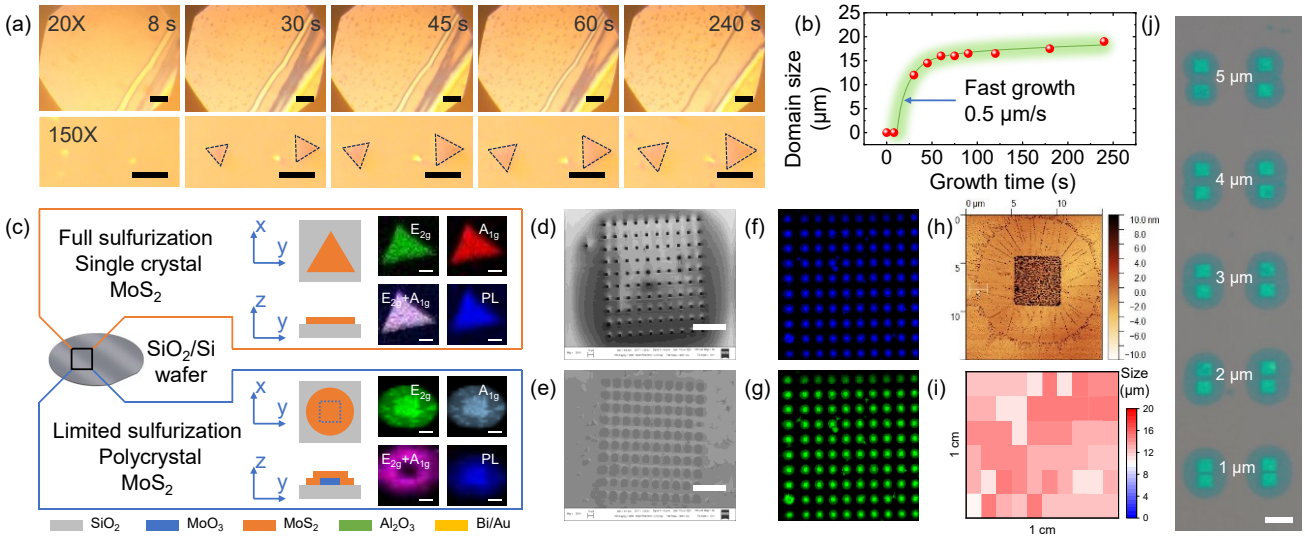


Fig. 1 Synthesis and characterization of wafer-scale location-on-demand MoS_2 growth on SiO_2/Si substrates. (a, b) Time-resolved in-situ microscope images to visualize the MoS_2 growth evolution, and the extracted growth rate up to 0.5 μm/s. Scale bar: 50 μm (top) and 20 μm (bottom). (c) Comparison of single crystal "triangle" MoS_2 and polycrystal "circular" MoS_2 synthesized by full sulfurization and limited sulfurization, respectively, and the corresponding Raman/PL spectroscopy mapping. Scale bar: 5 μm. (d, e) SEM images of a MoO_3 seeding array before growth and a MoS_2 array after growth. Scale bar: 60 μm. (f, g) E_{2g} and A_{1g} Raman spectroscopy mapping for the large-scale synthetic MoS_2 arrays. (h) AFM mapping of a circular MoS_2 with the MoO_3 seeding at the center. (i) Spatial uniformity of the MoS_2 sizes across a 1 × 1 cm^2 area. The average size is about 14 μm. (j) Microscopy image of the circular MoS_2 thin films which merges as the spacing distance of the MoO_3 seedings scales down from 5 to 1 μm. Scale bar: 10 μm.

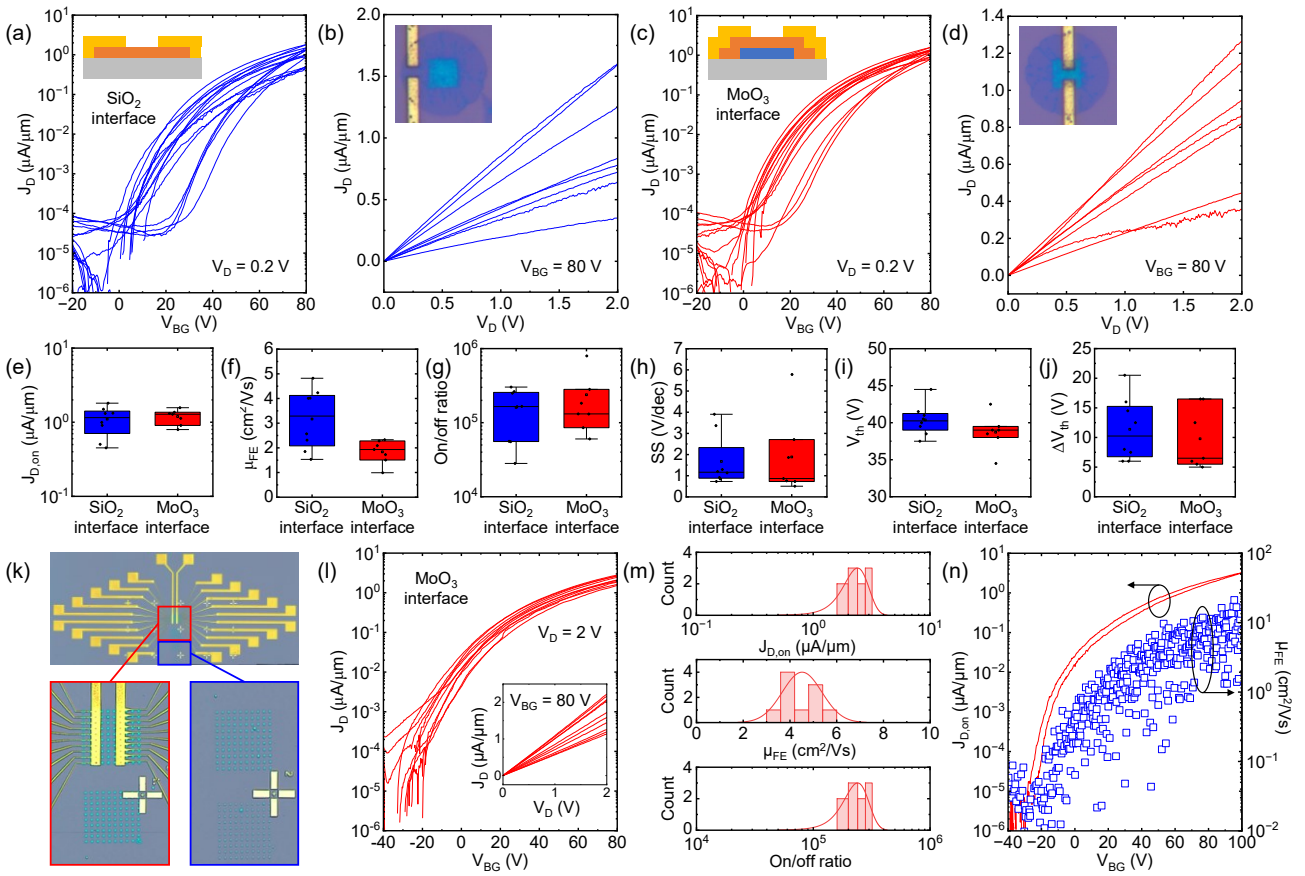


Fig. 2 Comparison of synthetic MoS_2 FETs with SiO_2 and MoO_3 dielectric interfaces. (a-d) Output and transfer characteristics of the MoS_2 FETs with SiO_2 and MoO_3 dielectric interfaces. Insets: The corresponding cross-sectional schematics and microscope images of the devices. (e-j) Statistical analysis of the MoS_2 FETs with SiO_2 and MoO_3 dielectric interfaces, including $J_{D,on}$, μ_{FE}, on/off ratio, SS, V_{th}, and ΔV_{th}. (k-m) Microscope images and transfer characteristics of the wafer-scale MoS_2 FET arrays with MoO_3 dielectric interfaces, and the corresponding statistical analysis including $J_{D,on}$, μ_{FE}, and on/off ratio. Inset of (l): The corresponding output characteristics. (n) The best back-gate MoS_2 FET with MoO_3 dielectric interfaces possesses $J_{D,on}$ of 3 μA/μm, μ_{FE} of 20 cm^2/Vs, and an on/off ratio up to 10^6.

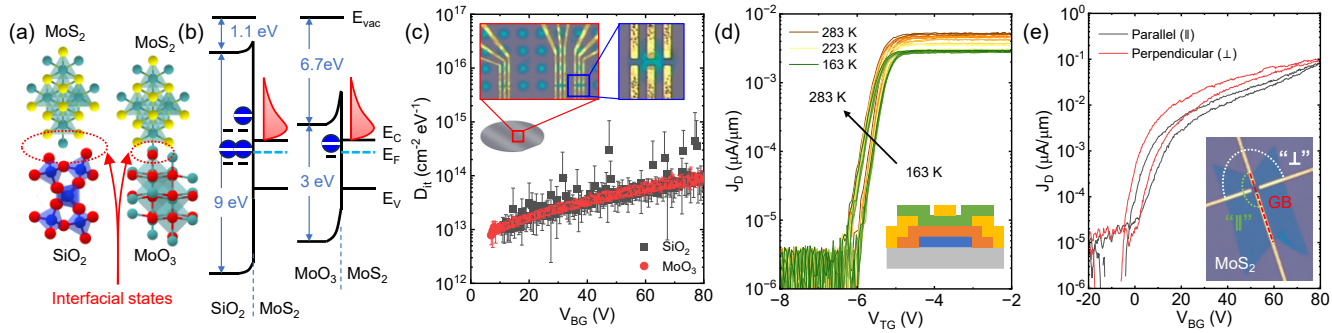


Fig. 3 Impacts of dielectric interfaces and GBs on MoS_2 FET performance. (a-c) Schematics, energy band diagrams, and the extracted D_{it} as a function of V_{BG} for MoS_2/SiO_2 and MoS_2/MoO_3 interfaces. Here E_C, E_V, E_F, and E_{vac} are the conduction band minimum, valence band maximum, Fermi level, and vacuum level, respectively. Inset of (c): Microscope images of wafer-scale MoS_2 FET arrays. (d) Temperature-variable transfer characteristics of a top-gate MoS_2 FET with the MoO_3 bottom dielectric and Al_2O_3 top dielectric. (e) Comparison of MoS_2 FETs with the channels parallel with and perpendicular to a GB.

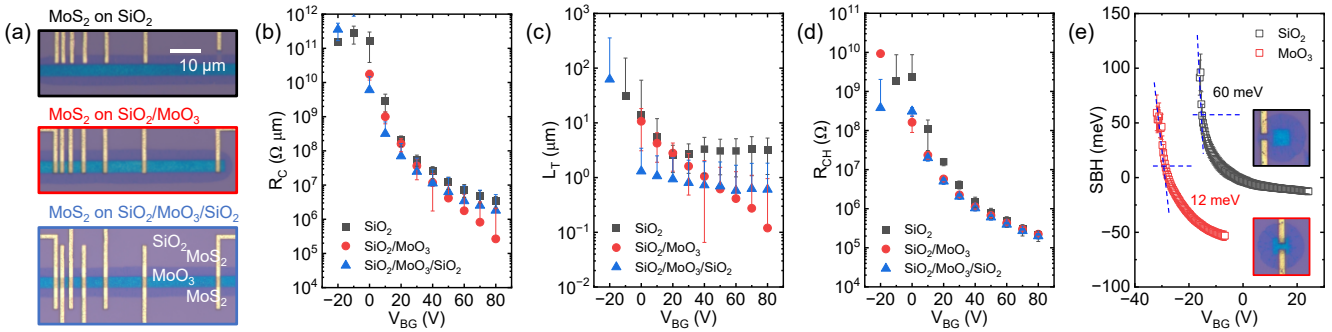


Fig. 4 Metal contacts with SiO_2 and MoO_3 dielectric interfaces. (a-d) Microscope images of MoS_2 TLM devices with different SiO_2 and MoO_3 dielectric interfaces, and the extracted R_C, L_T, and R_{CH}. (e) The extracted SBH as a function of V_{BG}. Inset: Microscope images of the corresponding MoS_2 FETs.

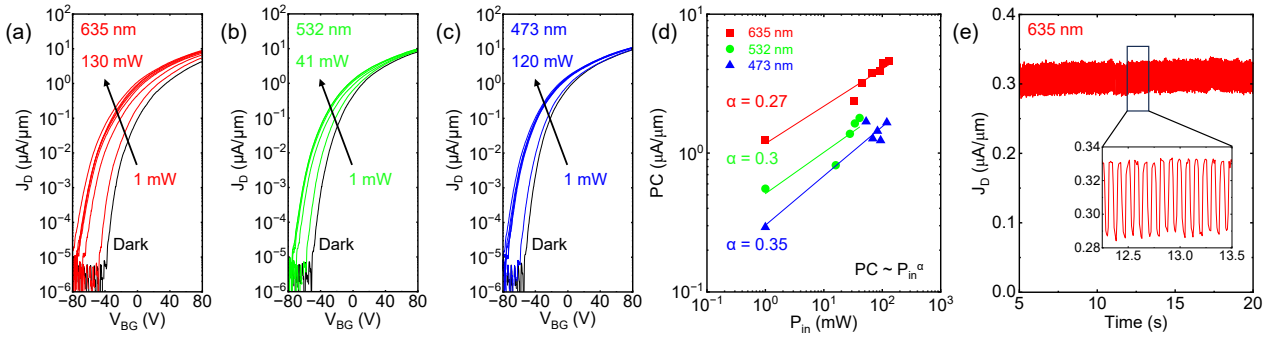


Fig. 5 Photoresponse of MoS_2 FETs as photodetectors. (a-c) Transfer characteristics of a MoS_2 FET under red, green, and blue laser illumination. (d) The extracted PC as a function of the input power (P_{in}). (e) Time-resolved high-speed photo-switching characteristics under the red laser illumination.

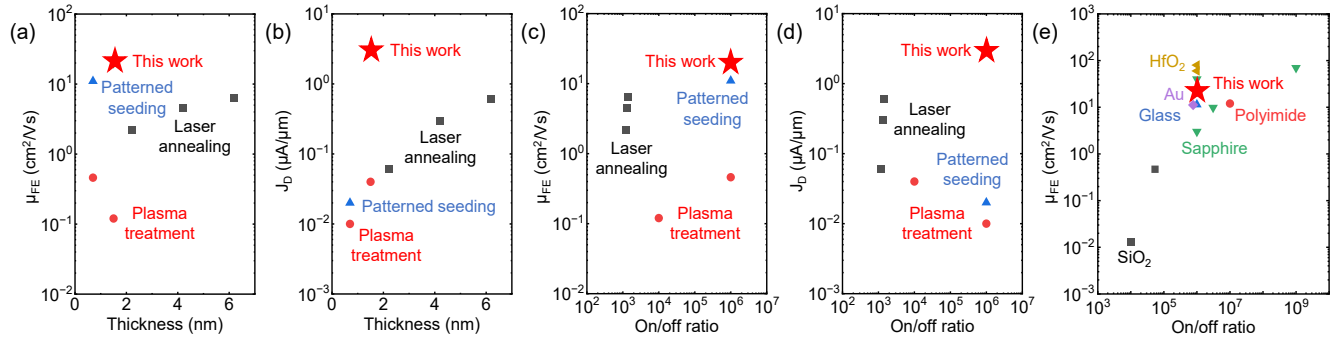


Fig. 6 Benchmarking of the MoS_2 FETs in this work with other state-of-the-art synthetic MoS_2 FETs in terms of $J_{D,on}$, μ_{FE}, on/off ratio, and thickness. (a-d) Benchmarking with other location-on-demand selectively synthesized MoS_2 FETs on SiO_2 substrates. (e) Benchmarking with other MoS_2 FETs grown on special substrates (e.g., SiO_2, HfO_2, sapphire, Au, polymer, and glass) which have no location-on-demand selectivity or require a wet/dry transfer process. The references include: S. Park et al., *ACS Nano*, 14, 8485 (2020); X. Chen et al., *Nanoscale*, 8, 15181 (2016); H. J. Kim et al., *Small*, 13, 1702256 (2017); G. H. Han et al., *Nat Commun.*, 6, 6128 (2015); N. B. Shinde et al., *ACS Appl. Nano Mater.*, 3, 7371 (2020); T. W. Kim et al., *Nanotechnology*, 28, 18LT01 (2017); J. Mun et al., *ACS Appl. Electron. Mater.*, 1, 4, 608 (2019); P. Yang et al., *Nat Commun.*, 9, 979 (2018); H. Yu et al., *ACS Nano*, 11, 12001 (2017); Q. Wang et al., *Nano Lett.*, 20, 7193 (2020); Y. F. Lim et al., *ACS Nano*, 12, 1339 (2018); B. M. Bersch et al., *2D Mater.*, 4, 025083 (2017); P. Yang et al., *ACS Nano*, 14, 5036 (2020); K. S. Kim et al., *Nature*, 614, 88 (2023).

Ultra-low power cryogenic field effect transistor utilising high mobility compressively strained germanium on silicon

M. Myronov[1], P. Waldron [2], and S. Studenikin[2]
[1]Physics Department, The University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK, e-mail: M.Myronov@warwick.ac.uk
[2]National Research Council of Canada, 1200 Montreal Rd., Ottawa, Ontario K1A 0R6, Canada

***Abstract*—**An ultra-low power cryogenic field effect transistor (Cryo-FET), utilizing a high mobility compressively strained germanium on silicon (cs-GoS) material platform, is presented. The device demonstrates beyond state of the art electrical characteristics at cryogenic temperature of 4.2 K, including superior stability, absence of any hysteresis in I-V characteristics, low subthreshold swing, small leakage current, and very low power dissipation in pW range suitable for building large cryo-electronic circuits containing over 1 million transistors dissipating <100μW and still being withing the cooling power budget of regular dilution refrigerators operating down to <100mK. Our findings highlight the potential of cs-GoS Cryo-FETs for advanced and scalable cryogenic applications such as quantum and classical computing, and deep space exploration.

I. Introduction

The growing demand for ultra-low power and high-performance electronic devices at cryogenic temperatures is driven by advancements in quantum computing, space technology, and cryogenic sensing.[1,2] Silicon, the conventional semiconductor on the market, commonly used in electronics faces significant challenges in maintaining required performance and efficiency at these temperatures, in particular, due to carrier freeze-out effect of dopants, charge instabilities, and large heat dissipation, which is not compatible with cryogenic instrumentation limited by the fundamental thermodynamic laws. This paper demonstrates the unique potentials of emerging cs-GoS material platform for Cryo-FET and spin qubit devices applications, building on recent research advancements [3, 4] that highlight the material's superior properties. The cs-GoS grown on full size silicon wafers, up to 200 mm diameter, is fully compatible with silicon foundries and offers the unique combination of properties that are important for quantum and cryogenic electronics applications.[3] The most important properties include the record high carrier mobility of holes and favorable band structure. The compressive strain in germanium enhances its hole mobility up to 4.3 x10^6 $cm^2V^{-1}s^{-1}$ and reduces the hole's effective mass, m*, down to 0.035m_0, making it an ideal candidate for low-temperature electronics.[4] In particular, smaller m* results in larger hole de-Broglie wavelength that makes electronic devices less sensitive to fluctuations and more reproduceable.[3]

II. Materials Synthesis

For the reported research, undoped cs-GoS heterostructures were grown by reduced pressure chemical vapor deposition (RP-CVD) on a relaxed $Si_{0.15}Ge_{0.85}$ buffer on a standard Si(001) wafer of 150 mm diameter. A schematic cross section of the heterostructure and fabricated FET-like gated Hall bar, with its source (S) and drain (D) ohmic contacts and gate (G) stack is shown in Fig. 1. All epilayers were intentionally undoped. Accumulated by the negative gate voltage, holes are confined in the 30 nm thick undoped and compressively strained Ge (cs-Ge) quantum well (QW), acting as an active p-channel for mobile holes, positioned ~300 nm below the surface.

III. Devices Microfabrication

Double-gated Hall bars, enabled to work in a Cryo-FET mode, were fabricated using standard UV lithography, dry etching and thin film deposition techniques. Fig. 2 shows an optical image of the device in a shape of a gated Hall bar with its channel oriented along the <110> in-plane crystallographic direction defined by the mesa structure etched in a Cl_2/Ar plasma. The Hall bar's channel width is 100 μm and the distance between source and drain contacts is 1000 μm. In order to reach an ultra-low power dissipation, we employ a high-quality undoped cs-GoS material stack, which is naturally not conductive at cryogenic temperatures. Therefore, to replace commonly used doping technology, we introduce a specially designed S and D accumulation gates (Figs. 1-2) allowing us to generate mobile carriers in the contact regions next to the active cs-Ge channel, as shown in Fig.1.

The alloyed AlSiGe ohmic contacts were prepared by evaporating a 120 nm thick Al film and annealing it at ~275 °C in N_2 ambient for 30 min. The injection contacts operating in the enhancement mode show low resistivity and excellent linear ohmic behaviour at cryogenic temperatures. Fig. 3 shows activation characteristics of the S and D contacts separately, with all other contacts being grounded. The S-D contacts accumulation gate is isolated from the contact metallisation and the Schottky gate by a 50 nm AlO_3 dielectric deposited by Atomic Layer Deposition at 200 °C. Superior reproducibility of contacts is evident from Fig.3 with a threshold voltage V_{ACC}=−230 mV. Further measurements of Cryo-FET characteristics are performed at V_{ACC} = −350 mV. The top

accumulation Schottky gate is made of 20 nm Ti followed by a 200 nm Au layer.

IV. Electrical Characterization

The Cryo-FET in a shape of a gated Hall bar with additional potential probes (2-3 in Fig.2) allows us to measure intrinsic material's properties like free-carrier mobility, carrier density, their mean free path, etc. [4], which are important for understanding and modeling these new devices. At the same time, we use our device in FET mode in order to measure its input and output characteristics. The electrical performance of the Cryo-FET device was carried out at temperature T=4.2 K.

There is a common issue of undesirable charges at the interface of a semiconductor and a dielectric, which leads to large instabilities and threshold voltage shifts in gated devices, including FETs.[4-6] In order to achieve superior gate control and stability we avoid using dielectrics in the Cryo-FET channel region and, instead, introduce the Schottky gate for cs-GoS devices, for the first time. Fig. 4 shows an I-V characteristic of Schottky gate current I_G versus gate voltage V_G in forward direction. The minimum leakage current is below the detectable limit of the experimental measurement setup. Some small leakage current appears at V_G below −150 mV, within a few pA's, and starts growing exponentially for voltages below −220 mV, following expected Schottky contact behavior in enhancement mode. It should be noted, this is an extremely low value for a relatively large area of 0.1 mm^2 of our test device. Also, we observed very reproduceable sharp peaks, evident in Fig. 4, for 4 back-and-forth sweeps, which may be due to resonant tunneling through deep level states. Very high hole mobility of 1.5×10^6 cm^2 V^{-1} s^{-1} is measured at V_G= −350 mV, as seen in Fig. 5. As a consequence, holes mean free path, shown in Fig. 6, reaches 8 μm. It indicates that a Cryo-FET device with a gate length below this value will operate in a ballistic regime, which may lead to even lower heat dissipation.

A 2D color-map plot of typical input characteristics of the Cryo-FET is shown in Fig. 7. $I_{SD} - V_{SD}$ line-traces at different V_G, extracted from this map, are presented in Fig 8. Due to very high mobility, I_{SD} saturates very fast at very low $V_{SD} \sim -10$ mV. Figure 9 shows a hysteresis check of the Cryo-FET characteristics in semi-logarithmic scales, and Fig. 10 in linear scales. These results indicate the superior performance of the Cryo-FET device: no observable hysteresis, very low threshold voltage, V_{TH} = −15 mV, and very low sub-threshold swing (SS) = ~3 mV/dec. The Cryo-FET reveals reliable gate control with undetectable minimum leakage currents. The off current of the FET is below 1 pA, which is limited by the cryogenic measurements experimental setup including electronics, circuits, cables and wiring. It is necessary to note that no temperature rise was observed during the measurements, discussed above, indicating an ultra-low power dissipation of the Cryo-FET device.

V. Performance Comparison with Conventional Materials

A comparative analysis was conducted between cs-GoS Cryo-FETs and FETs based on traditional semiconductor materials. [2,7] The Cryo-FET shows superior performance thanks to the very high hole mobility and material stack quality resulted in very low subthreshold swing, very low Schottky gate leakage current in the enhancement mode, very low V_{TH}, absolute absence of any hysteretic behavior, and superior stability of all FET characteristics at cryogenic temperatures. The Cryo-FET characteristics were repeatedly obtained during several days and no measurable drift of any characteristic was observed. We estimate that the Cryo-FET can operate in a very low dissipation power regime, with estimated ~50 pW of Joule heat dissipation. It means, an ULSI circuit containing, e.g., one million of such transistors would dissipate just ~50 μW heat power, which is within the cooling power of modern cryogenic-free dilution refrigerators operating down to < 100 mK. These advantages position the cs-GoS as a promising material platform for cryogenic electronic applications. Demonstrated unique properties of the cs-GoS Cryo-FET open up new possibilities for cryogenic classical and quantum electronic systems. Potential applications include low-power quantum computing circuits, cryogenic sensors, and deep space electronics. The high mobility and ultra-low power consumption of cs-GoS FETs are particularly beneficial for high-speed cryogenic electronics where energy efficiency and performance are critical.

VI. Conclusion

This paper presents the development and characterization of ultra-low power and high-performance p-type Cryo-FET based on the new cs-GoS material platform. The exceptional electrical properties of the cs-GoS material stack at cryogenic temperatures make it a very promising candidate for next-generation cryogenic classical and quantum electronics. Future work will focus on further optimization of devices, fabrication technology, and their integration with other cryogenic components such as blocks of qubits performing error corrections or other quantum computing algorithms.

Acknowledgment

The Authors thank the financial support of Quantum Sensing Program of NRC Canada and EPSRC UK.

References

[1] R. Nikandish, E. Blokhina, D. Leipold, and R. B. Staszewski, "Semiconductor Quantum Computing: Toward a CMOS quantum computer on chip," *IEEE Nanotechnology Magazine,* vol. 15, no. 6, pp. 8-20, 2021.

[2] G. Kiene *et al.*, "A 1-GS/s 6–8-b Cryo-CMOS SAR ADC for Quantum Computing," *IEEE Journal of Solid-State Circuits,* pp. 1-12, 2023.

[3] M. Myronov, P. Waldron, P. Barrios, A. Bogan, and S. Studenikin, "Electric field-tuneable crossing of hole Zeeman splitting and orbital gaps in compressively strained germanium semiconductor on silicon," *Communications Materials,* vol. 4, no. 1, p. 104, 2023.

[4] M. Myronov *et al.*, "Holes Outperform Electrons in Group IV Semiconductor Materials," *Small Science,* vol. 3, p. 2200094, 2022 202.

[5] J. Ferrero *et al.*, "Noise reduction by bias cooling in gated Si/Si_x Ge_{1-x} quantum dots," *Applied Physics Letters,* vol. 124, no. 20, 2024.

[6] M. Meyer *et al.*, "Electrical Control of Uniformity in Quantum Dot Devices," *Nano Letters,* vol. 23, no. 7, pp. 2522-2529, 2023/04/12 2023.

[7] X. Xue *et al.*, "CMOS-based cryogenic control of silicon quantum circuits," *Nature,* vol. 593, no. 7858, pp. 205-210, 2021/05/01 2021.

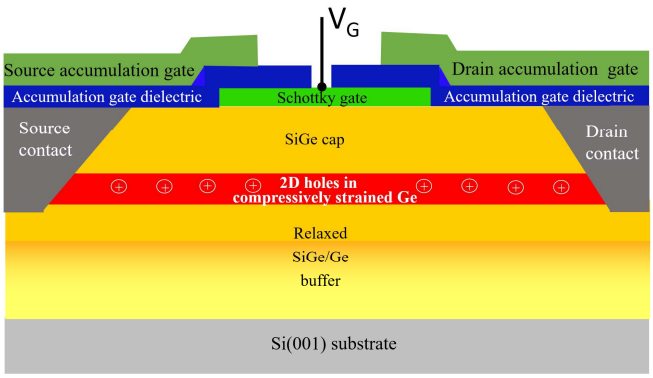


Fig. 1. Cross-section schematic of the Schottky Cryo-FET on undoped cs-GoS epiwafer, employing double-gates technology. The drain and source accumulation gates replace doping or ion-implantation techniques for contact activation, which would deteriorate electrical performance of the device at cryogenic temperatures.

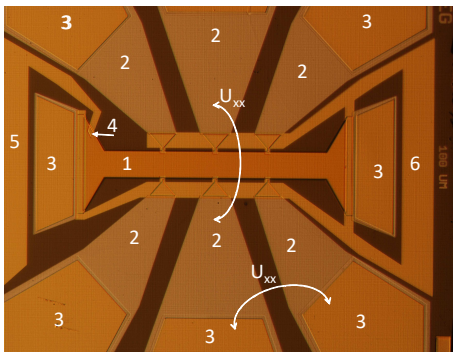


Fig. 2. An optical image of the double-gate Hall bar device which also is used to work as a Cryo-FET. Additional potential probes (2) allow to measure potentials U_{xx} and U_{xy} and extract transport characteristics [4] of the active channel controlled by the Schottky-contact gate (1) in enhancement mode. (3) Ti/Au bonding pads to source-drain and potential probe contacts; (4) via through the gate dielectric to contact Schottky gate; (5) Schottky gate bonding pad; (6) the contacts accumulation gate. The Cryo-FET p-channel width (W = 100 µm) to length (L = 1000 µm) ratio is 0.1.

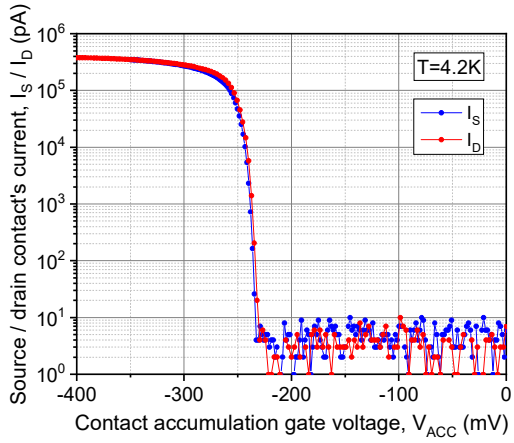


Fig. 3. Source-drain contacts initialization. Individual initialization traces of S and D contacts with all other contacts grounded, Schottky V_G=−100 mV. dielectric thickness 50 nm. In all experiments V_{ACC}= −350 mV.

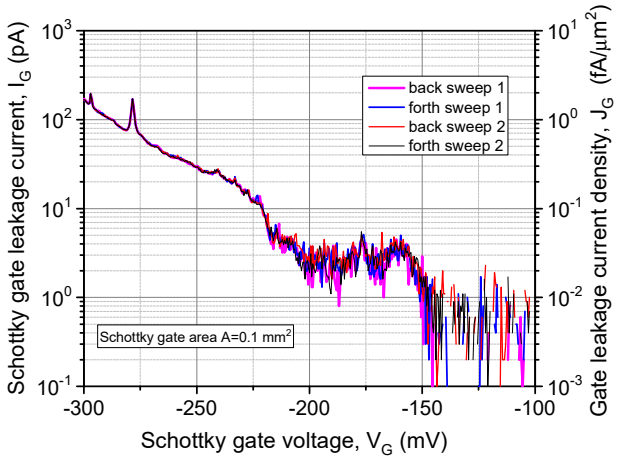


Fig. 4. Schottky-contact gate current in forward bias, so-called the enhancement mode operation. Very small current is detected as an indicator of superior quality of the Schottky contact to the cs-GoS material stack, opening new potentials for cryogenics electronics.

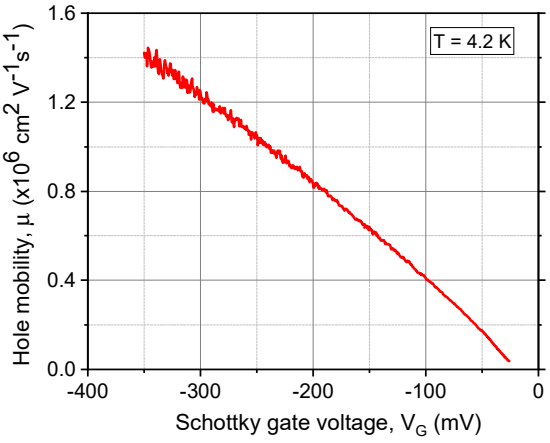


Fig. 5. Hole mobility of free carriers in the p-channel of the Cryo-FET device at T=4.2K.

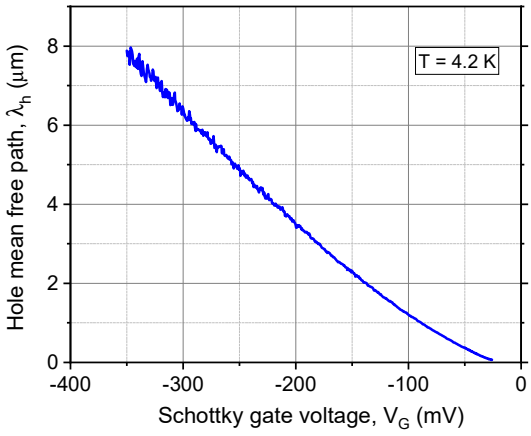


Fig. 6. Mean free path of free carriers as a function of the Schottky gate voltage in forward bias direction, i.e. the enhancement operation mode.

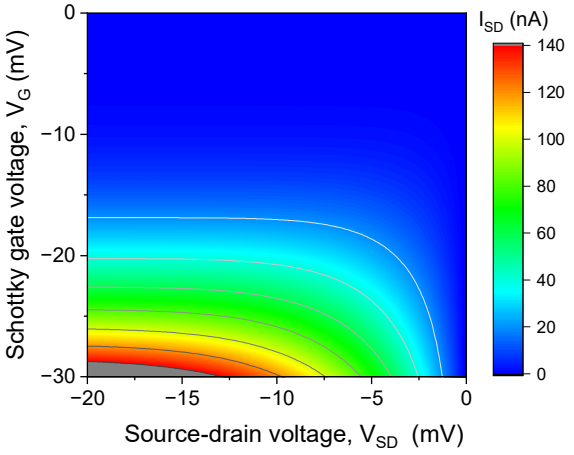


Fig. 7. 2-D color map of the cs-GoS Cryo-FET p-channel enhancement mode source drain current, $I_{SD,}$ characteristics as a function of V_G and V_{SD}, measured at T=4.2K.

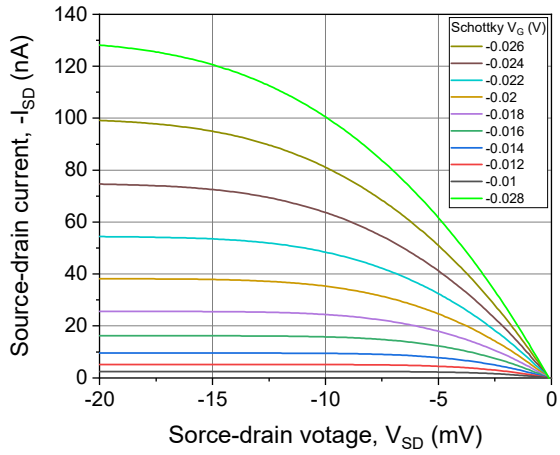


Fig. 8. Typical cs-GoS Cryo-FET p-channel enhancement mode I_{SD} – V_{SD} characteristics at gate voltage, V_G, varied from −10 mV to −28 mV, measured at 4.2 K. I_{SD} increases sharply and saturates at very low V_{SD}, for a given V_G, due to very high hole mobility, shown in Fig. 5.

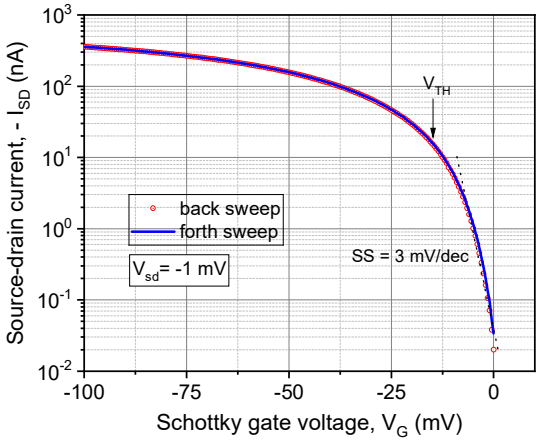


Fig. 9. Typical cs-GoS Cryo-FET p-channel enhancement mode forth and back sweep I_{SD} – V_G characteristics at low drain bias voltage I_{SD} = −1 mV, measured at 4.2 K. I_{SD} is plotted on a logarithmic scale. Both traces lay absolutely on top of each other, i.e., no noticeable hysteresis shifts are detected, indication of the absence of any carrier traps in the whole active Cryo-FET region including bulk and interface. I_{SD} increases sharply due to very high hole mobility, see Fig. 5. The dotted line illustrates the determination of the subthreshold swing SS = 3 mV/dec.

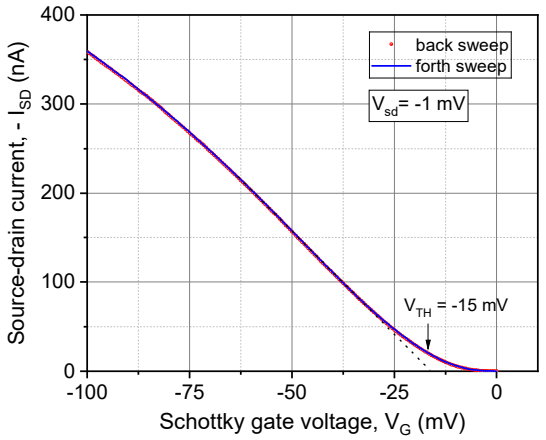


Fig. 10. Typical cs-GoS Cryo-FET p-channel enhancement mode forth and back sweep I_{SD} – V_G characteristics at low drain bias voltage I_{SD} = −1 mV, measured at 4.2 K. I_{SD} is plotted linear scale. No hysteresis is visible. I_{SD} increases sharply due to very high hole mobility, shown in Fig. 5. The dotted line illustrates the determination of the linearly extrapolated threshold voltage V_{TH} = −15 mV. Detailed understanding of the cs-GoS Cryo-FET requires careful modeling and theoretical attention.

New design concept of the 4H-SiC planar MOSFET with the narrowest cell-pitch down to sub-3 μm

Zhi Lin[1*], Da Wang[1], Huan Ning[1], Yuxi Zhang[1], Miao Chen[1], Shengdong Hu[1] and Jian Wu[2]
[1]School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China
[2]Jie Ping Fang Semiconductor (Shanghai) Co., Ltd., Shanghai, China
*E-mail: linzhi@cqu.edu.cn

***Abstract*—**The specific on-resistance ($R_{ON,\ SP}$) of the low-voltage 4H-SiC planar MOSFET can be reduced by shrinking its cell-pitch. In this paper, a new design concept to shrink the cell-pitch of the 4H-SiC planar MOSFET is proposed. The narrowest cell-pitch is shrunken to sub-3 μm by removing the P-plus region and contact metal of some cells. A dumbbell-sharped layout is adopted to realize the proposed 1.2 kV-class SiC MOSFET. Its on-state capability is close to the hexagonal device and its blocking capability is comparable to the linear device. Compared with the traditional linear device, its $R_{ON,\ SP}$ is reduced from 3.13 mΩ·cm^2 to 2.60 mΩ·cm^2 based on a 4.8 μm process platform. And its breakdown voltage only decreases by 15 V. It provides a solution to achieve a small cell-pitch for the SiC planar MOSFET.

I. INTRODUCTION

With the innovation of the device structure and the development of the process technology, the performance of the 4H-SiC MOSFET has been continuously improved [1-5]. However, the high efficiency energy conversion requires further improvement of its performance. At present, an urgent need of the 4H-SiC planar MOSFET is to reduce its specific on-resistance ($R_{ON,\ SP}$). By reducing $R_{ON,\ SP}$, the device with the same rated current will have a smaller chip area. This brings at least two benefits. Firstly, the cost of a single chip will go down because more chips can be fabricated on the same wafer. Secondly, the parasitic capacitances, as well as the switching loss, will decrease. Then, the efficiency of the electrical energy conversion will increase. For low-voltage (e.g. 1.2kV) SiC planar MOSFETs, the channel resistance occupies the largest proportion of $R_{ON,\ SP}$, as shown in Fig. 1. So, many efforts have been made to reduce the channel resistance. The two main methods are respectively increasing the mobility of carriers in the channel [6] and reducing the cell-pitch W_{cell}. The latter is illustrated in Fig. 2. Now, the cell-pitch of advanced SiC planar MOSFETs is less than 4 μm and approaching 3 μm. However, reducing the cell-pitch to sub-3 μm brings great challenges to the process technology. One solution is to use the trench gate structure. However, its reliability is not as good as the planar device. To alleviate the process requirement of reducing the cell-pitch, new device structures must be developed. In this paper, we propose a new design concept to reduce the cell-pitch of the 4H-SiC planar MOSFET. The narrowest cell-pitch is less than 3 μm based on a 4.8 μm process platform.

II. DEVICE CONCEPT AND STRUCTURE

Theoretically, in order to shrink the cell-pitch W_{cell}, all sizes can be reduced, including the JFET width W_{JFET}, the channel length L_{CH}, the gate-source overlap length L_{OV}, the gate-to-contact space L_{GC}, and the contact width L_{CT} (Fig. 2). However, these reductions are inevitably constrained by device performances and the process capability. For example, W_{JFET} is limited by the spacing of hard masks. And, reducing W_{JFET} will increase the JFET resistance. A too small L_{CH} will cause a high leakage current. L_{OV} and L_{GC} are decided by the alignment accuracy of lithography. Reducing L_{CT} increases the difficulty of metal filling. Fig. 3 shows schematically the proposed structure, in which the P-plus regions and contact metals of some cells are removed. That is, L_{GC} and L_{CT} on both sides of the cell are directly reduced to 0 μm. Therefore, the cell-pitch $W_{cell}' = W_{cell} - (2L_{GC} + L_{CT})$ is greatly reduced. For example, the minimum cell-pitch of a popular commercial process platform is 4.8 μm. Once the P-plus region and contact metal are removed, the cell-pitch W_{cell}' can be shrink to 2.8 μm. Other sizes keep unchanged and the design rules are not violated. The proposed structure alleviates the requirement of the process control accuracy.

Of course, in the shrunken cells, the P-well regions should not be floating. In practice, their P-well regions can be grounded periodically in the *z*-direction. The layout arrangement is shown in Fig. 4. It is like a dumbbell shape. The ends of the dumbbell are open. P-plus regions are placed and connected to the source pad through the contact metal. Fig. 4 also shows layouts of the conventional linear and hexagonal layouts. The channel density of the proposed layout is larger than that of the linear layout and smaller than that of the hexagonal layout. In the dumbbell-sharped layout, the P-well regions, as well as channels, are zigzags. Its JFET width keeps the same everywhere. The electric field crowding effect at the P-well corner is much less than that in the hexagonal layout, which is benefit for its blocking characteristic.

The simulated cross-section of the dumbbell-sharped device along the line AA' in Fig.4 and on-state current flowlines are shown in Fig. 5(a). The corresponding electron current densities in the JFET region and under the P-well region are displayed in Fig. 5(b). Since the P-well region of the shrunken cell is narrowed, more current flows under it. It is reasonable to shrink the cell further by reducing the N-plus length $2L_{OV}$. Fig. 6 shows the simulated $R_{ON,\ SP}$ at various cell-pitches by using calibrated models. Obviously, the smaller

W_{cell}', the smaller $R_{ON, SP}$. However, W_{cell}' cannot be reduced without limit, or the N-plus region will disappear, as illustrated in Fig. 7. Modern SiC planar MOSFETs usually use the self-aligned technology to control precisely the channel length. As shown in Fig. 7, after the P-well implantation step, spacers are deposited at all sides of the hard masks. If two hard marks are too close, spacers on the adjacent sides will merge. In the following N-plus implantation step, the phosphorus ions will be blocked by the merged spacer. They cannot reach the underlying semiconductor region. Then, the N-plus region disappears. Another important result from Fig. 6 is that $R_{ON, SP}$ can be reduced to below 2 mΩ·cm^2, if all cells are shrunken to 2.4 μm.

III. Experimental Results and Discussion

Three type of devices using linear, hexagonal and dumbbell-sharped cells respectively are fabricated. Their layouts follow the same design rule. Their active areas are all 0.2 mm^2. Fig. 8-13 demonstrate the measured static characteristics of them. Fig. 8 compares the transfer characteristics of three devices with different cell layouts. The extracted V_{th} of the linear, hexagonal and dumbbell-sharped devices are 2.7 V, 2.5 V, 2.4 V, at I_D = 100 μA, respectively. The current of the hexagonal and dumbbell-sharped devices are larger than that of the linear device. This is in line with the channel density.

Fig. 9 compares the blocking characteristics of the three devices. Their extracted breakdown voltages (BV) at various leakage currents are compared in Fig. 10. They are all exceeds 1400 V. Both the dumbbell-sharped device and the linear device exhibit hard breakdown behaviors. But the hexagonal device has a soft behavior near the breakdown voltage. Its leakage current is larger, and its BV is smaller, than those of the other two devices. This is caused by the sharp corner of the P-well region. The breakdown voltage of the dumbbell-sharped device is 15 V lower than the linear device, which is influenced by the corner of zigzag P-well region under the gate polysilicon. But since its JFET width are equal everywhere, the electric field crowding effect at the P-well corner is not serious. So, its BV is slightly affected.

Fig. 11 shows the measured output characteristic curves of the three devices, with the gate bias stepping from 4 to 20 V. Some curves at selected gate bias voltages are compared in Fig 12. The on-state currents of both the dumbbell-sharped device and the hexagonal device are much larger than those of the linear device. This proves the effect of the shrunken cell pitch. But the difference of the dumbbell-sharped device and the hexagonal device varies with the gate bias voltage. Their curves almost coincide at V_{GS} = 12 V and 14 V. When V_{GS} > 14 V, the hexagonal device has larger on-state currents than the dumbbell-sharped device. This is in line with the channel density. However, when V_{GS} < 12 V, the dumbbell-sharped device even has larger on-state currents than those of the hexagonal device. Fig 13 compares the extracted $R_{ON, SP}$ at V_{GS} = 12 V and I_D = 1 A. $R_{ON, SP}$ of the linear, hexagonal and dumbbell-sharped devices are respectively 3.13 mΩ·cm^2, 2.47 mΩ·cm^2 and 2.60 mΩ·cm^2.

The above parameters are summarized in Table 1. Although both BV and $R_{ON, SP}$ of the dumbbell-sharped device are between these of the linear device and the dumbbell-sharped device, its figure of merit is the highest among the three devices.

Without doubt, the cell-pitch of the SiC planar MOSFET will be shrunken continually in the future. Then, the equivalent cell-pitch of the proposed structure can be shrunken accordingly. As a simple estimation, the equivalent $W_{cell,eq}$ can be calculated as $W_{cell,eq} = (W_{cell} + W_{cell}') / 2 = W_{cell} - (2L_{GC} + L_{CT}) / 2$, where $2L_{GC} + L_{CT}$ is the gate open. For the process platform used in this paper, W_{cell} = 4.8 μm and W_{cell}' = 2.8 μm, then $W_{cell,eq}$ = 3.8 μm. Suppose that W_{cell} is reduced to 3.0 μm, and $2L_{GC} + L_{CT}$ is reduced to 1.2 μm, then W_{cell}' = 1.8 μm and $W_{cell,eq}$ = 2.4 μm.

IV. Conclusion

It is shown that the equivalent cell-pitch, as well as the specific on-resistance, of the 4H-SiC planar MOSFET can be reduced by removing the P-plus region and contact metal of some cells without violating the design rules. The narrowest cell-pitch of 2.8 μm is realized on a 4.8 μm process platform. Based on the new design concept, the linear cell layout is turned into a dumbbell-sharped cell layout. Its $R_{ON, SP}$ is reduced from 3.13 mΩ·cm^2 to 2.60 mΩ·cm^2, which is close to the hexagonal device. Meanwhile, its blocking capability is only slight reduced by 15 V, resulting in the highest figure of merit among the three devices. The equivalent cell-pitch can be reduced further once the normal cell-pitch shrinks.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 62074020; and in part by the Natural Science Foundation of Chongqing, China, under Grant CSTB2022NSCQ-MSX1532.

References

[1] T. Suematsu, T. Suto, Y. Mori, H. Shimizu, A. Shima, and Y. Tanaka, "Development of Vertical-Channel Fin-SiC MOSFET for 3.3 kV Applications," *Proc. of ISPSD2024*, pp. 1-4, 2024.

[2] S. Asaba, M. Furukawa, Y. Kusumoto, R. Iijima, and H. Kono, "Design guidelines for SBD integration into SiC-MOSFET breaking RonA-diode conduction capability trade-off," *Proc. of IEDM2022*, pp. 198-201, 2022.

[3] D. Kim, S. DeBoer, S. Y. Jang, A. J. Morgan, and W. Sung, "Improved Blocking and Switching Characteristics of Split-Gate 1.2kV 4H-SiC MOSFET with a Deep P-well," *Proc. of ISPSD2023*, pp. 350-353, 2023.

[4] B. J. Baliga, "Silicon Carbide Power Devices: Progress and Future Outlook," *IEEE J. Emerg. Sel. Topics Power Electron.*, 11(3), pp. 2400-2411, 2023.

[5] H. Yu, J. Wang, J. Zhang, S. Liang, and Z. J. Shen, "Theoretical Analysis and Experimental Characterization of 1.2-kV 4H-SiC Planar Split-Gate MOSFET With Source Field Plate," *IEEE Trans. Electron Devices*, 71(3), pp. 1508-1512, 2024.

[6] T. Kimoto, M. Kaneko, K. Tachiki, K. Ito, R. Ishikawa, X. Chi, D. Stefanakis, T. Kobayashi, and H. Tanaka, "Physics and Innovative Technologies in SiC Power Devices," *Proc. of IEDM2021*, pp. 761-764, 2021.

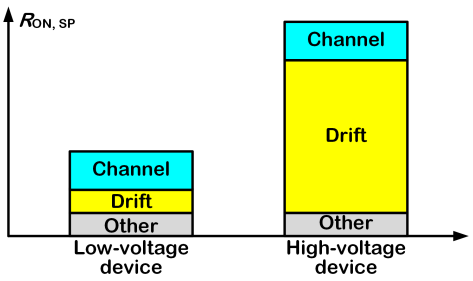


Fig. 1. Schematic specific on-resistance components within SiC planar MOSFETs.

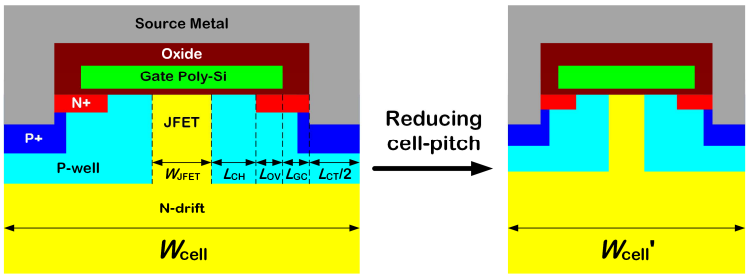


Fig. 2. Schematic diagram of the cell shrinking.

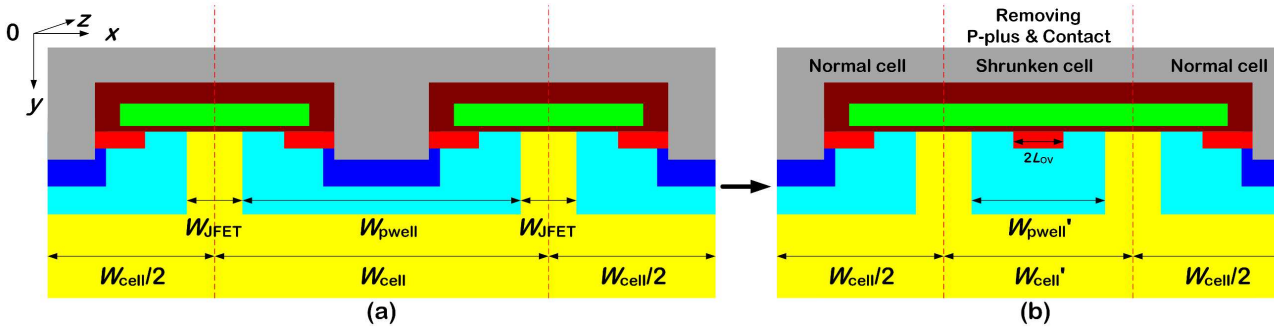


Fig. 3. Schematic cross-sections of (a) the conventional cells and (b) the proposed structure with the middle cell shrunken.

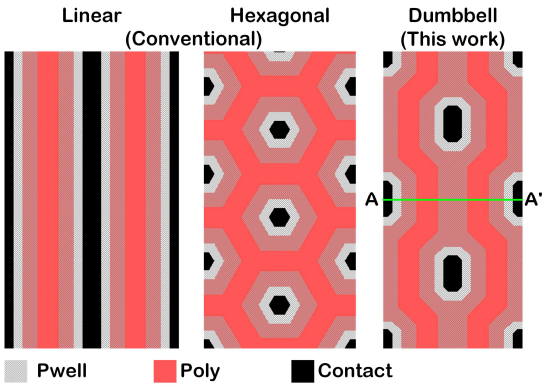


Fig. 4. Layout of the linear, hexagonal and dumbbell-shaped cells.

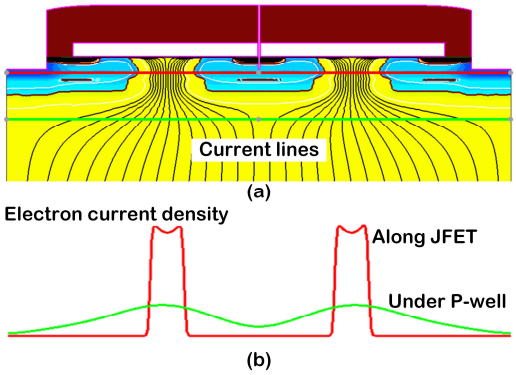


Fig. 5. Simulated (a) on-state current lines along AA' in Fig. 4 and (b) distribution of electron current density.

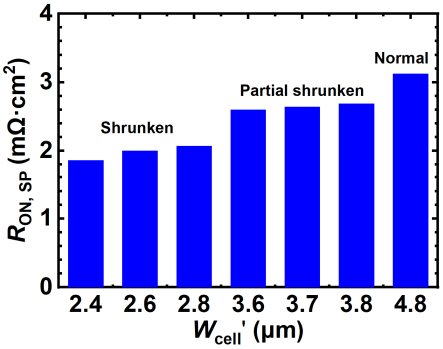


Fig. 6. Simulated specific on-resistance of various cell-pitches.

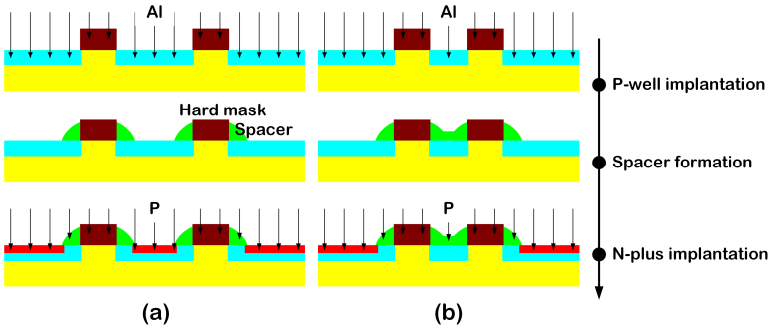


Fig. 7. Schematic process flow of the self-aligned technology: (a) good design and (b) bad design. N-plus implantation will be blocked by adjacent merged spacer if the hard masks are too close.

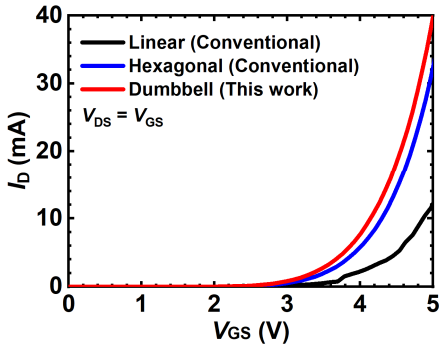


Fig. 8. Measured transfer characteristics of three devices.

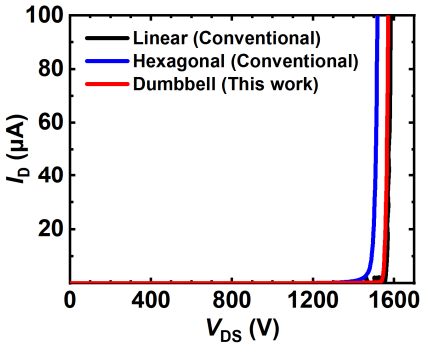


Fig.9. Measured blocking characteristics of three devices.

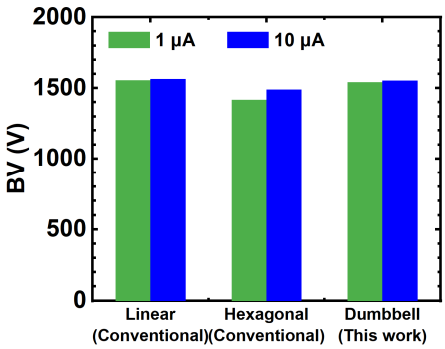
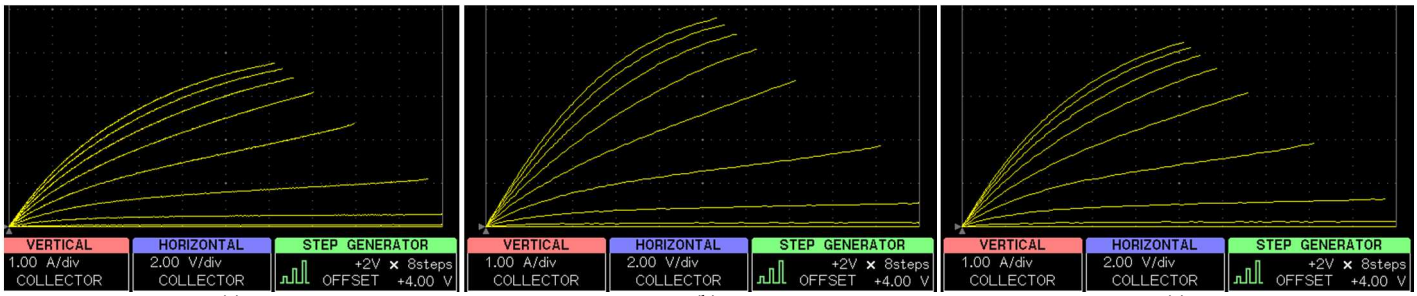


Fig. 10. Comparison of extracted breakdown voltages at I_D = 1 μA and 10 μA.



(a) (b) (c)

Fig. 11. Measured output characteristic curves of (a) Linear, (b) hexagonal, and (c) dumbbell-sharped devices. Gate bias voltages are step from 4 V to 20 V.

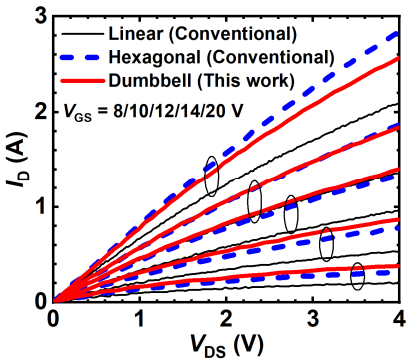


Fig. 12. Comparison of output characteristic curves at selected gate bias voltages.

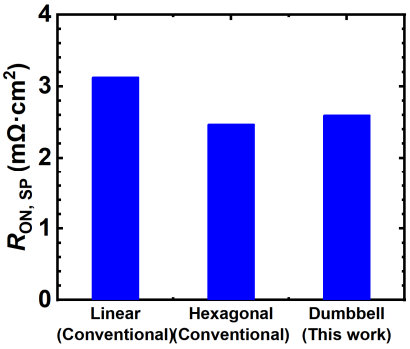


Fig. 13. Comparison of extracted specific on-resistance at V_{GS} = 20 V and I_D = 1 A.

Device	Linear (Conventional)	Hexagonal	Dumbbell (This work)
V_{th} (V)	2.7	2.5	2.4
BV (V)	1555	1417	1540
$R_{ON, SP}$ (mΩ·cm²)	3.13	2.47	2.60
FOM $BV^2/R_{ON, SP}$ (MW/cm²)	773	813	912

Table 1. Summary of electrical parameters.

Multiscale Modeling and Calibration of Hot-Carrier Stress Degradation in SiGe DRAM Peripheral MOSFETs

L. Silvestri[1], S. Jin[2], W. Jeong[2], H. Shim[2], W. Yim[2], H. Kwon[2], D. Oh[2], S. Son[2], N. Zographos[1], N. Kim[3], L. Sponton[1], Y. Lee[1]
[1]Synopsys Switzerland LLC, Zurich, Switzerland, email: luca.silvestri@synopsys.com
[2]SK hynix Inc, Icheon, Republic of Korea, [3]Synopsys Korea Inc, Gyeonggi-do, Republic of Korea

***Abstract*—**A simulation flow for hot-carrier stress (HCS) degradation that combines *ab-initio*, Spherical Harmonic Expansion of Boltzmann transport equation and TCAD simulations is developed and applied to DRAM peripheral MOSFETs. Moreover, automatic calibration strategies for the HCS degradation model are implemented and validated using a large set of experimental data from Si and SiGe n- and p-MOSFETs, thus giving insights into the physical mechanisms responsible for HCS degradation and showing its efficiency to rapidly design reliable DRAM peripheral transistors.

I. Introduction

A recent boost to low-power DRAM performance has been achieved by introducing the high-k metal gate (HKMG) technology to scale down the effective oxide thickness. HKMG pMOS adopt SiGe channel (cSiGe), because it allows threshold voltage modulation [1]. However, cSiGe is also known, despite of its various advantages, to worsen hot-carrier stress (HCS) degradation effects, due to higher hole mobility compared to Si channel [2]. Therefore, to continue the DRAM miniaturization trend, HCS degradation modeling is a major concern in designing memory peripheral devices, such as decoders, sense amplifiers and muxes.

Here we propose a simulation flow for HCS degradation in Si and cSiGe peripheral MOSFETs. It combines *ab-initio* atomistic simulations by QuantumATK [3] with the Spherical Harmonic Expansion of Boltzmann transport equation (SHE-BTE) and the HCS degradation simulations with the technology computer-aided design (TCAD) simulator Sentaurus Device by Synopsys [4]. The HCS degradation model is a physics-based analytical model that describes the microscopic interactions between hot carriers and the interface molecules via single-particle (SP) or multi-particle (MP) processes, as well as the field-enhanced thermal (TH) interaction with the lattice [5]. However, even though the HCS model was proven to accurately predict hot-carrier effects for several technologies [6], its parameters need to be calibrated first. The calibration procedure can be complex and time consuming. To bypass such complexity and efficiently generate a parameter set within physics-validated range, while retaining *ab-initio* based parameters, we developed two different automatic calibration strategies using Sentaurus Calibration Workbench (SCW) [7]. We emphasize that those strategies only require parameter search and gradient-based optimization steps, that avoid generating a large number of training simulation data, so that they can be rapidly applied to various DRAM technology nodes.

II. HCS Calibration Flow

A. Ab-initio bandstructure + SHE-BTE simulations

The simulation workflow is schematized in Fig.1. As a first step, a Si primitive unit cell or SiGe random alloy supercells are created. To better describe the random alloy characteristics, an atomistic alloy structure having total energy minimum is chosen for each Ge mole fraction (from 10 to 30 %) among 30 samples generated by the special quasi-random structure (SQS) method [8]. As an exchange-correlation method, HSE06 hybrid functional is employed, since SQS-HSE06 better depicts high-energy band tails important for the HCS modeling than the empirical-pseudopotential method (EPM) based on virtual crystal approximation (VCA). Band structures are then calculated based on SHE-BTE tensor meshes for bulk or strained irreducible wedges (IW) spanning the full Brillouin zone according to each symmetry operation to obtain the density of states $g(\mathrm{E})$ and group velocity $v(\mathrm{E})$.

The carrier energy distribution function $f(\vec{r},\mathrm{E})$ is obtained by solving the first-order SHE-BTE, which is used to calculate the scattering-rate integral of the SP and MP processes of the HCS degradation model with *ab-initio* $g(\mathrm{E})$ and $v(\mathrm{E})$. The SHE-BTE parameters for low-field scattering rate are calibrated to literature data [9], while default high-field impact ionization (ii) parameters [10] are modified to reproduce Si ii coefficients of van Overstraeten [11] and the trend of SiGe ii coefficients with and without alloy scattering [12]. In Table 1, the extracted SHE parameters are reported.

B. SCW Calibration Strategies

The SCW calibration strategies for ON-state HCS degradation of Si n-MOSFETs and OFF-state HCS degradation of Si and cSiGe p-MOSFETs are reported in Fig.1, as the most critical reliability cases. It is known that ON-state stress generates acceptor traps (negative charges when occupied), while OFF-state stress generates both acceptor and donor traps (positive charges when occupied) [13]. For nMOS acceptor traps monotonically degrade both I_{ON} and I_{OFF} with stress time. For pMOS instead, the effect of both acceptor and donor traps combines such that I_{OFF} increases, while I_{ON} first increases for short stress time and then decreases for longer stress time. Under the assumption that hot electrons generate acceptor traps while hot holes

donor traps [14], the HCS parameters for electrons only are calibrated for ON-state degradation of nMOS, while both electrons and holes parameters are calibrated for pMOS. Moreover, the trap energy distributions and the interface trap mobility model, that accounts for the Coulomb interaction of interface charges with carriers in the channel, are also calibrated.

The strategies follow the logic of calibration work performed by expert engineers and consist of ad-hoc sequencies of evaluation steps (to set specific parameter values) and search steps (to span parameter spaces), that aim at finding a good starting point for the subsequent gradient-based optimization steps. The nMOS strategy starts with a sequence of search steps for the acceptor trap energy distribution and the Coulomb mobility parameters, that aim at reproducing the subthreshold and on-state shape of the Id-Vg curves at large stress time, respectively. To this purpose one HCS component is activated, TH in this case, and *nu_th* is set to a large value to artificially generate *N0* interface traps all along the interface. For the second step all the HCS components are activated and specific parameters for MP, TH and SP are scanned. The final optimization step includes the main fitting parameters of the electron HCS, Coulomb mobility and trap energy model.

The pMOS strategy is separated in two parts. The first one focuses on the acceptor (steps 1-4), while the second one on the donor traps (steps 5-7). During the first part, the HCS components for holes are switched off and the electron components only are activated, and vice versa for the second part. The first step consists of searches for the acceptor trap energy distribution, that aim at reproducing the I_{ON} increase at short stress time. The SP parameters are then searched, after including I_{OFF} targets. A first optimization step focuses on I_{ON} increase, while the second one finalizes the acceptor trap parameters calibration to reproduce I_{ON} increase at short stress time and I_{OFF} increase. In the second part, searches are first run followed by an optimization step for the hole MP parameters. A final optimization step with all the relevant parameters for donor traps is run to reproduce the I_{ON} decrease at large stress time.

III. Results

A large set of degradation measurements (Id-Vg curves in linear and saturation regimes at different stress time, stress biases and temperature) of Si MOSFETs with asymmetric junctions and cSiGe pMOSFETs are used to validate the calibration strategies. Some results are illustrated in Fig. 2-5, showing an overall good agreement between simulations and measurements. The pMOS simulations show that the acceptor traps are generated faster and contribute to both I_{OFF} and I_{ON} increase. For longer stress time, donor traps are also generated, that partially compensate the acceptor traps and contribute to mobility degradation, thus decreasing I_{ON}. The detrimental effect on low-field mobility is confirmed by the fact that I_{ON} degradation is significantly more pronounced in linear regime than in saturation (Fig. 5). The generated interface charge profiles along the channel are shown in Fig. 6 for nMOS. As expected, the MP component is the main mechanism for the trap formation of ON-state degradation. Electrons accelerated towards the drain gain energy. The profile shows a peak close to the drain junction, also due to SP events. The electric field due to large Vgs contributes to generate a few traps at the source side via TH mechanism. In Fig. 7, the profiles for pMOS are reported, showing the SP and TH processes being dominant for acceptor traps formation. The SiGe electron $f(\vec{r}, \mathrm{E})$ shows high energy tail in the middle of the channel, which increases electron SP processes, partially responsible of the larger I_{OFF} variation compared to Si (Fig.4). In OFF-state, the TH mechanism dominates due to large vertical field. The MP and TH, for long stress time (not shown here), are the main degradation contributions for the donor traps. Leakage currents can still generate hot carriers [15]. In Table 2, the extracted model parameters for SiGe pMOS are reported.

The SCW calibration strategies are also tested with experimental data from different devices. In Fig. 8 the results obtained by applying the strategy to HCS in nMOSFETs from 40 nm embedded 1T NOR memory technology from [16] are reported, showing a good agreement with the experiments.

IV. CTGS Design

As an application example of the calibrated model, TCAD simulations are performed to understand the HCS degradation trend obtained experimentally by varying the contact-to-gate space (CTGS) in Si peripheral pMOS (Fig.9). While I_{OFF} variation is only weakly affected (Fig.10), a reduction of CTGS below a certain threshold causes a significantly more pronounced detrimental effect on I_{ON} by HCS (Fig.11). The shorter CTGS results in a slightly steeper junction and shorter channel that cause larger electric field and hole energy, respectively (Fig.9).

V. Conclusions

For SiGe DRAM technology development, attention must be paid to HCS degradation. The ab-initio to TCAD simulation flow with automatic parameter calibration presented here demonstrates to be an essential tool for accurate physical modeling of HCS degradation and SiGe peripheral transistor design.

References

[1] M. Sung et al., IEDM 2015, p.680.
[2] J. Franco et al., in: “Hot Carrier Degradation in Semiconductor Devices”, Switzerland: Springer, p. 259, 2015.
[3] *QuantumATK,* Version V-2024.03, Synopsys QuantumATK, 2024.
[4] *Sentaurus™ Device User Guide*, Version V-2024.03, Synopsys Inc., 2024.
[5] S. Reggiani et al., IEEE TED 60, p. 691, 2013.
[6] P. Pfäffli et al., Microel. Rel. Vol. 88–90, p. 1083, 2018.
[7] *Sentaurus™ Calibration Workbench User Guide*, Version V-2024.03, Synopsys Inc., 2024.
[8] S. Smidstrup et al., J. Phys. Condes. Matter 2020, 32, 015901.
[9] M. V. Fischetti et al., JAP 80, p. 2234, 1996.
[10] A. G. Chynoweth, Phys. Review, Vol. 109, 5, p. 1537, 1958.
[11] R. Van Overstraeten et al., Solid-State Electronics, 13, p. 583, 1970.
[12] T. V. Dinh et al., ULIS 2009, p. 77.
[13] N.-H. Lee et al., IRPS 2019, p. 1.
[14] P. E. Nicollian et al., IRPS 2007, p. 197.
[15] D. Varghese et al., IEEE TED 54, p. 2669, 2007.
[16] G. Torrente et al., IIRW 2015, p. 134.

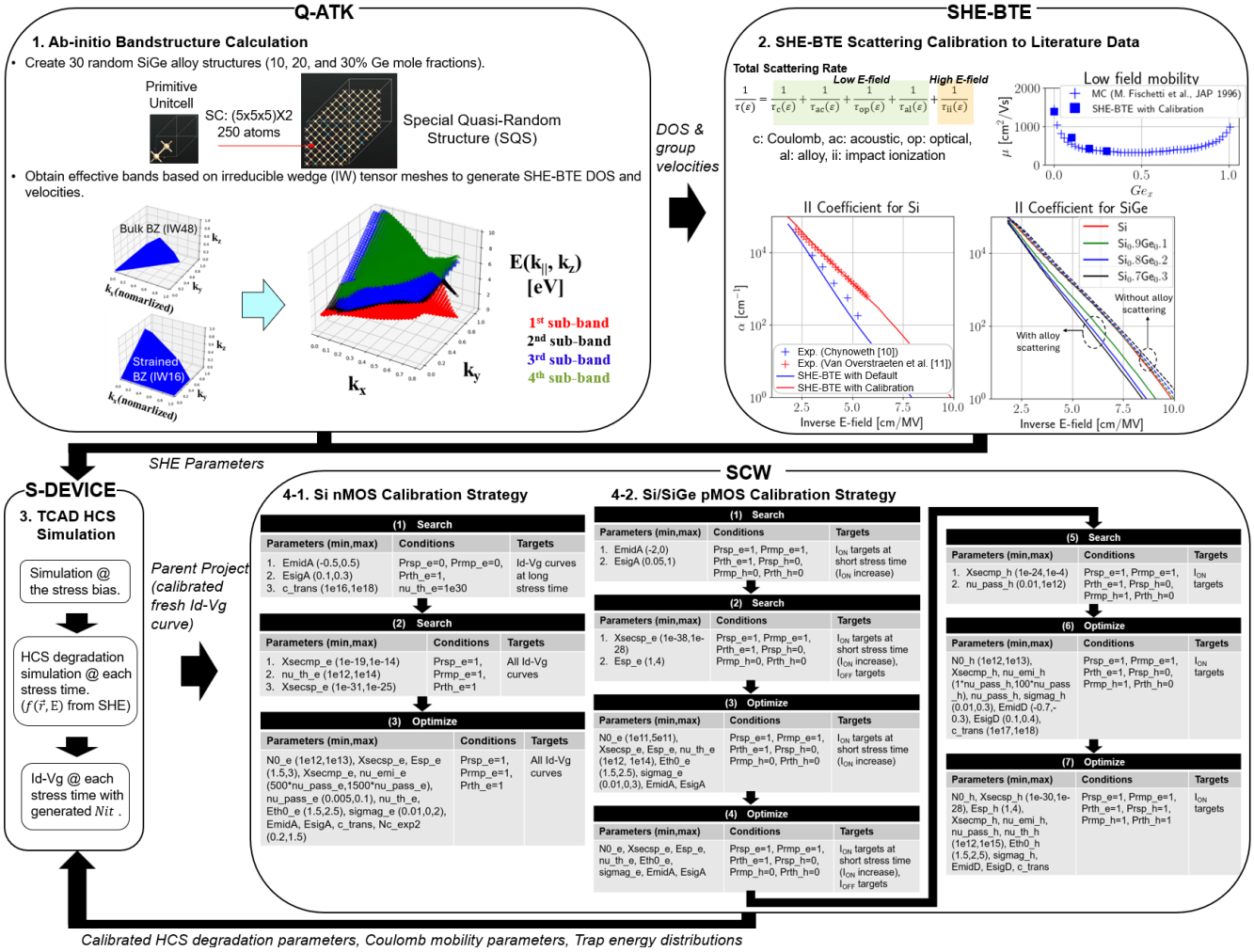


Fig. 1. Schematic of the simulation/calibration flow used for the study of HCD degradation in Si and SiGe DRAM peripheral n- and p-MOS. *Ab-initio* (1) and SHE-BTE simulations (2) provide TCAD with SHE material parameters. The TCAD simulation setup for HCS degradation simulation (3), calibrated to reproduce the fresh Id-Vg curves, represents the parent project for SCW calibration flows for n- (4-1) and p-MOS (4-2). The HCS model parameters *Prsp*, *Prmp* and *Prth* are used as switches for the corresponding components of the model, SP, MP and TH, respectively (the suffix "e" refers to electrons, "h" to holes). The parameter ranges are also reported. For pMOS two targets are defined for each Id-Vg experimental curve: one considering a Vg range in subthreshold regime (I_{OFF} target) and one in ON-state regime (I_{ON} target). The run time is about 6 hours and 1 day for the nMOS and pMOS calibration flows, respectively.

SHE Params [unit]	Si	Ge
ε_{op} [meV]	62	53.3
s_{ii1} [s^{-1}]	5.23e11	5.23e11
s_{ii2} [s^{-1}]	4.83e11	4.83e11
s_{ii3} [s^{-1}]	5.14e11	5.14e11
ε_{ii1} [eV]	1.128	0.9205
ε_{ii2} [eV]	1.572	1.572
ε_{ii3} [eV]	1.75	1.75
v_{ii1}	2.106	2.106
v_{ii2}	0.517	0.517
v_{ii3}	0.793	0.793
ΔU [V]	-	1.2

Table 1. Calibrated SHE parameters for optical phonon scattering (ε_{op}), alloy scattering (ΔU), and impact ionization scattering (all the remaining) for Si and Ge. The parameters are linearly interpolated in S-Device for SiGe with different Ge mole fraction.

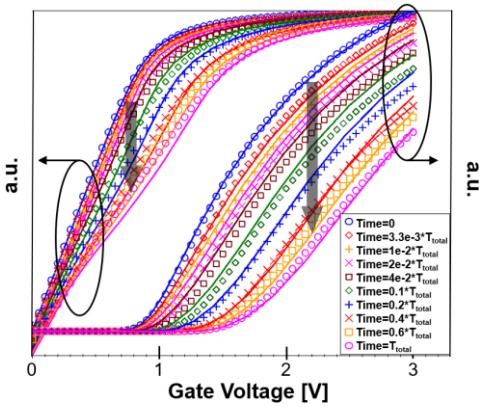


Fig. 2. Si nMOS Id-Vg characteristics at different stress time in linear and logarithmic scale. Symbols: measurements, lines: TCAD simulations. Measurements are performed at Vd=0.1 V and T=298 K, after stressing the device at Vg_{str}=1.75 V and Vd_{str}=2.75 V (Vs_{str}=Vb_{str}=0 V).

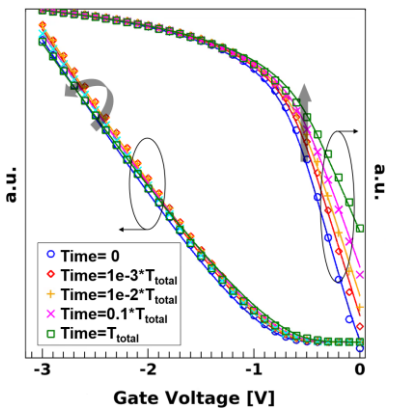


Fig. 3. SiGe pMOS Id-Vg characteristics at different stress time in linear and logarithmic scale. Symbols: measurements, lines: TCAD simulations. Measurements are performed at Vd=-3 V and T=398 K, after stressing the device at Vg_{str}=4.3 V, Vb_{str}=4.3 V and Vd_{str}= Vs_{str}=0 V.

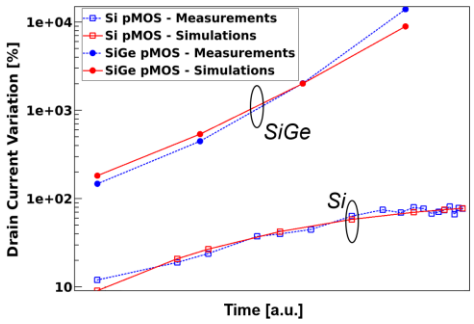


Fig. 4. Si and SiGe pMOS drain current variation at Vg=0 V as a function of stress time. Measurements are performed at Vd=-3 V and T=398 K. Stress conditions for SiGe pMOS are reported in Fig.3. For Si pMOS, Vd_{str}=-4.1 V, Vg_{str}=Vb_{str}=Vs_{str}=0 V.

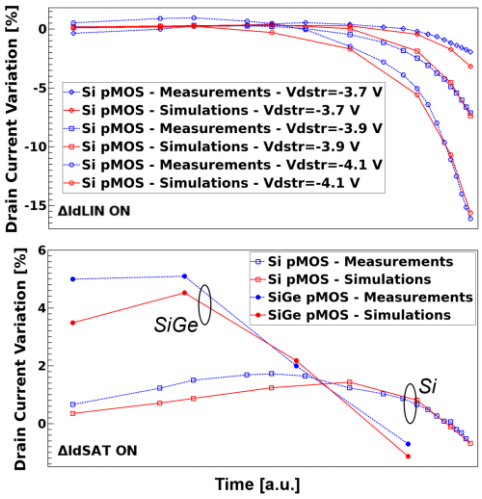


Fig. 5. Si and SiGe pMOS drain current variation at Vg=-3 V as a function of stress time and drain stress bias Vd_{str} (Vg_{str}=Vb_{str}=Vs_{str}=0 V). Measurements are performed at Vd=-0.1 V (top) and at Vd=-3 V (bottom) and T=398 K. Stress conditions for saturation current are reported in Fig. 3 and 4 for SiGe and Si pMOS, respectively.

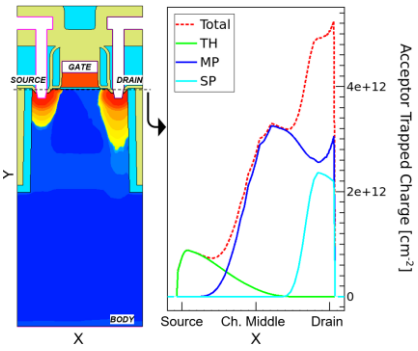


Fig. 6. Left: TCAD structure of nMOS with asymmetric junctions. The simulated interface charge profiles in the channel after T_{total} time of ON-state degradation are shown on the right for the three components of the HCS degradation model (SP, MP and TH).

HCS Parameters		
SiGe Params [Unit]	***Electron***	***Hole***
N0 [cm^{-2}]	1.936e12	4.944e11
Eth0 [eV]	1.51	2
nu_emi [s^{-1}]	default	0.988
nu_pass [s^{-1}]	default	1.103
nu_th [s^{-1}]	9.33e13	9.864e12
Esp [eV]	3.935	3.567
Xsecsp [cm^{-2}]	1.316e-33	2.928e-29
Xsecmp [cm^{-2}]	default	7.335e-5
sigmag [eV]	0.132	0.086
Coulomb Mobility Parameters		
SiGe Params [Unit]		
c_trans [cm^{-3}]	2.397e18	
c_exp [1]	0.5	
Nc_exp2 [1]	1.5	
l_crit [cm]	1e-6	
Trap Energy Distributions		
Acceptor	***Donor***	
• Uniform: EmidA= -0.204 [eV], EsigA= 0.146 [eV] • Fixed charges	Gaussian: EmidD=-0.425 [eV], EsigD= 0.249 [eV]	

Table 2. Calibrated HCS, Coulomb mobility and trap energy parameters for SiGe pMOS. Half of the acceptor traps lay inside the valence band and are considered negative fixed charges.

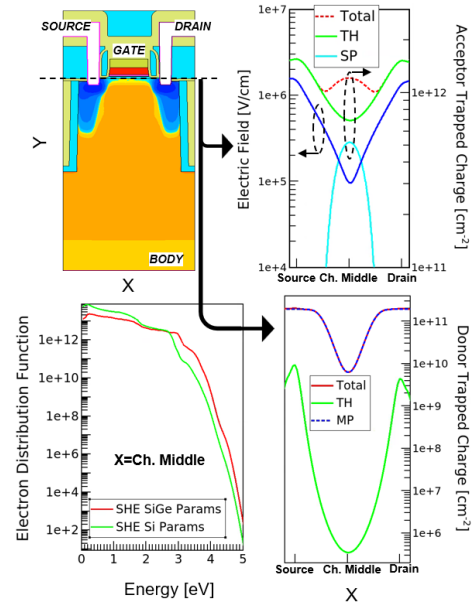


Fig. 7. Top left: TCAD structure of $cSi_{0.7}Ge_{0.3}$ pMOS. The simulated interface acceptor and donor charge profiles in the channel after T_{total} time of OFF-state degradation are shown on the right for the three components of the HCS degradation model (SP, MP and TH). Bottom left: $f(\vec{r}, E)$ in the middle of the channel obtained with default Si and $Si_{0.7}Ge_{0.3}$ SHE parameters.

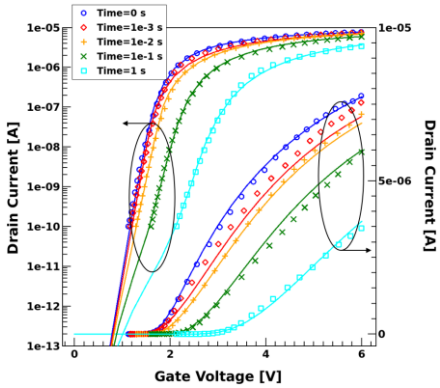


Fig. 8. nMOS Id-Vg characteristics at different stress time in linear and logarithmic scale. Symbols: measurements from [16], lines: TCAD simulations. Measurements are performed at Vd=0.05 V and T=298 K, after stressing the device at Vg_{str}=6 V and Vd_{str}=4 V.

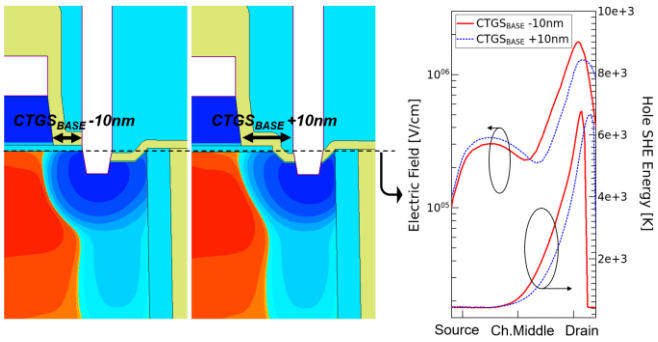


Fig. 9. Si pMOS TCAD structure with different CTGS. The electric field and the average hole energy along a cutline just below the Si/SiO_2 interface are also shown at Vg_{str}=Vb_{str}=Vs_{str}=0 V, Vd_{str}=-3.85 V and T=398 K.

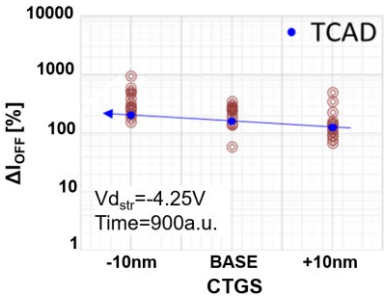


Fig. 10. Si pMOS drain current variation as a function of CTGS at Vg=0 V and Vd=-0.1 V. Red symbols are measurements at different CTGS for nominally identical devices. The devices are stressed for 900 a.u. at Vg_{str}=Vb_{str}=Vs_{str}=0 V, Vd_{str}=-4.25 V.

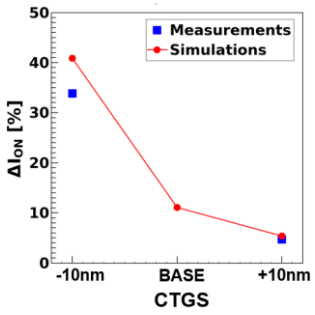


Fig. 11. Si pMOS drain current variation as a function of CTGS at Vg=0 V and Vd=-3 V, after stressing the device for 63000 a.u. at Vg_{str}=Vb_{str}=Vs_{str}=0 V, Vd_{str}=-3.85 V and T=398 K.

Innovative Cool-Stacking Technology for High Performance and Energy-Efficiency SoIC®

Terry Ku, C. H. Tsai, C. C. Hsieh, J.C. Twu, R. F. Tsui, S.W. Lu, C. S. Liu, and C. H. Douglas Yu
Taiwan Semiconductor Manufacturing Company, Ltd., Hsinchu, Taiwan, R.O.C., email: cchsiehr@tsmc.com

***Abstract*—**TSMC SoIC® has been proven as an enabling technology for high performance and energy-efficiency system integration [1]. More effective thermal management can enable broader applications of the stacking technology. An innovative bonding technology, Cool-Stacking SoIC®, is presented to greatly improve the thermal performance of a 3D stacking device. Boosted heat dissipation capability, superior lateral heat spreading for hot spot mitigation, and enhanced inter-die communication speed and energy efficiency are accomplished simultaneously by integrating a thermally conductive and mechanically friendly thin film (Cool layer or C-layer). Compared with standard SoIC®, the thermally enhanced bonding scheme effectively reduces the overall thermal resistance by 57% and improves the energy efficient performance (EEP) by up to 31%. The proposed Cool-Stacking technology unleashes system integration innovations for future AI/HPC applications adopting logic-on-logic or DRAM-on-logic stacking configuration.

I. Introduction

Heterogeneous integration of integrated circuit (IC) chips has been widely adopted in numerous high-performance computing and AI applications in which high bandwidth and energy efficient inter-die communication is of vital importance. Among potential heterogeneous integration schemes, 3D stacking of IC chips can deliver superior performance in a relatively small footprint. Nevertheless, the compact 3D stacking form factor inevitably imposes an adverse influence on thermal management, making effective heat dissipation an ever-challenging task compared with its 2D/2.5D counterparts.

3D IC architectures include logic on logic [1-5] and DRAM on logic [6-9]. Logic-on-logic architectures can be xPU die on I/O die or xPU die on xPU die while DRAM on logic be High Bandwidth Memory (HBM) or DRAM cube on xPU. Among these architectures, multiple DRAM (m-DRAM) dies on a computing die is highly desirable by emerging generative AI (G-AI) applications as the 3D configuration can offer high inter-die communication bandwidth and energy efficiency [8-9]. In such applications, the higher power GPU/AI accelerators are usually on the bottom of the stacked structure to alleviate the challenges in power integrity (PI), which causes adverse impact on heat dissipation. Hence, design trade-offs must be made between computing power (thermal design power) and memory bandwidth/capacity.

II. 3D Stacking Devices

A 3D stacking device consists of multiple active IC chips assembled layer by layer in the vertical (out-of-plane) direction. The primary heat dissipation occurs along paths in the same direction as the footprint of the stacking device and is usually much larger than its stack-up height. As indicated in Fig. 1, apart from the silicon and BEOL layers, the interconnect layers may contribute a significant portion to the overall thermal resistance of the multi-stacking die. Microbumps (Fig. 2a) have been successfully implemented to connect neighboring chips in 3D stacking applications, such as High-Bandwidth Memory (HBM) [10]. However, the polymer-filled (>70% of the stacking area), thick (>10 um) interconnect layer makes the microbump stacking scheme thermally unfavorable. 3D stacking with metallic bonding pads in thin dielectric and bonding layers (Fig. 2b & 2c) are viable alternatives from the perspective of thermal management.

A. Thermally Enhanced SoIC® Bonding Scheme

An ideal bonding scheme streamlines both signal and heat flow between dies with minimal temperature overhead while possessing a compatible process and integration flow. In this study, a process integration-friendly high κ (>50 W/mK) thin film (Cool layer or C-layer) is employed in the cool-stacking bonding scheme (Fig. 2c). Integration flow and bonding structures are delicately modified to accommodate the adoption of the C-layer without sacrificing the inter-die bondability. More importantly, energy efficient performance (EEP) of the inter-die communication is not compromised.

Dielectric thin films with high thermal conductivity (κ) are advantageous to the low thermal resistance requirement in die stacking. AlN and diamond are potential candidates with high thermal conductivity [11]. However, incorporating such materials into the existing process could be challenging as both materials are mechanically dissimilar to the materials (e.g. Si and SiO_2) commonly used in ICs. For instance, the coefficient of thermal expansion (CTE) mismatch-induced stresses could be of high concern in the bonding process. Thinning down the bonding layers is an intuitive approach to lower the out-of-plane thermal resistance. On the contrary, heat spreading in the in-plane direction of a local hot spot is more effective in a thicker film. A bonding scheme of a designated thickness is favorable to realizing a lower overall thermal resistance.

B. Integration Flow and Device Fabrication

The integration flow of Cool-Stacking SoIC® starts with a thin (submicrons to several microns) layer of dielectric film (e.g. SiO_2) formed on both top and bottom dies. The film on the active side of the top die is usually thicker than that on the back side of the bottom die to compensate any incoming surface roughness. A C-layer film of several microns thick is then

integrated onto the active side of the top die wafer. Submicron-thick bonding layers are deposited on both wafers, followed by the formation of Cu bonding pads. The structure and formation of the Cu pad and its associated processes are critical to thermal performance, energy efficient performance and bonding quality. Planarization of the bonding interfaces is carefully carried out before SoIC® bonding.

A test vehicle (TV) that mimics standard SoIC® bonding structure was fabricated as the benchmark for bonding and thermal performance comparison. A dielectric layer was replaced with C-layer to represent the thermally enhanced bonding structure. The bonding quality of both bonding schemes was evaluated using C-mode scanning acoustic microscopy (CSAM) before test samples were singulated. No obvious bonding defects were found through CSAM inspection on the bonded wafers (Fig. 4).

III. Performance Characterization

A. Thermal Performance

The effective thermal resistance (in K-mm^2/W) of the reference TV and C-layer TV were carefully characterized. Numerical models were constructed based on the empirical data to study the effectiveness of Cool-Stacking bonding in typical 3D stacking applications. The thermally enhanced bonding scheme provides a viable means to cutting down the effective thermal resistance of the die-to-die stacking structure by 57% compared with the standard SoIC® bonding scheme or by 73% compared with the microbump stacking technology (Fig. 4). The lower thermal resistance of the inter-die stacking structure manifests its system integration superiority in 3D stacking devices with higher die count, which is a pivotal feature in high-performance computing and high-capacity HBM applications.

Two potential high-performance computing applications, fine-grain memory on accelerator and typical HBM4 3D stackings, are used to illustrate the thermal advantage in terms of lower junction temperature elevation and heat spreading. Some general assumptions in the simulation include CoWoS®-S as the carrier and liquid-cooled cold plate as the system cooling solution. The fine-grain memory on accelerator structure has a GPU stacked underneath 8 layers (8-hi) of fine-grain DRAMs. The GPU has a uniform power density of 0.5 W/mm^2, and the DRAMs with a hot-spot power density of 0.21 W/mm^2. Table I indicates that the C-layer SoIC bonding scheme is the only stacking scheme that can meet the junction temperature (T_j) limit of 100 ºC for GPU and 95 ºC for DRAMs. The C-layer scheme significantly enhances the heat spreading by comparing the temperature contours of GPU die (Fig. 5) and the very top DRAM die (Fig. 6).

The typical HBM4 structure has a non-uniform power base die stacked underneath 8-hi, 12-hi and 16-hi DRAM stacks [13]. Every DRAM die is assumed to have a uniform power of 0.01W/mm^2. A sensitivity study has been carried out to evaluate the hot-spot power density impact of base die of which 0.5 W/mm^2, 0.6 W/mm^2, 0.7 W/mm^2, and 0.8 W/mm^2 are applied on certain hot spots for different HBM stacking configurations. The base die T_j comparison for 8-hi, 12-hi and 16-hi are shown in Fig. 7, 8 and 9, respectively. The C-layer SoIC® bonding significantly outperforms the other schemes with higher maximum allowable hot-spot power densities on the base die (Table II). Not only does it reduce the junction temperature elevation of the base die, but also progressively enhances the temperature reduction as the die count of the 3D stacking increases. Temperature contour of three interconnect technologies for a typical 16-hi HBM4 3D stacking is shown in Fig. 10 for the bottom base die and Fig. 11 for the top DRAM die. The proposed bonding scheme also generates a more uniform temperature contour on both the bottom base die and the top DRAM die, implying a more effective lateral heat spreading along the major heat dissipation path.

B. Electricl Performance and Energy efficient

Fig. 12 shows the schematic network of parasitic resistance and capacitance in a typical F2B die-to-die interconnect structure. The smaller the R_{bond} and R_{TSV} are, the better the electrical performance. Parasitic capacitances, C_{bond} and C_{TSV} are dominated by the materials between the metal structures, such as bonding pads, TSVs, and the silicon substrate.

Energy efficient performance, a performance indicator in interconnect speed and energy efficiency, is numerically characterized for the baseline SoIC® bonding scheme and two thermally enhanced SoIC® bonding schemes with comparable thermal performance. Detailed die-to-die interconnect schemes, including base die TSVs, layer-by-layer bonding structure, and top metal layers of the top die, are included in the analysis. EEP-associated electrical performance is documented in Table III for microbump and three SoIC® bonding schemes. Innovative C-layer bonding structure optimize not only R_{bond} and R_{TSV}, but also C_{bond} and C_{TSV} simultaneously, leading to a higher interconnect speed and energy efficiency.

The proposed bonding schemes deliver a better PDN impedance and IR drop by reducing the parasitic resistance by about 23% compared with the baseline SoIC® scheme. Moreover, components in the bonding schemes, such as dielectric/bonding layer thickness and bonding pad size, are deliberately designed to ensure the energy efficiency in die-to-die communication is not compromised. On the other hand, the interconnect speed (1/RC) of the proposed bonding schemes can be enhanced by about 30%. The resulting EEP, defined as the product of interconnect bandwidth density and energy efficiency, is improved by up to 31%, meaning the inter-die signal transmission consumes less power when thermally enhanced SoIC® bonding scheme is employed.

IV. Conclusions

An innovative cool-stacking SoIC® bonding scheme has been proposed to improve the thermal performance of the stacking structures. Compared with the standard SoIC®, the proposed bonding scheme reduces the effective thermal resistance of the bonding structure by 57%. Up to 31% improvement in EEP can be achieved by bonding scheme optimization while maintaining comparable thermal performance. The study successfully demonstrates a promising 3D bonding scheme enabling high-power thermal management and energy-efficient signal transmission for high-performance computing and AI applications.

ACKNOWLEDGEMENT

The authors are grateful to tsmc collaborators from Pathfinding for System Integration and Integrated Interconnect and Packaging for their help and support.

REFERENCES

[1] M. F. Chen et al., IEEE ECTC, p. 594, 2019. [2] S. W. Liang et al., IEEE ECTC, p. 1090, 2022. [3] R. Mathur et al., IEEE ECTC, p. 541-547, 2020. [4] S. Sinha et al., IEDM, p. 15.1.1, 2020. [5] Das Sharma et al., Nat Electron 7, p. 244, 2024. [6] M. F. Chen et al., " IEEE Trans. on Electron Devices, 67, p. 5343, 2020. [7] D. C. H. Yu et al., IEEE Trans. on Electron Devices, 69, p. 7167, 2022. [8] B. Dally, Insight from NVIDIA Research, 2022. [9] P. M. Kogge et al., IEEE/ACM PMBS, p. 26, 2022. [10] S. Y. Hou et al., IEEE Trans. on Electron Devices, 64, p. 4071, 2017. [11] W. Y. Woon et al., IEDM, p. 1, 2023. [13] K. Kim. VLSI SC-C-6, 2024

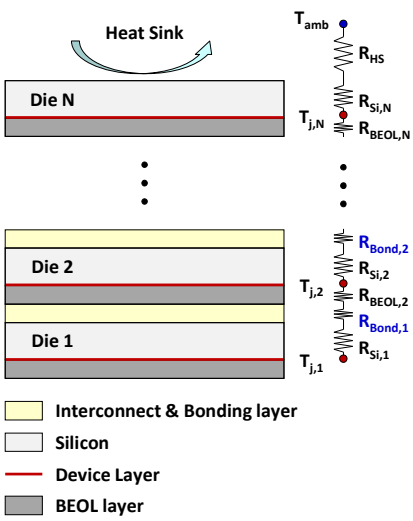


Fig. 1. Thermal resistance network of a typical F2B 3D stacking device.

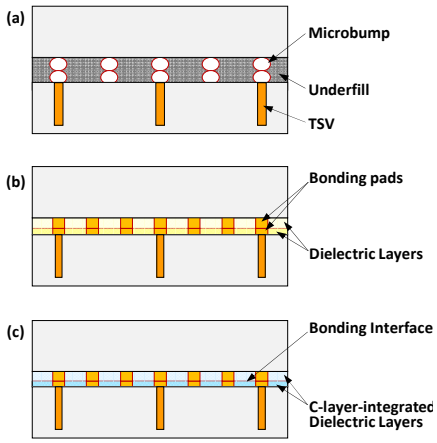


Fig. 2. Schematics of three die-to-die stacking technologies for 3D IC integration. (a) microbump F2B stacking scheme, (b) standard SoIC® F2B bonding scheme, and (c) C-layer SoIC® F2B bonding scheme.

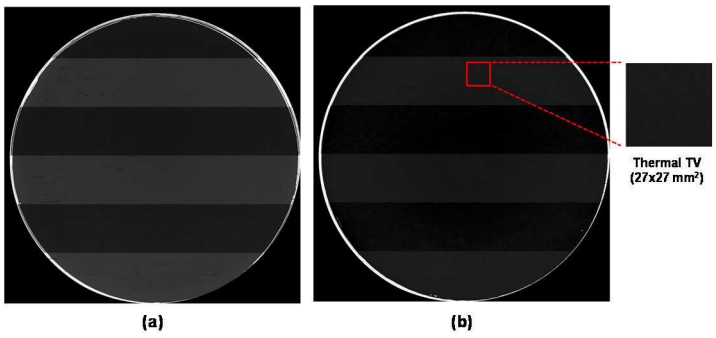


Fig. 3. CASM photo of bonded wafers with different bonding schemes. (a) standard SoIC®-mimic bonding scheme, and (b) C-layer-inserted bonding scheme.

F2B Stacking Technology	MicroBump	Standard SoIC®	C-layer SoIC®
T_j of GPU	123.4 ºC	108.9 ºC	94.4 ºC
T_j of bottom DRAM	119.1 ºC	105.9 ºC	93.2 ºC

Table I. Junction temperature elevation comparison of three F2B die-to-die stacking technologies to build an 8-hi fine-grain structure. Only C-layer survives 8-hi fine grain by keep the Tj of the very bottom DRAM under 95oC as shown in comparison of three F2B die-to-die stacking technologies to build an 8-hi fine-grain structure.

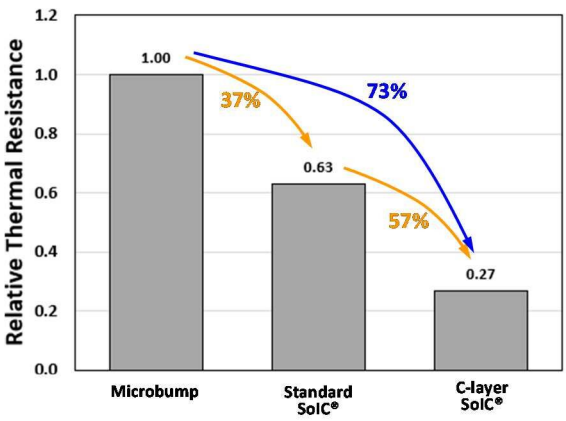


Fig. 4. Comparison of effective thermal resistance of the interface bonding structure in three F2B die-to-die stacking technologies.

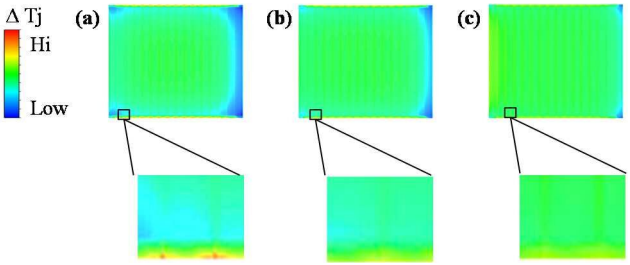


Fig. 6. Temperature contour plots of the top DRAM die of 8-hi fine grain stacking configuration: (a) microbump F2B bonding scheme, (b) standard SoIC® F2B bonding scheme, and (c) C-layer SoIC® F2B bonding scheme.

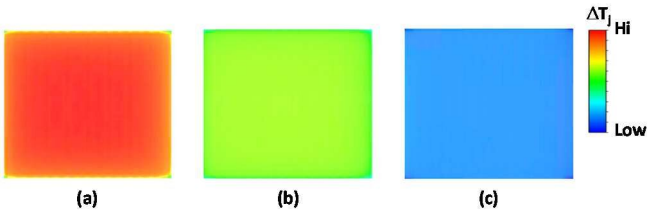


Fig. 5. Temperature contour plots of the bottom GPU die of 8-hi fine grain stacking configuration: (a) microbump F2B bonding scheme, (b) standard SoIC® F2B bonding scheme, and (c) C-layer SoIC® F2B bonding scheme.

F2B Stacking Technology	MicroBump	Standard SoIC®	C-layer SoIC®
8-hi HBM	0.65 W/mm^2	0.79 W/mm^2	1.28 W/mm^2
12-hi HBM	0.50 W/mm^2	0.60 W/mm^2	1.05 W/mm^2
16-hi HBM	0.40 W/mm^2	0.50 W/mm^2	0.86 W/mm^2

Table II. Maximum allowable power densities of hot spots meeting base die T_j limit of 95ºC for three die-to-die F2B stacking technologies in typical HBM4 stacking configuration.

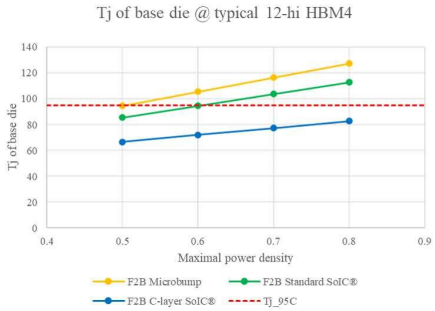


Fig. 8. The base die junction temperature of typical 12-hi HBM4 stacking configuration @ hot spot power density @ 0.5W/mm^2, 0.6W/mm^2, 0.7W/mm^2, and 0.8W/mm^2: (a) microbump F2B stacking scheme (b) standard SoIC® F2B bonding scheme, and (c) C-layer SoIC® F2B bonding scheme.

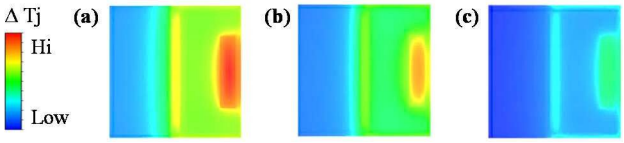


Fig. 10. Temperature contour plots of the base die (0.7 W/mm^2) in a 16-hi HBM4-like stacking configuration: (a) microbump F2B bonding scheme, (b) standard SoIC® F2B bonding scheme, and (c) C-layer SoIC® F2B bonding scheme.

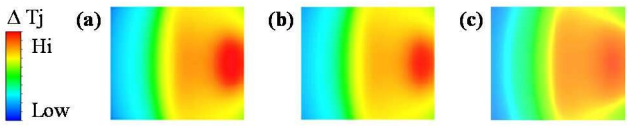


Fig. 11. Temperature contour plots of the top DRAM die in16-hi HBM4-like stacking configuration: (a) microbump F2B stacking scheme, (b) standard SoIC® F2B bonding scheme, and (c) C-layer SoIC® F2B bonding scheme.

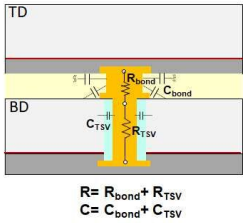


Fig 12. Schematic network of parasitic resistance and capacitance in a typical F2B die-to-die interconnect.

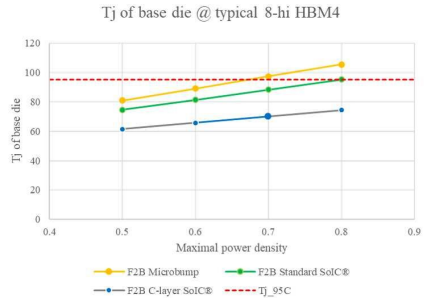


Fig. 7. The base die junction temperature of typical 8-hi HBM4 stacking configuration @ hot spot power density @ 0.5W/mm^2, 0.6W/mm^2, 0.7W/mm^2, and 0.8W/mm^2: (a) microbump F2B stacking scheme (b) standard SoIC® F2B bonding scheme, and (c) C-layer SoIC® F2B bonding scheme.

Fig. 9. The base die junction temperature of typical 16-hi HBM4 stacking configuration @ hot spot power density @ 0.5W/mm^2, 0.6W/mm^2, 0.7W/mm^2, and 0.8W/mm^2: (a) microbump F2B stacking scheme (b) standard SoIC® F2B bonding scheme, and (c) C-layer SoIC® F2B bonding scheme.

	MicroBump	STD SoIC®	C-Layer SoIC®-1	C-Layer SoIC®-2
Parasitic Resistance, R	0.47X	1.0X	0.77X	0.77X
Parasitic Capacitance, C	6.44X	1.0X	1.01X	1.0X
Speed, 1/(RC)	0.33X	1.0X	1.29X	1.31X
Energy Efficiency (EE)	0.16X	1.0X	0.99X	1.0X

Table III. Comparison of electrical performance and energy efficiency of different SoIC® bonding technologies.

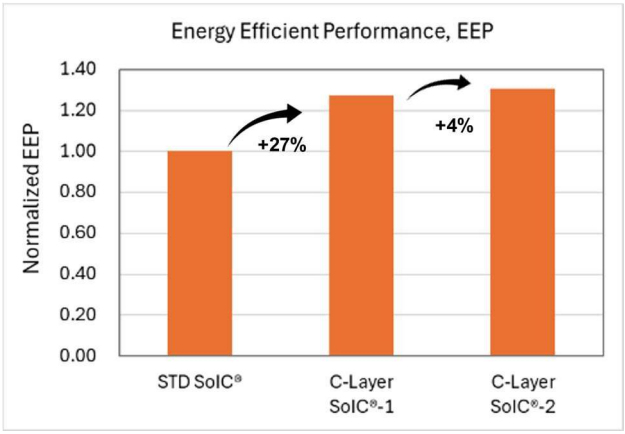


Fig. 13. EEP of die-to-die communication for three SoIC® bonding schemes. An improvement of 27% is achieved for C-layer SoIC®-1 and further optimized to 31% for C-layer SoIC®-2 compared with standard SoIC®.

2nm Platform Technology featuring Energy-efficient Nanosheet Transistors and Interconnects co-optimized with 3DIC for AI, HPC and Mobile SoC Applications

Geoffrey Yeap, S.S. Lin, H.L. Shang, H.C. Lin, Y.C. Peng, M. Wang, PW Wang, CP Lin, KF Yu, WY Lee, HK Chen, DW Lin, BR Yang, CC Yeh, CT Chan, JM Kuo, C-M Liu, TH Chiu, MC Wen, T.L. Lee, CY Chang, R. Chen, P-H Huang, C.S. Hou, YK Lin, FK Yang, J. Wang, S. Fung, Ryan Chen, C.H. Lee, TL Lee, W. Chang, DY Lee, CY Ting, T. Chang, HC Huang, HJ Lin, C. Tseng, CW Chang, KB Huang, YC Lu, C-H Chen, C.O. Chui, KW Chen, MH Tsai, CC Chen, N. Wu, HT Chiang, XM Chen, SH Sun, JT Tzeng, K. Wang, YC Peng, HJ Liao, T. Chen, YK Cheng, J. Chang, K. Hsieh, A. Cheng, G. Liu, A. Chen, HT Lin, KC Chiang, CW Tsai, H. Wang, W. Sheu, J. Yeh, YM Chen, CK Lin, J. Wu, M. Cao, LS Juang, F. Lai, Y. Ku, S.M. Jang, L.C. Lu

Global R&D Center, Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu, Taiwan, R.O.C.

Abstract

A leading edge 2nm CMOS platform technology (N2) has been developed and engineered for energy-efficient compute in AI, mobile and HPC applications. This industry-leading N2 logic technology features energy-efficient gate-all-around nanosheet transistors, middle-of-line and backend-of-line interconnects with densest SRAM macro of ~38Mb/mm^2. N2 delivers a full node benefit from previous 3nm node [4] in offering 15% speed gain or 30% power reduction with >1.15x chip density increase. N2 platform technology, equipped with new Cu scalable RDL, flat passivation and TSVs, co-optimizes holistically with 3DFabric™ technology enabling system integration/scaling for AI/mobile/HPC product designs. N2 successfully met wafer-level reliability requirements and passed 1000hrs HTOL qual with high yielding 256Mb HC/HD SRAM, and logic test chip (>3B gates) consisting of CPU/GPU/ SoC blocks. Currently in risk production, N2 platform technology is scheduled for mass production in 2H'25. N2P, 5% speed enhanced version of N2 with full GDS compatibility, targets to complete qualification in 2025 and mass production in 2026.

Introduction

Advanced CMOS technology has been the key enabler for semiconductor product innovations. Since the generative AI break-through moment in Q1'23, AI together with 5G-advanced mobile and HPC have ignited the industry with an insatiable appetite for best-in-class advanced energy-efficient logic technology [1]. Our industry leading 2nm platform technology (N2) is one such advanced logic technology. This paper describes the state-of-art N2 technology successful transition into NS platform technology and acceleration of >140x energy-efficient compute from 28nm to N2 as shown in Fig. 1. We also present system technology co-optimization (STCO) innovation in design rules, standard cell, SRAM and interconnects co-optimization with 3DFabric™. N2 technology has been verified on our development/qual test vehicle. N2 met all the wafer-level reliability requirements and completed the full 1000hours HTOL qualification with high yielding 256Mb HD/HC SRAM and logic test chip (>3B gates). Now in risk production, N2 is on track for mass production in 2H'25. N2P with 5% additional speed and full GDS compatibility targets to complete qual in 2025 and mass production in 2026.

N2 NanoFlex™ [3] Technology Architecture

The N2 2nm platform technology is defined and developed to meet PPACt (Power, Performance, Area, Cost, and Time-to-market) [2]. STCO is emphasized with smart scaling features instead of brute-force design rule scaling which drastically increases process cost and inadvertently causes critical yield issues. Extensive STCO coupled with smart scaling of major design rules (e.g., gate, nanosheet, MoL, Cu RDL, passivation, TSVs) was performed in optimizing this 2nm technology to achieve target PPA. This development also involves co-optimization with 3DFabric™ SoIC 3D-stacking and advanced packaging technology (INFO/CoWoS variants) thereby accelerating system integration/scaling for AI/mobile/HPC product designs.

N2 NanoFlex™ [3] standard cell innovation offers not only nanosheet width modulation but also the much-desired design flexibility of the multi-cell architecture. N2 short cell lib for area and power efficiency. Selective use of tall cell lib lifts frequency to meet design target Combining with six-Vt offerings spanning across 200mV, N2 provides unprecedented design flexibility to satisfy a wide spectrum of energy-efficient compute applications at the best logic density. N2 delivers a full node scaling with attractive PPA values at projected cost and time-to-market: ~15% speed gain or ~30% power reduction with >1.15x chip density scaling (Fig. 2-3).

Energy-efficient Nanosheet Transistors, MoL and BEOL Interconnects

Multiple generations of Si FinFet with fin depopulation were in use from 16nm to 7nm (2-fin) node. High mobility channel transistors with industry-first zero-thickness dipole based true multi-Vt (7-Vt), cut metal-gate and gate-contact over-active innovations extended FinFet architecture into N5 node [2]. FinFlex™ DTCO coupled with other key enhancements successfully extracted another full node PPA benefits in N3, last FinFet node [4].

N2 platform technology successfully completes the transition from FinFet into energy-efficient nanosheet technology. Figure 4 shows optimized nominal gate-length NS transistors with excellent DIBL and subs-threshold swings. Long gate-length NS transistors achieve near-ideal 60.1mV/dec swings. Fig. 5 shows the six-Vt's ranging from the extreme low-Vt to standard-Vt in ~200mV span for N2 N/P FETs. Si data is very close to matching ring speed@standby-power at all six Vts. This multi-Vt capability is enabled with 3rd-generation (since N5) dipole-based multi-Vt integration with both n-type and p-type dipoles.

Much process and device enhancements are focused on engineering not just the transistor drive currents through sheet interface/thickness, junction engineering, dopant diffusion/activation and stress engineering, but more on Ceff reduction to drive best-in-class energy efficiency. All these enhancements lead to much improved I/CV speed gain of 70% and 110% respectively for NS N/P FETs. N2 nanosheet technology exhibits substantially better Perf/Watt than FinFET at low Vdd range of 0.5V-0.6V (Fig. 7). Emphasis is placed on low Vdd perf/watt uplift through process and device continuous improvements resulting in 20% speed gain and 75% lower stand-by power at 0.5V operation. N2 NanoFlex coupled with multi-Vt provides unprecedented design flexibility to satisfy a wide spectrum of energy-efficient compute applications at the most competitive logic density.

Overall technology energy efficiency and performance are also critically dependent on MoL, backend and far-backend interconnects. With innovative materials and processing, VG Rc reduces significantly by 55% with barrier-less all-tungsten MoL. The low resistance MoL combined with capacitance reduction features achieve a total of ~6.2% INV D4 ring oscillator speed gain (Figure 8). Optimized M1 with novel 1P1E EUV patterning led to close to 10% std cell capacitance reduction and a saving of multiple EUV masks. Substantial My RC and Vy Rc reductions seen on the tightest 193i 1P1E workhouse metal/via layers (Fig. 12). In summary, N2 MoL and BEOL RC reduce by ~>20% contributing significantly to energy-efficient compute.

Seamless Integration with 3DFabric Tech

This 2nm platform technology, including the new Cu RDL with flat passivation and TSVs, co-optimizes holistically with 3DIC enabling system integration/scaling for AI/mobile/HPC product designs (Fig. 11-12). Attention is paid to optimize materials and processing in backend/far-backend for global warpage and local planarity for robust integration with 3D stacking. N2 also optimizes pTSV/sTSV (for power/signal) in terms CD/pitch/density for F2F/F2B stacking with SoIC bond pitch scaling from 9µm/6µm down to 4.5µm.

SRAM, Logic Test Chip and Qual/Reliability

For advanced nodes, SRAM bitcell scaling has become a challenge. With N2 NanoFlex and improved on-off current, DTCO is employed to maximize #bitcell/bitline, bitline loading and SRAM peripheral layout efficiency resulting in densest 2nm SRAM macro density ~38Mbmm^2 (Fig. 13). N2 HC/HD pull-down Nfet with better Vt-sigma than FinFet resulting in ~20mV lower HC Vmin and 30~35mV lower HD Vmin (Fig. 14). HD 256Mb SRAM shmoo plot in Fig. 15 illustrates full read and write down to ~0.4V. With innovative well engineering and junction isolation, N2 has better latch-up trigger voltage for both logic and SRAM than FinFet (Fig. 15). Higher Vtrig in N2 leads to additional logic density and more effective DVS screening for product quality. N2 test chip demonstrates healthy CPU/GPU functionality and passed the GPU Vmin-power spec shown in Fig. 16.

N2 256Mb HC/HDSRAM consistently demonstrated healthy defect density resulting in >80% / >90% avg/peak yields (w/o repairs). Fig. 20 shows 256Mb SRAM passed 1000hour HTOL qualification with ~110mV margin.

Additional HPC features such as super high performance MiM (SHP-MiM) with ~>200fF/mm2 capacitance density is offered for higher Fmax by minimizing transient drooping voltage. High-speed SerDes test chip also demonstrated fully functioning 14Gb/s LPDDR6 and 10Gb/s HBM3E interfaces.

References:
[1] Y.J. Mii, IEDM, plenary, 2024.
[2] G. Yeap et al., IEDM, 2019.

[3] "TSMC 30th North America Tech Sym. with Innovations Powering AI with Silicon Leadership," TSMC PR, 2024/04/24
[4] S.-Y. Wu et al., IEDM, 2022.

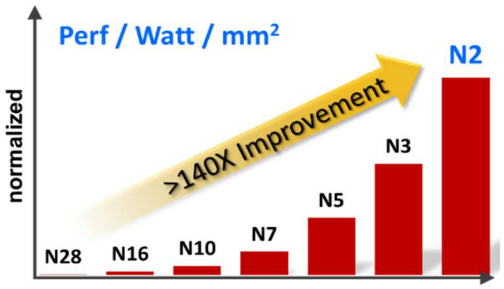


Fig.1 N2 NanoFlex™ accelerates energy-efficient compute

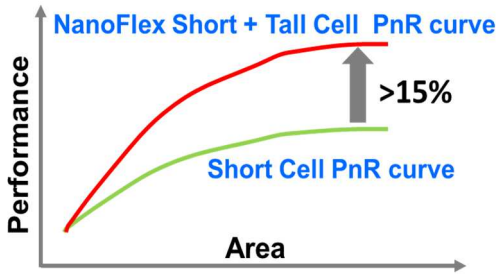


Fig.2 N2 NanoFlex innovation modulates NS width for best PPA. Short cell library for area and power efficiency. Selective use of tall cell library lifts frequency to design target

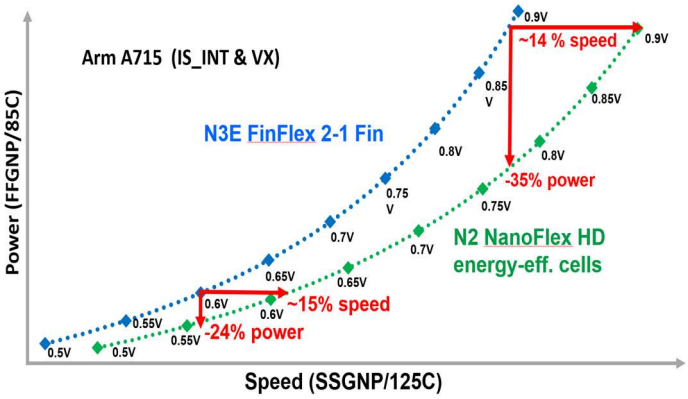


Fig.3 N2 NanoFlex HD cells gain 14~15% speed@power vs. N3E FinFlex 2-1 fin cell across Vdd range: 35% power saving at higher voltage and 24% power saving at lower voltage

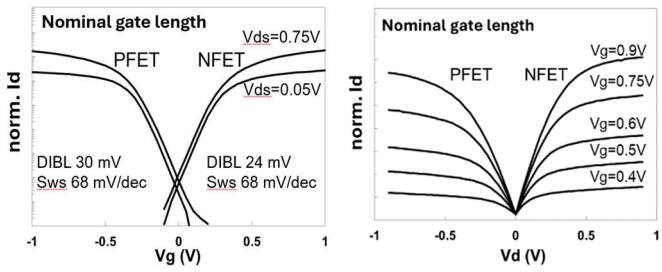


Fig.4 N2 gate & drain characteristics w/ excellent DIBL/Sws

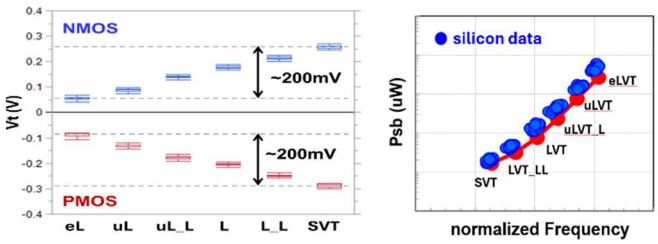


Fig.5 Six-Vt's with ~200mV range for low leakage and high-perf. optimization. Si data closes to matching ring speed @standby-power

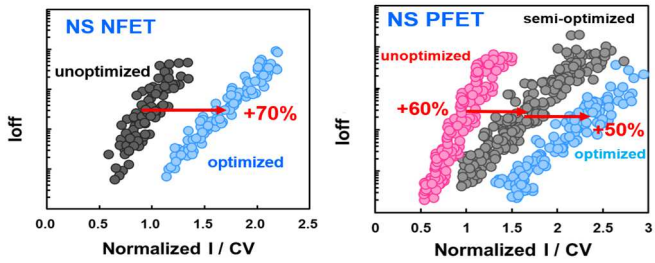


Fig.6 Not only drive current/mobility enhancement, more so on Ceff reduction: N/P +70% and +110% gain in I/CV speed

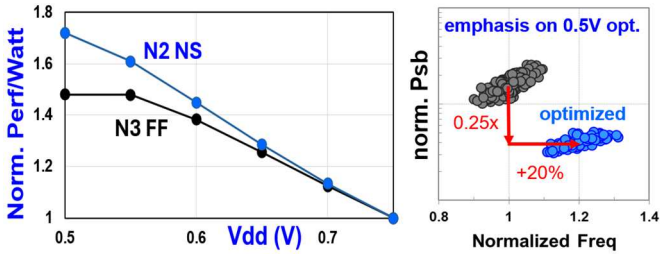


Fig.7 NS vs FF: better Perf./Watt at 0.5V~0.6V. Sp. emphasis at low-Vdd: +20% speed and 75% lower std-by power

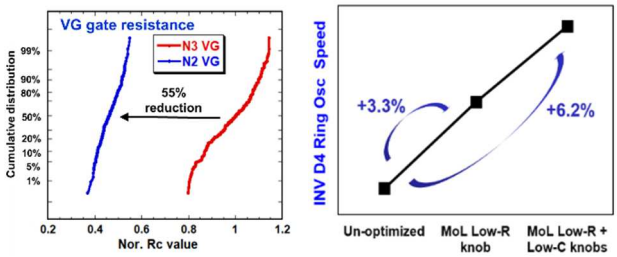


Fig.8 MoL low R (VG, VD and MD) and lower Ceff optimization leading to 6.2% speed gain

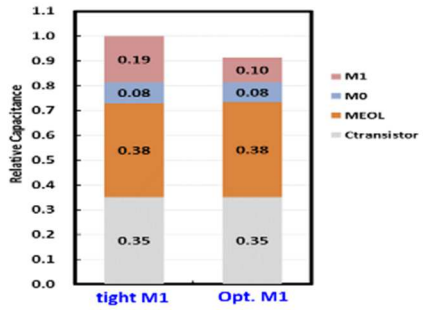


Fig.9 Optimized M1 with novel 1P1E EUV patterning leads to 9% Ceff reduction and a saving of multiple EUV masks

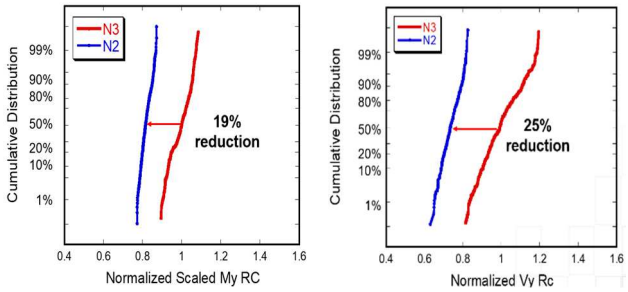


Fig.10 Significant My RC and Vy Rc reduction on N2 tightest 193i 1P1E workhouse metals/vias

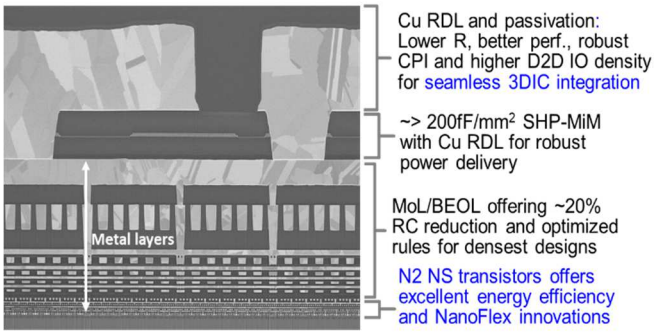


Fig.11 N2 new Cu RDL and passivation provide seamless integration with 3DFabric™ (SoIC, INFO and CoWoS) tech

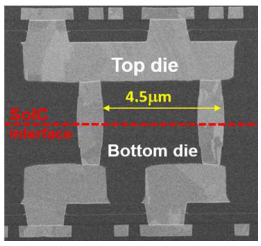


Fig.12 N2 optimizes pTSV and sTSV for F2F/F2B stacking w/ SoIC bond pitch →4.5 µm

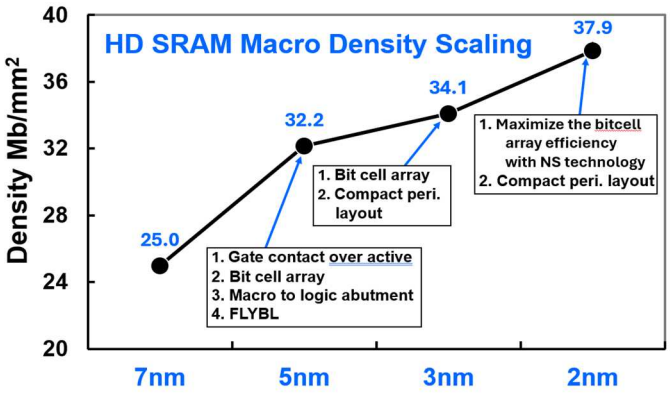


Fig.13 N2 offers highest SRAM macro density ~38Mb/mm²

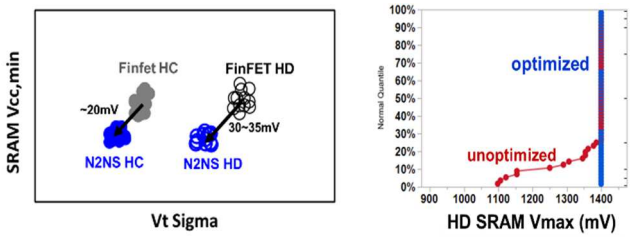


Fig.14 N2NS with better DIBL and Vt-sigma leading to lower Vmin for more energy-efficient compute. HD Vmax>1.4V

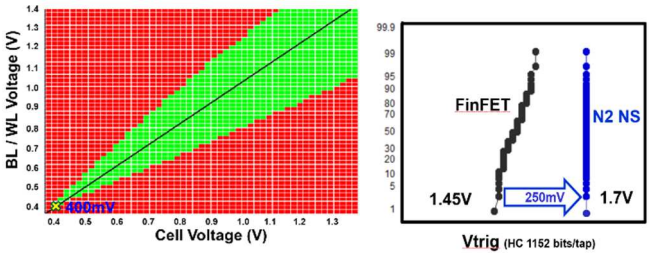


Fig.15 256Mb HD SRAM shmoo to 0.4V. N2 >1.7V Vtrig: higher logic density and effective DVS for product quality

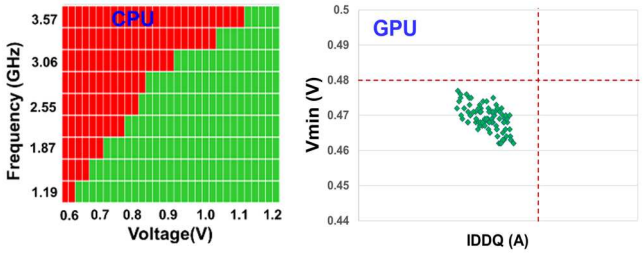


Fig.16 Shmoo plots of CPU block and GPU Vmin vs. IDDQ in the high yielding logic test chip in N2 qual vehicle

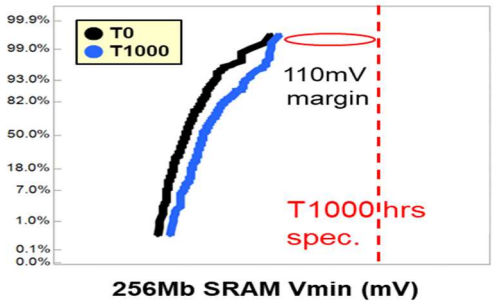


Fig.17 N2 technology met all wafer-level reliability requirements and passed 1000hrs HTOL spec

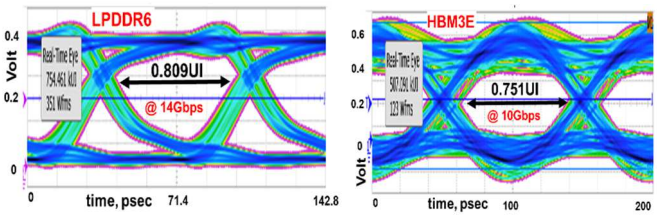


Fig.18 High-speed SerDes test chip in N2 vehicle showing fully functioning LPDDR6 @14Gb/s and HBM3E @10Gb/s

Broadband PureB Ge-on-Si photodiodes in 3.5 µm Ge-filled windows

V.V. Hassan[1], A. Attariabad[1], T. Knezevic[2], N. Rosson[3], J. Italiano[3], and L.K. Nanver[1]
[1]MESA+ Inst. for Nanotechnology , University of Twente, Enschede, The Netherlands, email: v.v.hassan@utwente.nl
[2]Ruđer Bošković Institute, Zagreb, Croatia, [3]Lawrence Semiconductor, Tempe, AZ, USA

Abstract—With the purpose of enabling butt-coupling to SiN waveguides, broadband PureB Ge-on-Si photodiodes were fabricated by filling 3.5-µm-deep oxide windows to a lightly-n-doped Si substrate. Conditions were found that allow growth of good quality Ge and B with high selectivity, resulting in Al-contacted photodiodes with low dark currents and near-ideal responsivity measured at 406, 670, 1310 and 1550 nm to be 0.15, 0.23, 0.58, 0.49 A/W, respectively, for 5.2-µm-thick Ge.

I. Introduction

In a recent publication [1] the successful use of pure boron deposition on Ge was demonstrated to enable the fabrication of nm-shallow p^+-n-like junctions. These junctions were used to fabricate PureB Ge-on-Si photodiodes with low dark currents and near-ideal responsivity for wavelengths from 400 – 1550 nm. The Ge was grown in 1-µm-shallow oxide windows with a low-temperature (400°C), selective chemical-vapor-deposition process that rendered fully-faceted Ge islands. The Ge was capped *in-situ* with B at 700°C. It formed a barrier to an Al-contacting layer but the compactness decreased with increasing defect densities in the Ge.

In this paper we focus on the filling of several micron deep oxide windows. The objective is to enable integration of the photodiodes in a TriPlex photonic integrated circuit (PIC) platform [2] by butt-coupling of Ge detectors to the waveguides. The waveguide fabrication involves several micron thick thermal oxides and long high-temperature steps. Therefore, to obtain high-crystallinity islands, Ge must be deposited at the end of the waveguide processing in very deep oxide windows to the Si, and complete filling is the most optically efficient way to achieve butt-coupling. For this the Ge deposition temperature was increased to 700°C, which not only resulted in filling of the windows but also increased the n-type auto-doping of the Ge from the substrate, phosphorus doped to about 10^{15} cm^{-3}. In itself, this very light doping of the Ge was not problematic, but the P-segregation on the Ge surface before B-deposition degraded the B-layer quality.

This n-dopant segregation issue was only one of several conflicting aspects in the processing that needed to be considered before the goal of fabricating broadband near-ideal photodiodes could be reached. These will be discussed in the following.

II. The B-Ge Interfacial Hole Layer

An important characteristic of PureB diodes is the sheet resistance along the B-semiconductor interface, $\rho_{B\text{-}semi}$, which is determined by the hole concentration accumulated at the semiconductor surface. The B-layer itself has high resistivity and does not significantly contribute to the conductance along the interface. For 700°C B-deposition on lightly-doped n-Si substrates, a $\rho_{B\text{-}Si}$ of ~10 kΩ/sq was consistently found [3]. In [1], the $\rho_{B\text{-}Ge}$ was in all cases found to be ~ 3 kΩ/sq, showing the mobility advantage of Ge. The doping of Ge with B is generally found to be hampered by the very low solubility and diffusivity of B in Ge. This is supported by the simulations displayed in Fig. 1: our experimentally found $\rho_{B\text{-}Ge}$ was a least a decade lower than could be expected from doping with B through methods involving diffusion from a B source. In fact, fabricating low-saturation-current diodes with the p-doping profiles in Fig. 1 would not be possible because integral doping of ~10^{12} cm^{-2} is too low. In contrast, for the PureB Ge diodes studied in [1] the Gummel no. corresponded to 4×10^{14} cm^{-2}, which is comparable to highly-doped, efficient emitters.

III. Experimental Procedures

The basic process flow is illustrated in Fig. 2. Photodiodes were fabricated and optoelectrically characterized on samples with 3 different types of Ge islands grown in windows etched in either 1- or 3.5-µm-thick oxide. The Ge was capped with B grown to a targeted thickness of 10 nm in all cases. The Ge and B depositions were performed consecutively in an ASM Epsilon 2000 CVD reactor equipped with germane (GeH_4) and diborane (B_2H_6) carrier gases. The cleaning parameters and deposition conditions for each of the 6 sample types are listed in Table I. The LB400 and LB700 samples, received a low-temperature (LT) bake in H_2 at 900°C for 3 min, while the HB700 samples where H-baked at a high-temperature (HT) of 1120°C for 1.5 min. The bulk Ge was grown at either 400°C or 700°C and B was deposited at 700°C. An Al contact metallization was patterned by lift-off and alloyed at 400°C.

Optoelectronic measurements were performed at 25°C using an in-house setup built around a Karl Suss MicroTec PM300 wafer-prober, as described in detail in [1]. A Keithley 4200 parameter analyzer was used to measure *I-V* characteristics, the latter also while exposing the diodes to laser light from a 4-channel laser source, Thorlabs MCLS1, using fiber-coupled laser diodes. The available wavelengths were λ = 406, 670, 1310, and 1550 nm. The optical power received by the diodes was measured using Thorlabs slide power sensors.

IV. Material and Electrical Characterization

Scanning electron microscopy (SEM) and optical images of the 6 types of Ge-islands are shown in Figs. 3 and 4, respectively. The LB400 and HB700 Ge growth was highly

selective which meant that loading effects caused the island thickness to increase with decreasing window size and increase of the surrounding oxide area. This is also true for the B-layer thickness, hence the color differences seen optically. In contrast, the LB700 samples display very poor selectivity and all windows are uniformly grey, indicating that the B was distributed fairly uniformly over the wafers. On the oxide Ge deposition is seen as large balls about the same height, or even higher, than the Ge filling the windows.

The growth of such balls depends on the presence of nucleation centers on the oxide, presumably in the form of Si clusters that apparently were removed by the HT bake. This is feasibly due to the well-known reaction $Si + SiO_2 \rightarrow 2SiO\uparrow$. The SiO is volatile, thus removing Si from the surface. For the same reason the HB700 images also display an undercut at the oxide window perimeters. This is an undesirable effect, but it did not, as seen in Fig. 5, create extra leakage currents in reverse bias as might be expected due to increased carrier trap density.

As desired, the LB700 and HB700 windows are filled with Ge, the HB700 with rounded corners due to loading effects that became mushroom-like upon overgrowth. On the other hand, the Ge "starved" LB700 windows were filled with faceted Ge surfaces. The LB400 Ge islands were flat-top pyramids for which the complete Ge surface was capped with B. The LB700 Ge had oxide sidewalls but both diode types had practically identical *I-V*'s in both forward and reverse. As opposed to this, the HB700 diodes had low-voltage current levels that were more than a decade lower than for the LB diodes, and also did not scale with area as seen in Fig. 6.

V. Optical Characterization

The responsivity of all sample types are compared in Fig. 7. The LB samples display quite similar values. A comparison to simulations performed as a function of the Ge thickness is shown in Fig. 9. The LB700d values with a Ge thickness of 5.2 μm lie very close to the simulated values over the whole wavelength range. As the Ge thickness is reduced, the 406 and 670 nm responsivity does not change much with thickness.

On the other hand, the responsivity of the HB samples was all but completely destroyed at 1310 and 1550 nm, while being only somewhat lower for the 2 short wavelengths. This was found to be related to poor B-layer quality as verified by the high B etch rates in Al-etchant (Table 3 and Fig. 8). The $\rho_{B\text{-}Ge}$, is also significantly lower than the ~ 3 kΩ/sq found for the LB devices. In accordance with responsivity lowering due to Al-mediated migration of Ge and Si described in detail in [4], the overall results indicate that the poor B quality has allowed Al to contact the Ge through weak spots, particularly at the diode perimeter. This leads to Al p-doping of the Ge and shifts the diode junction from the Ge into the Si, thus lowering the electron current and making the diodes non-responsive to the longer wavelengths.

The mechanism responsible for deteriorating B-layer quality is thought to be the segregation of P on the Si during the HT bake. The LT bake reduces this effect. However, with Ge growth at 700°C, any segregated P will slowly be built into the Ge, leaving a segregation layer on the surface to be capped with B. From PureB Si studies, it was found that n-dopants on the surface deteriorated the B-layer quality [5]. The LB700 samples had slightly degraded B-layer properties as compared to the LB400 samples but there was no sign of Al-migration and the responsivity was high for the thick LB700d Ge. The LB700s had very thin 0.8 μm Ge and the responsivity suffered from the proximity to the defected Ge-Si interface.

The oxide-related issues encounter here in connection with the H-bake step have solutions that only involve changes in the oxide processing. The n-dopant segregation is, however, a more fundamental problem that will become particularly critical if low-ohmic substrates are applied, for example, to lower series resistance, or to eliminate the Ge-Si interfacial barrier as discussed in [4], thus allowing photodiode operation at a few volts instead of high values like the -20 V used here to reduce the barrier. Since 400°C Ge growth significantly reduces the segregation effect, growing a thick 400°C layer before filling with a 700°C layer may be a solution. Otherwise (undesirable) *ex-situ* surface cleaning could be employed. Since n-dopants on the Ge surface did not directly influence the *I-V* characteristics, the function as material barrier could also be allocated to standard layers such at TiN. This would also be a solution if, in the long run, the B-layer quality, that was consistently much poorer that the very robust layers routinely grown on Si [6], becomes a reliability issue.

Overall, the results are exceptional in that Ge filling of the micron deep windows was successful, a B-layer quality suitable for direct Al-contacting was achieved, and the PureB Ge-on-Si photodiodes fabricated in the thickest 5.2 μm Ge had low dark currents and near-ideal responsivity measured at 406, 670, 1310 and 1550 nm to be 0.15, 0.23, 0.58, 0.49 A/W, respectively.

Acknowledgment

This work was funded by EU HORIZON-RIA project 101070441, and partly by Croatian Science Foundation (HRZZ) under the project IP-2022-10-5294 and the European NextGenerationEU fund – National Recovery and Resilience Plan Development Research Grants (grant no. NPOO.C3.2.R2-I1.06.0025).

References

[1] L.K. Nanver, V.V. Hassan, A. Attariabad, N. Rosson, C.J. Arena, "Broad-band PureB Ge-on-Si photodiodes," IEEE EDL, 45-6, p. 1040, 2024.

[2] C.G.H. Roeloffzen, et al., "Low-Loss Si3N4 TriPleX Optical Waveguides: Techn. and Appl. Overview," IEEE JSTQE, 24-4, p. 4400321, 2018.

[3] L. Qi and L.K. Nanver, "Conductance along the interface formed by 400C pure boron deposition on silicon," IEEE EDL, 36-2, pp. 102-104, 2015.

[4] L. Marković, et al., "The impact of the Ge-Si interfacial barrier on the temperature-dependent performance of PureGaB Ge-on-Si p^+n photodiodes," Optics Express, Aug. 2024. DOI: 10.1364/OE.530466

[5] L.K. Nanver, K. Lyon, X. Liu, J. Italiano, and J. Huffman, "Material Reliability of Low-Temperature Boron Deposition for PureB Silicon Photodiode Fabrication," MRS Advances, vol. 3, pp. 3397–3402, 2018.

[6] S.D. Thammaiah, T. Knezevic, L.K. Nanver, "Nanometer-thin pure boron CVD layers as material barrier to Au or Cu metallization of Si," J. Material Science, Materials in Electronics, 32, pp. 7123–7135, 2021.

[7] L. Marković, T. Knežević, L. K. Nanver and T. Suligoj, "Modeling and Simulation Study of Electrical Properties of Ge-on-Si Diodes with Nanometer-thin PureGaB Layer," 2021 MIPRO, pp. 64-69, 2021.

[8] T.N. Nunley, et al., "Optical constants of germanium and thermally grown germanium dioxide from 0.5 to 6.6 eV via a multisample ellipsometry investigation," J. VSTB, Nanotechnol. Microelectron., Mater., Process., Meas., Phenomena, 34-6, 2016.

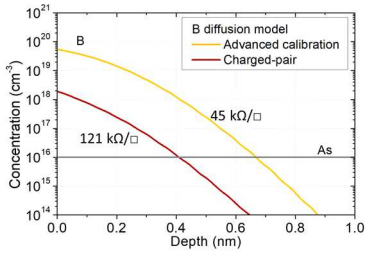


Fig. 1. Simulations performed with Synopsys Sentaurus Process simulator, in which either the default advanced calibration model with optimized parameters for the latest technology nodes or the charged-pair diffusion model with the parameters proposed in [7] were applied. A 5-nm-thick Ge-layer doped with B to a concentration of 10^{20} cm^{-3} was used as dopant source to mimic diffusion during a 700°C B deposition. Sheet resistance was simulated by using Sentaurus Device.

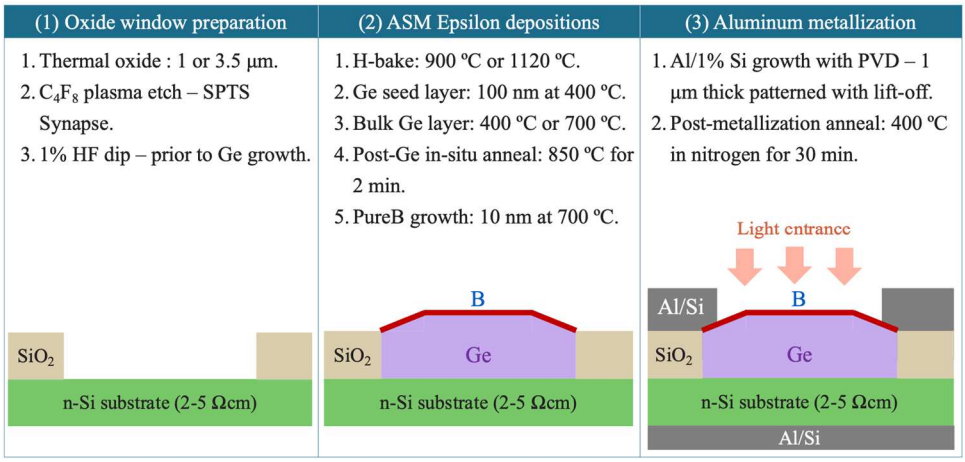


Fig. 2. Basic process flow for fabricating PureB Ge-on-Si photodiodes.

Sample type	Window depth (µm)	H-bake temperature/time	Dep. Temp (°C)	Selectivity	Ge island morphology	Ge Thickness 1x1cm^2 (µm)	Ge Thickness 25x25µm^2 (µm)	Oxide undercut
LB400s	1	900°C 3 min	400	Good	Fully facetted	1.34	1.45	-
LB400d	3.5	900°C 3 min	400	Good	Fully facetted	4.70	4.73	-
LB700s	1	900°C 3 min	700	Poor	Filled, surface facetted	0.27	0.79	-
LB700d	3.5	900°C 3 min	700	Poor	Filled, surface facetted	3.70	5.40	-
HB700s	1	1120°C 1.5 min	700	Good	Filled, surface rounded	0.25	0.75	~ 1 µm
HB700d	3.5	1120°C 1.5 min	700	Good	Filled, surface rounded	0.61	4.68	~ 1 µm

Table 1. Germanium deposition parameters and properties for 25×25 µm^2 windows.

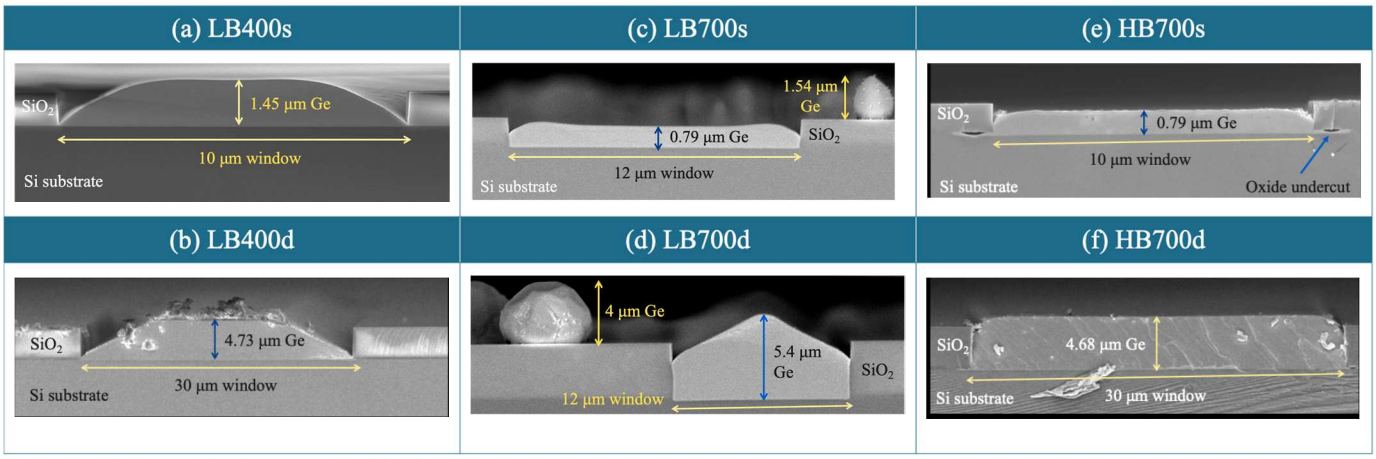


Fig. 3. SEM images of the six types of Ge islands.

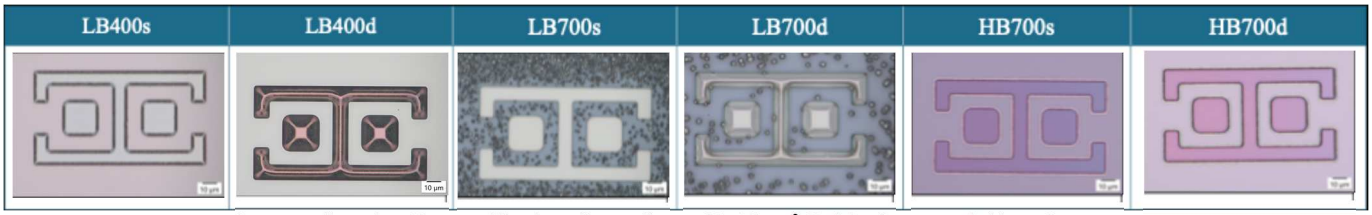


Fig. 4. Optical microscope images collected at 50x magnification of two adjacent 25×25 µm^2 Ge islands surrounded by a Ge moat.

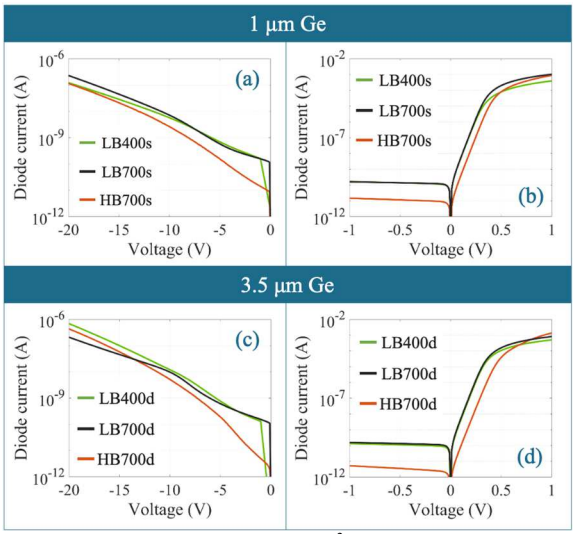


Fig. 5. IV characteristics for 25×25 μm² diodes.

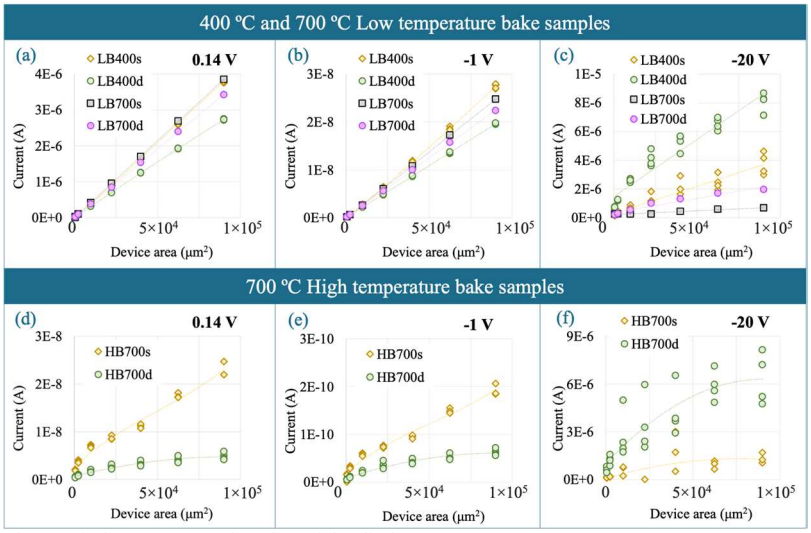


Fig. 6. Diode current as a function of the diode area.

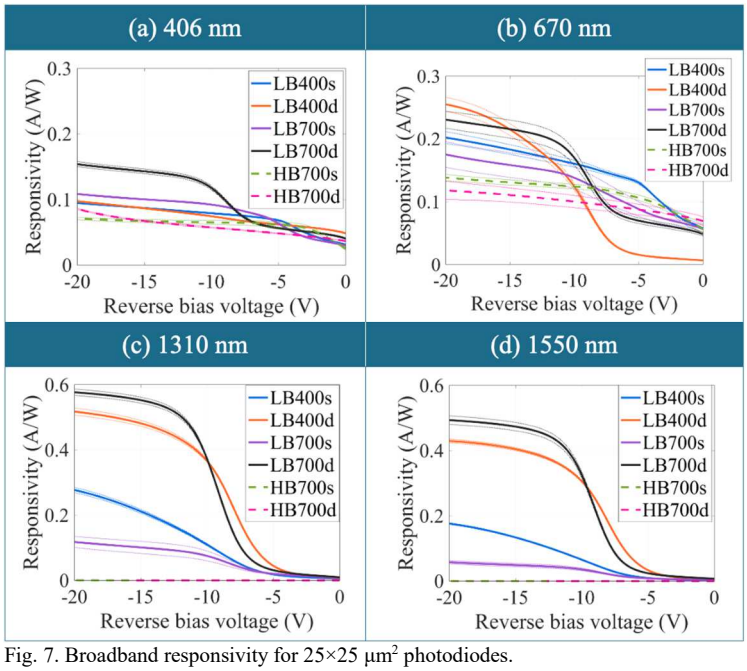


Fig. 7. Broadband responsivity for 25×25 μm² photodiodes.

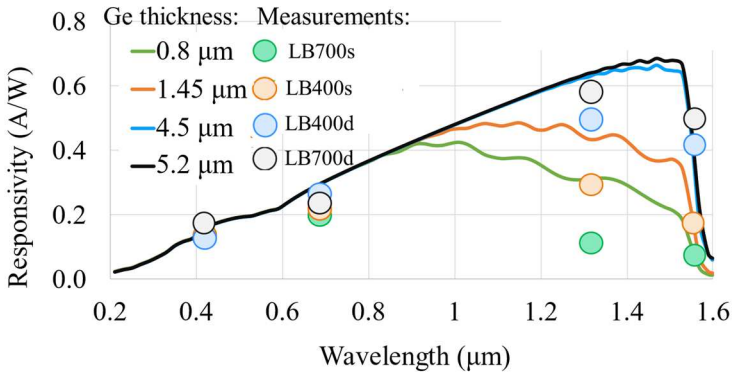


Fig. 9. Simulated responsivity as a function of wavelength, compared to measured values. The simulations were conducted using Sentaurus Device from Synopsys TCAD. The complex refractive index for silicon was adopted from the simulator, while the optical constants for Ge were taken from [8], and for B from [1]. Optical generation was simulated using the transfer matrix method.

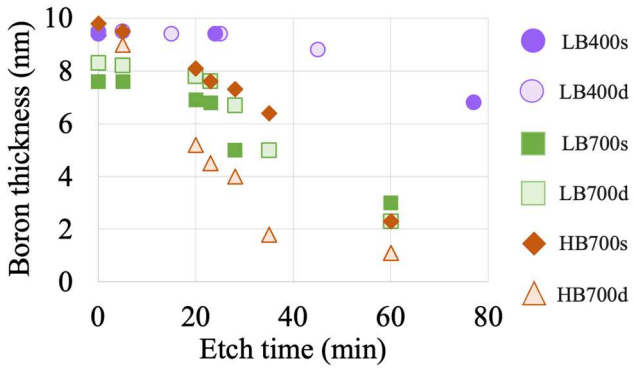


Fig. 8. Boron etch rate in Al etchant at room temperature

Sample type	I (0.14 V)	I (-1 V)	I (-20 V)	Responsivity (A/W) 1310 nm (-20 V)
LB400s	2.4×10^{-8}	1.7×10^{-10}	1.2×10^{-7}	0.28
LB400d	1.9×10^{-8}	1.4×10^{-10}	6.9×10^{-7}	0.52
LB700s	2.5×10^{-8}	1.6×10^{-10}	2.3×10^{-7}	0.12
LB700d	2.4×10^{-8}	1.6×10^{-10}	2.1×10^{-7}	0.58
HB700s	1.8×10^{-9}	1.5×10^{-11}	1.1×10^{-7}	0.0005
HB700d	4.0×10^{-10}	5.2×10^{-12}	4.4×10^{-7}	0.0005

Table 2. Optoelectrical characteristics of photodiodes with area 25×25 μm²

Sample type	Thickness + roughness (nm)[a]	ρ_{sh} (kΩ/□)[b]	Resistivity (kΩ-cm)[c]	Etch rate (nm/min)[a]
LB400s	9.4+1.5	2.90	12.2	0.03
LB400d	9.5+1.5	2.78	8.3	0.02
LB700s	7.6+1.3	2.73	5.2	0.08
LB700d	9.8+1.4	3.46	5.3	0.10
HB700s	9.8+1.4	2.29	6.5	0.13
HB700d	9.6+1.1	1.66	10.7	0.22

Table 3. Boron properties. Measurements on 1×1 cm² windows ([a]), 200×200 μm² Van der Pauw structures ([b]), 25×25 μm² diodes ([c]).

Low-temperature Behavior in Nanowire Transistors by Quantum Transport Simulation

S. Moslemi-Tabrizi[1], P. Schvan[1], U. Kapoor[2], P. Blaise[2], D. Lemus[3], and T. Kubis[3]
[1]Ciena Inc, Ottawa, Canada, email: smoslemi@ciena.com
[2]Silvaco Inc, Santa Clara, CA, USA, [3]University of Purdue, West Lafayette, IN, USA

***Abstract*—** The simulation of a silicon nanowire n-FET is performed using the NEGF quantum transport solver Victory Atomistic down to a temperature of 2 K. Inner spacers offering an optimal electrostatic confinement are used promoting several localized states appearing in the channel close to the bottom of the conduction band. The NEGF simulations allow probing of the NWFET behavior with various localized states at low temperatures, by gradually injecting electrons in the conductive channel and by taking into account the electron-phonon scattering mechanisms.

I. Introduction

The physical dimensions of FET transistors continue to shrink thanks to the VLSI integration and EUV lithography, this is toward a generalized usage of nanosheet and nanowire transistors made of silicon, (NSFET and NWFET), with a typical cross-section of a few nm^2 [1]. At the same time, there is a tremendous regain of interest in using silicon-based transistors at low temperatures for a few Kelvin and below, for high-performance computing, new electronic memories, and quantum computing [2]. This poses difficulties to conventional TCAD simulation of devices, being unable to capture a correct energetic landscape of electrons at a small scale, (i.e. beyond the effective mass approximation due to the strong confinement effect), and hardly converging at low temperatures. In that respect, the non-equilibrium Green's function (NEGF) method offers several advantages and can remedy TCAD difficulties, which is what we showcase in this work.

II. Atomistic Simulation Methodology

A. NEGF code

Throughout this study, we use the VictoryAtomistic tool [3] an evolution of Nemo5 [4] delivering a decisive speed-up in NEGF calculations thanks to a generalization of the modespace approximation technique [5]. Electronic bandstructure is described by a tight-binding $sp^3d^5s^*$ basis offering an accurate description of the electronic levels close to the Si bandgap plus a very good transferability to an NWFET geometry with a nanometric section. The silicon dangling bonds are fully passivated with hydrogen atoms thanks to an ad-hoc self-energy of passivation. An adaptive energy mesh is used with an average resolution of 1 meV for T = 300 K and 60 K, and with a resolution better than 0.2 meV for T = 20 K and T = 2 K. Each simulation is performed using a suitable low-rank approximation matrix [5] that preserves the calculation accuracy and decreases the simulation time by more than two orders of magnitude. This combination of techniques allows one to test many configurations and temperatures, alleviating difficulties previously encountered at low T below a few dozen K in NEGF [6].

B. Physical dimensions and parameters

A typical NWFET structure is visible in Figure 1, showing the essential geometric characteristics and its corresponding atomistic structure. The crystallographic direction along the transport direction is Si <100>. The channel is made of intrinsic Si with a gate-all-round (GAA) structure of 1nm EOT. There are two inner spacers on the source and drain sides. The dielectric thickness of the spacers is 1nm and one uses several dielectric constants. The source and drain are highly doped. All the essential characteristics are summarized in Table 1.

The electron-phonon interactions are included thanks to the self-consistent Born approximation. The physical parameters relative to the acoustic and optical branches are summarized in Table 2. We use silicon-bulk parameters as a first guess to probe the essential effect of phonons on the current characteristics and levels broadening.

C. NEGF self-consistency

Our NEGF simulator solves for a steady state of the electronic device. Its self-consistency is reached when the electronic density solved with the $G^<$ component of NEGF doesn't significantly change when iterated in Poisson's equation solver. Then one can have access to the non-equilibrium occupation numbers and electronic current through the channel, as illustrated in Figure 2. The electron-phonon coupling is naturally taken into account by the self-energy terms. It is worth noting that both the self-energy in the channel and the occupation numbers of each electrode depend directly on temperature.

III. Intermediate Temperatures Results

We study an NWFET of 2x2 nm^2 square cross-section, for which the bandgap becomes direct at the Γ point due to a strong confinement effect. We work at low Vds = 10 mV. The Id(Vg) transfer characteristics are shown in Figure 3 for two temperatures 300 K and 60 K. Thanks to the good electrostatic control offered by the GAA geometry, the subthreshold slopes are obtained at 64 mV/dec and 13 mV/dec for a current of 10^{-10} A at T= 300 K and 60 K respectively, (a bit higher than the predicted theoretical values essentially because of the e-ph coupling). Above the threshold voltage, the saturation current is lower at room temperature due to an increased e-ph coupling. More interestingly, one can see a kind of current fluctuation at

T = 60 K not visible at room temperature. This fluctuation is related to a few quantum localized states that appear below the top of the barrier, as illustrated in Figure 4 for the density of states (DOS) resolved in energy and space obtained at Vg = 0.6 V. Indeed these localized states contribute only very little to the current density as a function of the applied gate voltage. The smearing of levels at T = 300 K on several kB.T doesn't allow these levels to get any weight in the I(V) characteristics, and they start to be hardly perceived at T = 60 K. Moreover a level splitting of 5 meV is visible at T= 60 K for the lowest energy level. These encouraging preliminary results permit us to study the NWFET at even lower temperatures.

IV. Results at T = 20K and T = 2K

A. Electronic levels as a function of Vg at T = 20 K

The confined electronic levels start to detach from the conduction band for a gate potential higher than 0.5 V. Two levels below the top of the barrier are going deeper into the quantum well formed by the channel and the two spacers as illustrated Figure 5 showing the DOS resolved in energy and space. They are followed by other states higher in energy showing a different symmetry with supplemental lobes along the transport direction. Thanks to low Vds applied and the spatial and energetic proximity of the electrodes, in a steady state regime, the lowest energy levels start to be filled by a few electrons, as can be seen in Figure 6 showing the electronic density resolved in energy and space.

B. States occupancy from T = 20 K to T = 2 K

At T = 2 K, the two lowest levels in energy gain a supplementary splitting of 1 meV, see Figure 7. This does not affect the global charge of the levels estimated in Figure 8 for the two temperatures 20 K and 2K. This net charge is obtained by counting the number of electrons in the channel and removing the contribution from any other states than the levels detected in the same energy window and spatial location of interest. Interestingly, the coupling of the charge in the Vg = [0.48, 0.64] V window is almost independent of the temperature and is purely capacitive with a 2 aF proportionality.

C. Confinement function of geometry and dielectric spacers

Figure 9 shows the DOS resolved in energy and space for a slightly larger and rectangular cross-section of 2x3 nm^2. In comparison with the 2x2 nm^2 square cross-section, the splitting of the first energy levels increases from 5 meV to 7 meV. This underlies the importance of describing accurately the atomistic structure and the nontrivial symmetries of the atomic arrangements in the confinement directions.

Figure 10 illustrates the effect of employing longer dielectric spacers of 14 nm instead of 7 nm with two different values for the dielectric constant $\varepsilon = 12$ and $\varepsilon = 4$. The DOS resolved in energy doesn't show a much better confinement by employing longer spacers. Nevertheless, by decreasing the spacer dielectric constant, the energy confinement is increased by 5 meV, and the spatial localization is only slightly improved. This is explained by a deeper quantum well formed by the presence of the two spacers that are less coupled to the gate when the dielectric constant is lowered.

D. Energy relaxation of the electronic states due to phonons

By taking into account the self-energy of acoustic and optical phonons of silicon in interaction with electrons, the electronic levels of confined states are broadened with increasing temperature, see Figure 11. To resolve the four levels of lowest energy of the confined states, one needs to lower the temperature below 2 K, (note that the energy resolution of Figure 11 is 0.1 meV approx.). For the conduction band of silicon, the inclusion of the spin-orbit coupling interaction doesn't change significantly this result, (calculation not shown).

V. Conclusion

Thanks to an efficient combination of several numerical optimizations, a systematic exploration of the density and density of states resolved in energy and space becomes doable for NWFETs with the NEGF feature of Silvaco's Victory Atomistic. The results obtained at low temperatures show the importance of combining an atomistic resolution with a suitable basis to describe accurately the electronic levels. For the steady state of a Si NWFET oriented <100> and including inner spacers, several confined states appear near the threshold voltage. These states present a complex energetic profile close to the conduction band of silicon depending on several factors and physical parameters. They contribute very slightly to the overall current of the NWFET but can be filled by a few electrons at low Vds voltage. On one hand, the capacitive coupling with the gate can be quantified, on the other hand the short channel combined with the spacers contributes to the efficiency of the quantum well. For better control of low-T devices, due to electron-phonon coupling, a temperature below 2K is required to address each state individually by resolving the level broadening.

Acknowledgment

Part of the calculations were performed using Negishi's machine at Purdue's Rosen Center of Advanced Computation.

References

[1] A. Razavieh, P. Zeitzoff, and E. J. Nowak, "Challenges and Limitations of CMOS Scaling for FinFET and Beyond Architectures," in IEEE Transactions on Nanotechnology, vol. 18, pp. 999-1004, 2019.

[2] K.K. Likharev, "Single-electron devices and their applications," in Proceedings of the IEEE, vol. 87, no. 4, pp. 606-632, April 1999

[3] Victory Atomistic User's Manual by Silvaco International, (2023).

[4] S. Steiger, M. Povolotskyi, H.-H. Park, T. Kubis, and G. Klimeck, "Nemo5: a parallel multiscale nanoelectronics modeling tool", IEEE Trans Nano, vol. 10, p. 1464, 2011

[5] Lemus, D.A., Charles, J. & Kubis, T. Mode-space-compatible inelastic scattering in atomistic nonequilibrium Green's function implementations. J Comput Electron 19, 1389–1398 (2020)

[6] Lavieville R, Triozon F, Barraud S, Corna A, Jehl X, Sanquer M, Li J, Abisset A, Duchemin I and Niquet Y-M 2015 Quantum dot made in metal oxide silicon-nanowire field effect transistor working at room temperature Nano Lett. 15 2958–64

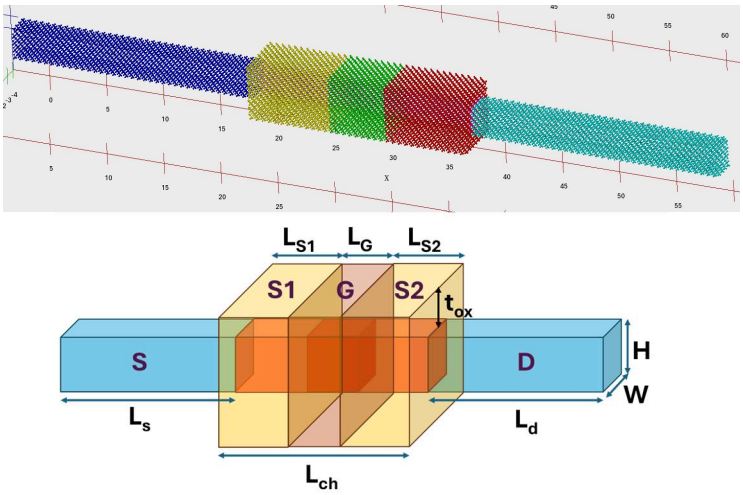


Fig. 1. Si nanowire GAA-FET structure of square cross-section including dielectric spacers. Top: atomistic structure of a 2x2 nm^2 nanowire with the source (blue) and drain (turquoise) of 20 nm length each, gate of 5 nm (green), and dielectric spacers (yellow and red) of 7 nm each. Bottom: schematic of the NWFET geometry with the various characteristic lengths.

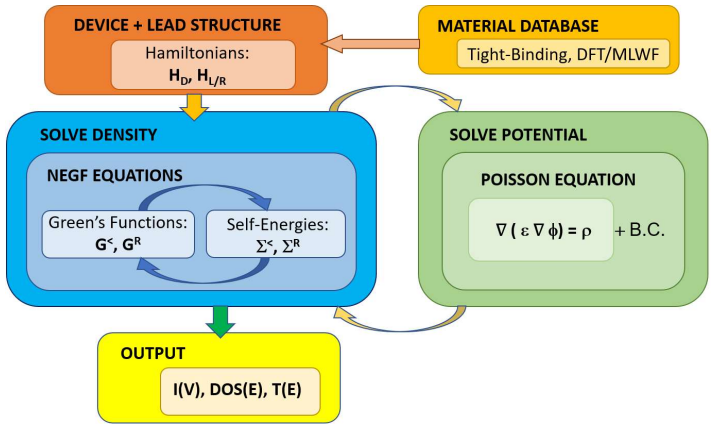


Fig. 2. Victory Atomistic flowchart showing the NEGF solver coupled with the Poisson equation solver. The Hamiltonian of the electrodes and the central device is built within a tight-binding formalism and with open boundaries conditions. The transfer characteristic is calculated when the self-consistency of NEGF and Poisson is reached.

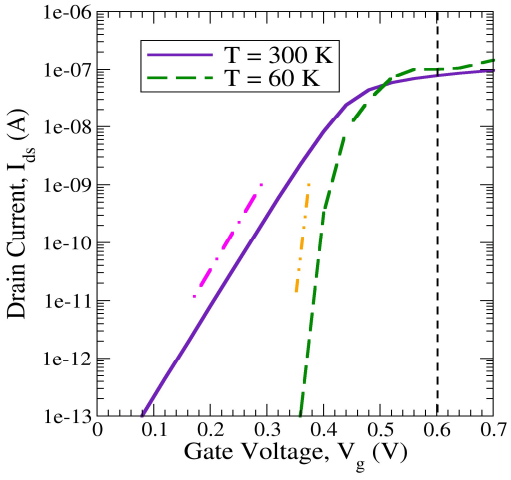


Fig. 3. Id(Vg) current transfer characteristic at T = 300 K and 60 K, Vds = 10 mV, along with the ideal slopes of 60 mV/dec (magenta) and 12 mV/dec (orange).

Param.	Value/range
Cross-section	2x2, 2x3 nm^2
L_S, L_D	20 to 40 nm
L_{CH}	19 to 33 nm
L_{S1}, L_{S2}	7 to 14 nm
L_G	5 nm
t_{OX}	1 nm
$\varepsilon_{S1,S2}$	4 to 20 (ε_0)
ε_G (SiO_2)	3.9 (ε_0)
Doping$_{(CH)}$	10^{15} (e.cm^{-3})
Doping(S,D)	10^{20} (e.cm^{-3})

Table 1. Physical dimensions of the Si <100> NWFET, dielectric properties and doping.

Param.	Value/range
Temperature	2, 20, 60, 300 K
V_{sound} (Si)	8433 m.s^{-1}
ρ(Si)	2.336 g.cm^{-3}
D_{ADP}	8.8 eV
ω_{op}	63 meV
D_{ODP}	110 eV.nm^{-1}
W_F	4.2 eV
V_{DS}	10 mV
V_G	0.0 to 0.7 V

Table 2. Physical parameters employed for the NWFET simulations.

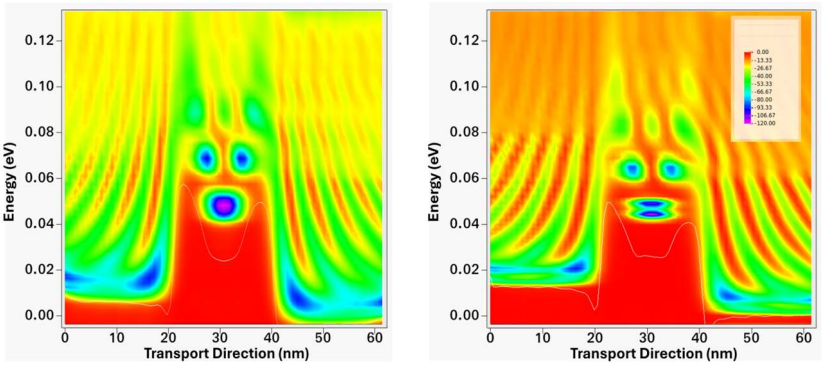


Fig. 4. Density of states resolved in energy and position along the transport axis obtained at T = 60K (left), T = 300K (right) with Vg = 0.6 V, Vds = 10 mV. The color scale is in arbitrary units from low density (red) to high density (violet). A few confined states are visible in the channel region below the gate around 50 meV. The white line is here to visualize the 1D electrostatic profile aligned with the bottom of the conduction band on both leads. At T = 60K a first splitting of 5 meV for the lowest energy level appears.

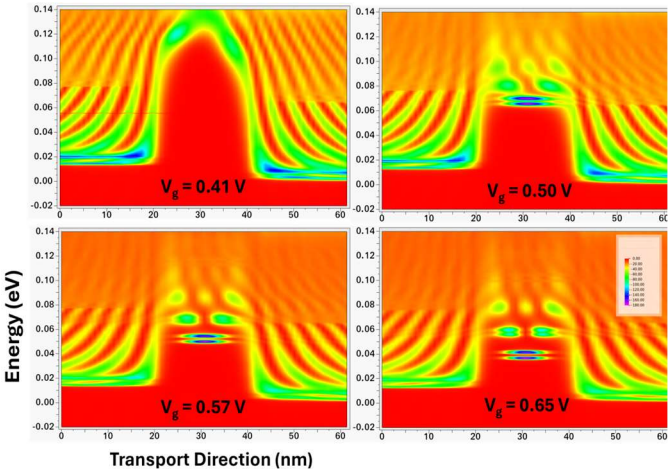


Fig. 5. Density of states resolved in energy along the transport axis at T = 20K, Vds = 10 mV, for four different values of Vg = 0.4, 0.5, 0.57, 0.65 V. Two localized states appear at Vg = 0.5 V slightly below the top of the barrier. These two states are pushed lower in energy as Vg increases, followed by other states showing a different symmetry.

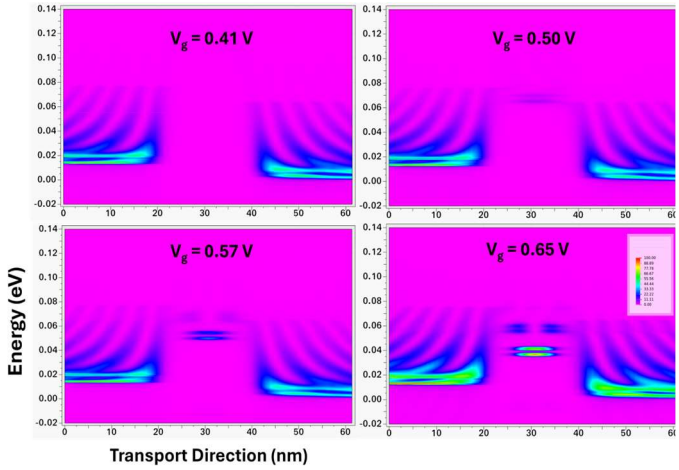


Fig. 6. Electronic density resolved in energy along the transport axis at T = 20K, Vds = 10 mV. The electronic levels are progressively filled using four different values of Vg = 0.4, 0.5, 0.57, 0.65 V, with a total number of electrons approximately equal to one at Vg = 0.57 V, and equal to two at Vg = 0.65 V.

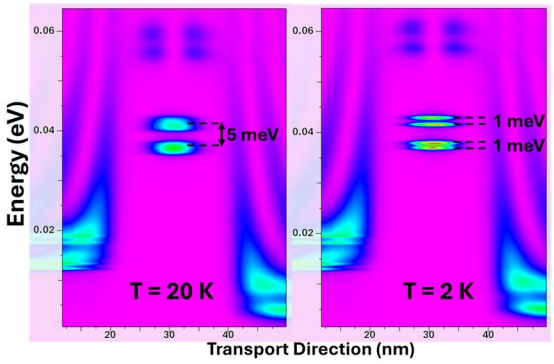


Fig. 7. Electronic density resolved in energy along the transport axis at T = 20K vs 2K, Vds = 10 mV, Vg = 0.65 V. At T = 20 K a level splitting of 5 meV is visualized. At T = 2 K, a second splitting of 1 meV occurs for the lowest energy levels, the progressive filling of these levels being unchanged.

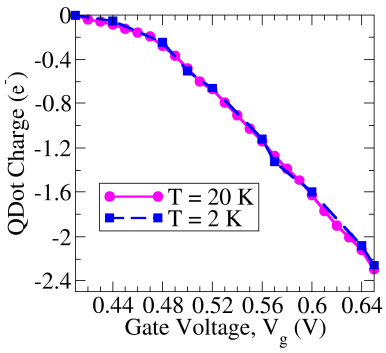


Fig. 8. Total electronic charge captured by the localized states of lowest energy function of Vg at T = 2K and 20K.The total charge follows a quasi-linear law for Vg=[0.48, 0.64] V.

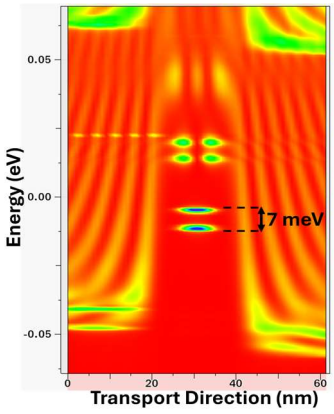


Fig. 9. Density of states resolved in energy along the transport axis at T = 20 K, Vds = 10 mV, Vg = 0.65 V for a rectangular NW of 2x3 nm^2 showing a level splitting of 7 meV.

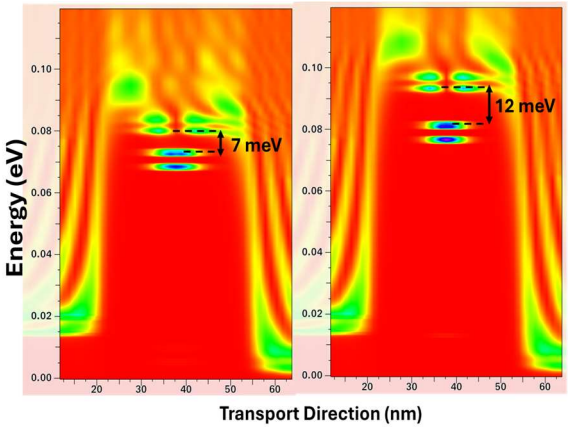


Fig. 10. DOS(E, x) along the transport axis at T = 20K, Vds = 10 mV, Vg = 0.57 V using long spacers of length 14 nm, and with two different dielectric constants ε = 12 (left), ε = 4 (right). A lower dielectric constant helps in confining the electronic levels.

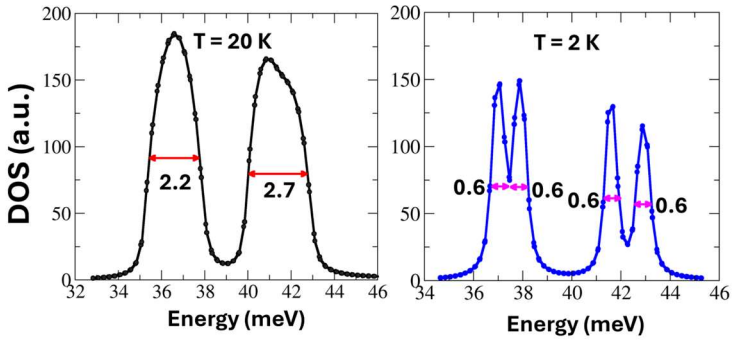


Fig. 11. Energy relaxation of the electronic states of lowest energy in the transistor channel due to the electron-phonon coupling for two temperatures at T = 20 K (left panel) and T = 2K (right panel), Vds = 10 mV, Vg = 0 .65V. Each level broadening is estimated with the full width at half maximum. Acoustic and optical phonon branches are taken into account in NEGF simulation within the self-consistent Born approximation.

Ultrafast Charge Trap-based Volatile Memory Cell with Schottky Barrier S/D and Thin Tunnel Oxide

Hyeongyu Kim[1*], Dabok Lee[2*], Hyun-Sik Choi[3], Yoojin Seol[1], Jonghyeon Ha[2], Sung-Il Chang[4], Kihyun Kim[1], Jungsik Kim[2], Won-Ju Cho[3], Zvi Or-Bach[4]

[1]Division of Electronic Engineering, Jeonbuk National University, Jeonju, Jeonbuk, 54896, Rep. of Korea

[2]Department of Electrical Engineering, Gyeongsang National University, Jinju, Gyeongnam, 52828, Republic of Korea

[3]Electronic Materials Engineering, Kwangwoon University, Seoul, 01897, Rep. of Korea

[4]MonolithIC3D, Klamath Falls, OR 97601, United States

*: Equal contribution; Corresponding E-mail: sung.il@monolithic3d.com

***Abstract*—**In this paper, we propose a novel 3D Charge Trap-based DRAM (CT DRAM) having great advantages on power consumption and heat dissipation. Schottky barrier-induced hot carrier injection and ultrathin tunnel oxide were used for fast write operation. Microwave anneal (MWA) shows better silicide formation (low resistance and high diode on/off ratio) than conventional rapid-thermal annealing (RTA) and its volume heating is quite effective for 3D integrated devices. We designed key parameters (Schottky barrier height, ONO thickness, and trap characteristics) with 3D TCAD simulation and fabricated the planar device having poly-Si channel to explore 3D integration. The device is expected to have a program/erase window larger than 2 V below 10 ns write pulse, and retention time longer than 10s@85°C, and >0.22V window after 10^{15} cycles.

I. Introduction

The current technology of 10-nm saddle-fin-based DRAM has reached the limit of miniaturization due to various reasons, such as the limitation of meeting the minimum line resistance and capacitance values and the reliability problem of bit flip due to row hammer [1-2]. Although various attempts are also being made to develop next-generation 3D DRAMs, 1T1C 3D DRAM is extremely hard to integrate, and 2T IGZO 3D DRAM is not cost-effective for over several hundreds of layers [1, 3]. Therefore, developing a DRAM structure with a straightforward 3D integration scheme like 3D NAND is essential to achieve higher bit density [4].

In this paper, we present a novel 3D charge trap-based memory cell structure as a next-generation DRAM, as shown in Fig. 1. It uses an ONO structure instead of a capacitor and a metal silicide source/drain (S/D) for hot carrier injection (HCI) [5]. Based on parameters optimized by TCAD simulation, 2D device was fabricated and their operating speed and reliability were measured to verify the feasibility of 3D memory cells.

II. Device Design of 3D DRAM

A. Structure design

Fig. 2. summarizes the main process flow of the 3D CT DRAM devices. The key processes are a thin tunnel oxide, donut-type poly-Si channel and metal silicidation. We deposit O/N/O and poly-Si sequentially on the nitride-indented mold with holes. Following poly-Si etch results in the formation of the separated donut-type poly-Si channels in the recessed region. The process sequence used to form the poly-Si islands is already verified through the mass production of the floating-gate type 3D NAND devices from Intel and it is a highly reliable process [6-7]. Metal silicide is formed at the both ends of the poly-Si channel in the SL and BL sides. The proposed structure has strong advantages in power consumption and heat dissipation, which are the most important aspects for high-performance DRAM like HBM. Fig. 3. (a) shows its power advantage. It consumes less power than the conventional 1T1C DRAM cell because it does not lose charges during the read operation and does not need to be restored. The CT DRAM without capacitors can achieve extremely high integration density and metal S/Ds is very effective to dissipate heat from the underlying logic die as shown in Fig. 3 (b).

Fig. 4. (a) shows the cross-sectional device structure of the 3D CT DRAM and Fig 4. (b) shows its program operation. It is very hard for the poly-Si channel transistors to make hot electrons at the drain side because of many grain boundaries and trap sites. But if we utilize the SB at the source side, we can make hot carriers easily even with the poly-Si channel. The electrons injected through the SB gain additional energy and result in impact ionization by the steep E-field on the source side to form electron-hole pairs (EHPs).

B. Optimization of device parameters

We optimized the process/device parameters of the 3D CT DRAM with 3D TCAD simulation. Fig. 5. (a) and (b) show the simulated I_d-V_g curves and the energy band diagrams along with various SB heights, respectively. The larger SB height is, the steeper band bending occurs on the source side but on-state current is decreased because of decreased carrier tunneling. Considering the program characteristics and the on-current, SB of around 0.25 eV is good for the device. For T_{oxb} between 0.9 nm and 1.5 nm, the variation in threshold voltage (V_{th}) and electrostatic potential is small as shown in Fig. 5. (c), which means that the difference in vertical and lateral E-fields required for HCI is also negligible as shown in Fig. 5. (d).

Fig. 6. (a) shows the retention time as a function of the trap density of charge trap layer (CTL). A high trap density increases the number of electrons trapped in the CTL, resulting in a higher initial V_{th}. Fig. 6. (b) shows the retention time as a function of trap energy level. High trap energy level improves

retention time because a deep trap is formed, requiring high energy for electrons to escape from the trap. In Fig. 6. (c), we can see that the thicker T_{oxb} leads to an increase in the energy to escape to the channel beyond the tunnel oxide, which improves the retention time. T_{oxb} around 1 nm is good enough considering the proper retention characteristics for DRAM. Fig. 6. (d) summarizes the mechanism of electron leakage.

The process target values were set based on the simulation result, and the detailed parameters are shown in Table 1.

III. Device Fabrication

The feasibility of the 3D CT DRAM was verified by fabricating and measuring CT DRAM devices with planar structures as shown in Fig. 7.

A. Schottky barrier silicide formation

In this study we formed silicides by using Microwave annealing (MWA). MWA utilizes electromagnetic radiation to heat the materials, offering the advantage of achieving uniform volume heating [8]. The sheet resistance (R_s) and junction characteristics of the silicide were examined as a function of microwave power. Fig. 8 depicts the sheet resistance of Ni and Co-silicide according to MWA conditions. Overall, Ni-silicide exhibits lower R_s compared to Co-silicide. Fig. 9. (a) and (b) show the on/off current ratio and SB heights (ϕ_b) of the SB diodes, respectively. On/off currents were extracted from 5 V and averaging reverse current, respectively. ϕ_b was extracted using the following equation (1).

$$\boldsymbol{\phi_b} = \frac{kT}{q} \ln\left(\frac{A^{**}T^2}{J_0}\right) \quad (1)$$

Where q, k, T, A^{**}, and J_0 are the unit electron charge, Boltzmann's constant, absolute temperature, equivalent Richardson's constant, and reverse saturation current density, respectively. In this study, we applied Ni, which exhibits a relatively higher on/off current ratio, SB heights, and lower R_s. SB height can be further adjusted by optimizing the stoichiometry of Ni silicide [9-10].

B. Device integration

Fig. 10 illustrates the fabrication of the 2D CT DRAM device. Buried oxide and active poly-Si layers were formed on the substrate. After the active region was formed, N/O layers and n+ poly-Si gate were formed. Then, ON spacer was formed to isolate gate and S/D regions during following self-aligned silicidation process. Finally, nickel was deposited using an Electron Beam Evaporator, and MWA was performed to form a silicide with low contact resistance. The unreacted Ni was removed.

IV. Results and Discussions

Fig. 11. (a) shows the cross-sectional TEM image of the fabricated device. NiSi was formed successfully at the gate, source, and drain. The S/D regions were fully silicided and its thickness is 20 nm. The gate spacer of 15 nm can be further reduced to decrease the non-overlap length between the gate and the NiSi S/D. Fig. 11. (b) illustrates the thickness of the ONO layer, with the tunnel oxide, trap layer, and blocking oxide layers of 1, 2.3, and 4 nm, respectively.

Fig. 12 illustrates the I_d–V_g curves of the 2D CT DRAM devices utilizing MWA at different time and power levels. As mentioned earlier, we used 600 W/1 min of MWA condition considering the lateral growth of NiSi as well as on-current of the devices. Fig. 13. (a) compares the program characteristics of the device with a tunnel oxide thickness of 1 nm to that with a thickness of 2.9 nm. The device with T_{oxb} of 1 nm showed as fast program speed as V_{th} shift of 2 V even at the program time of 20 ns@V_{GS} = 10 V. Fig. 13. (b) shows the erase characteristics of the device as a function of erase time and voltage. The device with a T_{oxb} of 1 nm demonstrated a fast erase speed with a V_{th} change of approximately -0.8 V at an erase time of 20 ns@V_{GS} = -6.5 V. Fig. 14. shows the program characteristics utilizing hot carrier injection at a short program time of 20 ns and lower gate biases. When programming using FN tunneling without applying drain voltage, the threshold voltage shift was very small, approximately 0.5 V at a program voltage of 9 V. As the drain voltage increases, a larger E-field is generated at the source end, resulting in significantly faster program characteristics.

Fig. 15 shows the retention characteristics at 85°C of the 2D CT DRAM device with a thin tunnel oxide. Even with the tunnel oxide as thin as 1 nm, more than 50% of the initial Vth window was retained after 0.2 seconds, which is sufficient for DRAM applications.

Fig. 16 shows the endurance (>$5x10^5$ Cycles) of the fabricated device as a function of pulse cycles. The fabricated 2D CT DRAM cell shows a V_{th} window of 0.22 V, ensuring sufficient reliability after 10^{15} P/E cycles.

V. Conclusions

This study investigated the CT device, its potential as a next-generation 3D memory solution. We showed that CT DRAM was promising when it was combined with ultrathin tunnel oxide and HCI from metal silicide S/D. Moreover, the 1T CT DRAM is one of the most promising candidates for 3D DRAM due to its simple process integration and easy heat dissipation through the metal S/D.

Acknowledgment

The authors would like thank H.-C. Hwang/B.-R. Yoon at NNFC for helping the device fabrication.

References

[1] J. W. Han, et. al., IEEE VLSI Tech., 2023
[2] M. Suh, et. al., IEEE TED, 2024
[3] A. Belmonte, et. al., IEEE IEDM, 2021
[4] J. Kim, et. al., IEEE IEDM, 2023
[5] S. -J. Choi, et al., IEEE VLSI Tech., 2009
[6] A. Khakifirooz, et. al., IEEE ISSCC, 2023
[7] K. Parat, et. al., IEEE IEDM, 2015
[8] Min, J.G., et. al., Nanomaterials, 2022
[9] K. Takahashi, et al., IEEE IEDM, 2004
[10] A. Lauwers, et. al., IEEE IEDM, 2005

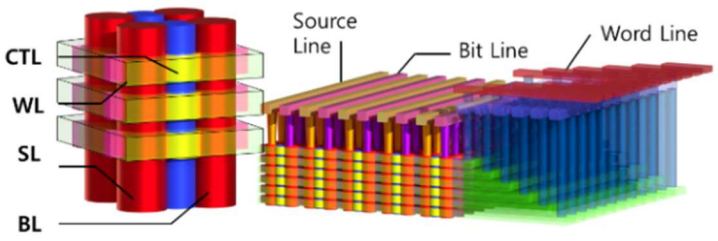


Fig. 1. Schematic illustration of the 3D CT DRAM

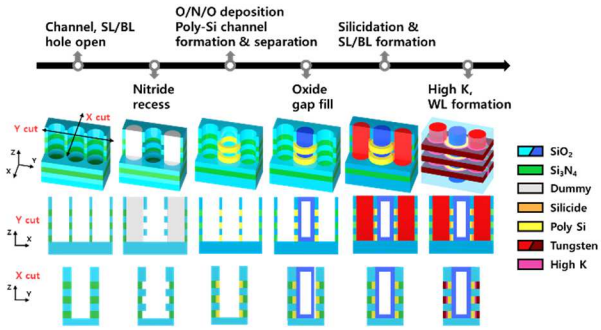


Fig. 2. Process integration of the 3D CT DRAM with SB S/D.

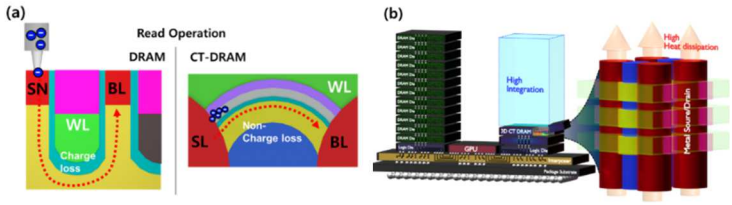


Fig. 3. Advantages of the CT DRAM compared to the conventional DRAM in view of (a) power and (b) 3D integration/heat dissipation.

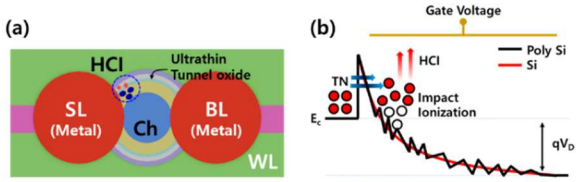


Fig. 4. (a) Cross-sectional view of the 3D CT DRAM with Schottky barrier (SB) S/D and ultrathin tunnel oxide (b) HCI at the source-side was used for the program operation at the device having a poly-Si channel.

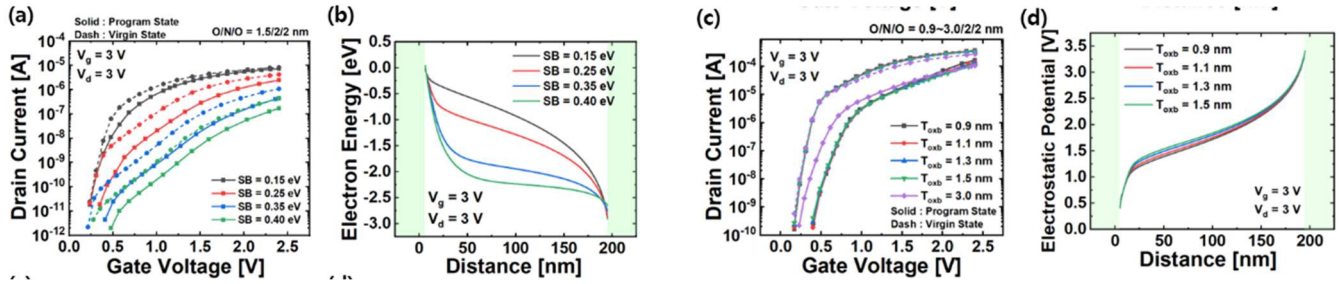


Fig. 5. I_d–V_g characteristics at different (a) SB (0.15 eV ~ 0.40 eV). The SB characteristics that vary with (b) programmed energy band. I_d–V_g characteristics at the different tunnel oxide thickness (0.9 nm ~ 1.5 nm) shown in (c). (d) The electrostatic potential hardly changes with tunnel oxide thickness.

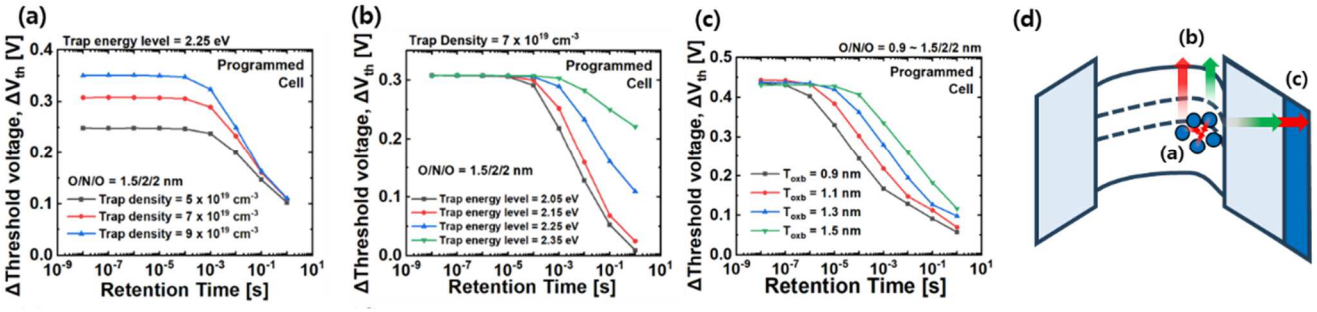


Fig. 6. The retention characteristics at (a) different trap density, (b) trap energy level, and (c) tunnel oxide thickness. (d) Mechanism of electron leakage.

Table 1. Optimized Device parameters for the 3D CT DRAM

Description (parameter)	Values
Blocking oxide thickness (T_{oxt})	2 nm
CTL(Si_3N_4) thickness (T_{CTL})	2 nm
Tunnel oxide thickness (T_{oxb})	1 nm
Channel length (L_g)	150~300 nm
Channel thickness (T_{ch})	7 nm
Schottky barrier height (SB)	0.25 eV
Nitride Trap density (N_{trap})	7×10^{19} cm^{-3}
Substrate doping concentration	Undoped
Gate work function	4.65 eV
Trap energy level	2.25 eV

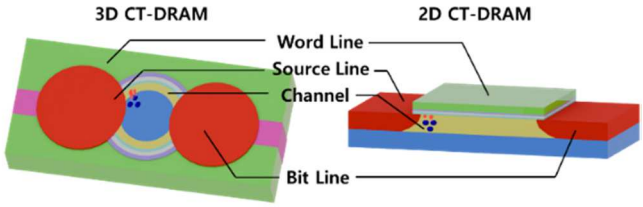


Fig. 7. Schematic of a planar CT DRAM to verify the feasibility of the 3D CT DRAM.

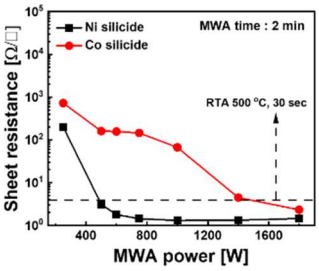


Fig. 8. Sheet resistance of Ni and Co silicide according to MWA power.

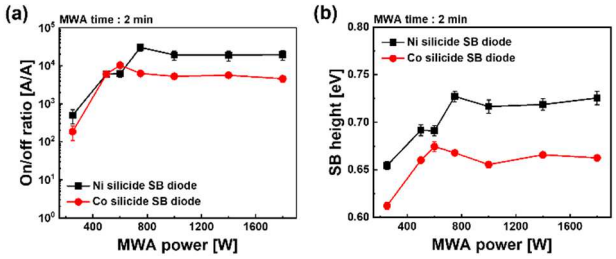


Fig. 9. (a) On/off current ratio, and (b) SB heights of Ni and Co-silicide SB diodes.

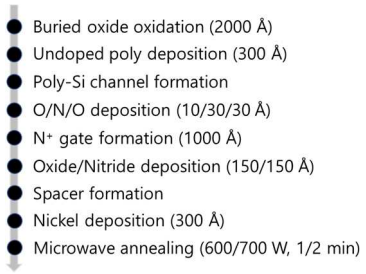


Fig. 10. Process integration of the 2D CT DRAM with SB S/D

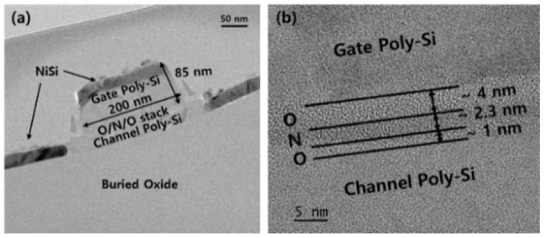


Fig. 11. TEM photographs of the fabricated 2D CT DRAM with SB S/D (a) Cross-sectional TEM (b) Magnified TEM of ONO layers.

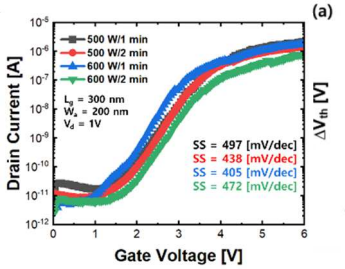


Fig. 12. Measured initial I-V characteristics of 2D CT DRAM device with L_g/W=300 nm/300 nm.

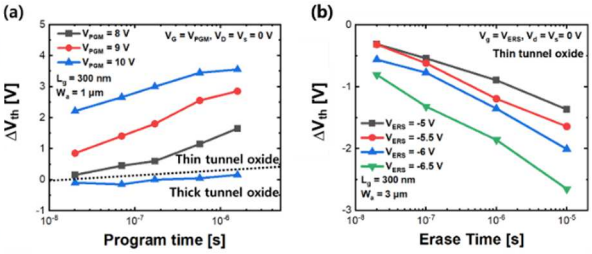


Fig. 13. Measured (a) program and (b) erase characteristics of the devices with thick and thin tunnel oxide.

\

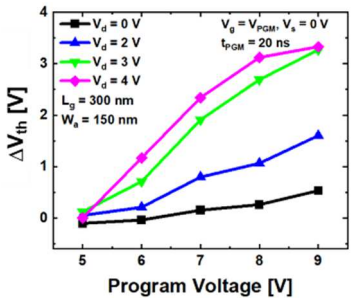


Fig. 14. Measured program operation characteristics using hot carrier injection with varying drain voltages.

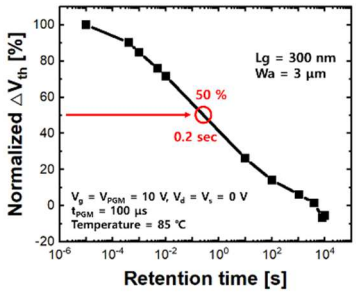


Fig. 15. Measured retention characteristics of the device with thin tunnel oxide at 85 ℃.

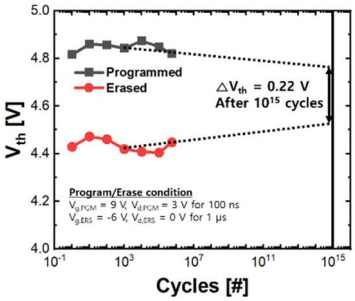


Fig. 16. Measured endurance characteristics of the device with thin tunnel oxide

Effects of Impurity Doping on Electrical Properties of Cubic Boron Nitride: A First-principles Study

Zirui He[1,2], Haojun Hu[1,2], Siqing Shen[2], Yongli Liang[2], Shang-Peng Gao[1,2,*], Hao Hu[2,†], and Meng Chen[2,‡]
[1]Department of Materials Science, Fudan University, Shanghai, China, email: *gaosp@fudan.edu.cn
[2]Advanced Silicon Technology Co., Ltd., Shanghai, China, emails: †hhu@ast.com.cn, ‡mchen@ast.com.cn

***Abstract*—**Cubic boron nitride (*c*-BN) is a promising III-V semiconductor with an ultra-wide band gap. Despite many experimental studies on dopants in *c*-BN, the reported results exhibit notable great discrepancy. We have studied several possible acceptors (Be, Mg, Zn, C, and Si) and donors (O, S, Se, C, and Si) with state-of-the-art first-principles calculations. We used a hybrid functional to overcome the band gap problem, and calculated the accurate ionization energy of each species. We carried out fully *ab initio* electron-phonon calculations to study the carrier mobilities with phonon and impurity scattering. Finally, we obtained quantitative results of conductivity and resistivity under different doping conditions and temperatures. This work can provide useful guidance for controlling the electrical properties of *c*-BN by doping.

I. Introduction

Cubic boron nitride (*c*-BN) is one of the most promising III-V semiconductors. It possesses high hardness, high thermal and chemical stability, as well as various exceptional physical properties, such as an ultra-wide band gap, low dielectric constant, high breakdown field, high Baliga figure of merit, wide range of transmittance wavelengths, and high thermal conductivity [1]. Therefore, *c*-BN can be applied in high-power and high-frequency electronic devices, deep-ultraviolet optoelectronics, quantum information, and extreme-environment sensors or electronics [2].

Doping is a common method to increase the carrier concentration in semiconductors. But on the other hand, doping also reduces the carrier mobility due to the ionized impurity scattering, which may become even more significant compared with the phonon scattering under heavy doping. Hence, both the carrier concentration and carrier mobility should be considered to evaluate the effect of doping on the electrical conductivity.

It has been experimentally demonstrated that both *n*-type doping and *p*-type doping are feasible for *c*-BN [1]. The ionization energies, which are important properties of dopants, have been measured for several impurities such as Be, Mg, Si, and S [1]. The mobilities of both carrier types have also been measured [1]. However, the reported results for ionization energies and mobilities are distinct within the literature. This discrepancy may stem from the differences in samples (type, quality, synthesis method, etc.) or in measurement techniques.

On the other hand, first-principles calculations based on density functional theory (DFT) can serve as effective and predictive tools. Nonetheless, it remains challenging to study these properties precisely. Standard DFT calculations underestimate the band gaps of semiconductors severely, which may introduce significant errors to the calculated ionization energy, as it depends on the relative position of the acceptor or donor level with respect to the band edge. For carrier mobility, semi-empirical models are still the most widely applied methods, which cannot be expected to produce quantitative results.

In this work, we adopt state-of-the-art computational methods to address these problems. A hybrid functional is applied, which can produce much more accurate band gaps, and thereby more reliable ionization energy. And a fully *ab initio* method, which considers all electron-phonon interaction and ionized impurity scattering, is used to calculate carrier mobility under different temperatures and doping concentrations. This work is expected to provide useful insights into the doping effect on the electrical properties of *c*-BN, and provides helpful guidance for its practical application.

II. Formation and Ionization Energies of Impurities in *c*-BN

The formation energy E_{form} of a defect in the charge state of q is given by [3]

$$E_{\text{form}}(q) = E^{\text{total}}_{\text{defect}}(q) - E^{\text{total}}_{\text{pristine}} - \Delta\mu + q(E_{\text{VBM}} + E_{\text{F}}) + E_{\text{corr}}, \quad (1)$$

where $E^{\text{total}}_{\text{defect}}$ and $E^{\text{total}}_{\text{pristine}}$ are the total energies of a supercell with the defect and a pristine supercell, respectively. $\Delta\mu$ is the change in the chemical potential due to atom removal or adding, E_{VBM} the energy of valence band maximum (VBM), E_{F} the Fermi energy, and E_{corr} the correction term to address the artificial interaction between the charged defect and its periodic images due to the supercell approximation.

The transition level is defined as the Fermi energy that makes $E_{\text{form}}(0) = E_{\text{form}}(q)$. And the ionization energy is the energy difference between the transition level and band edge.

The energy terms needed in (1) were obtained from DFT calculations using VASP [4], [5]. The HSE06 hybrid functional [6] was used to overcome the underestimation of the band gap by standard DFT calculations. $3 \times 3 \times 3$ cubic supercells were applied. And the correction terms for charged impurities were calculated based on the FNV scheme [7], [8].

We considered Be, Mg, and Zn as acceptors, and O, S, and Se as donors. The calculated ionization energies are given in Tab. I. The IIA and IIB elements studied (Be, Mg, and Zn) are good acceptors, while the VIA elements (O, S, Se) are good

donors. The results are generally in reasonable agreement with reported values [1]. And it can be found that the ionization energy gradually increases as the period number increases.

Meanwhile, we also considered two IVA elements, C and Si, as both acceptors and donors, depending on the lattice site they occupy (substituting N and substituting B, respectively). As listed in Tab. I, C can serve as a shallow acceptor and donor. However, Si can only be a shallow donor, as it introduces a rather deep acceptor level (0.74 eV) when substituting N. Additionally, the calculated formation energy indicates that Si on B is thermodynamically more stable than Si on N. For C, the relative stability of the two sites depends on the synthesis condition (B-rich or N-rich) and Fermi energy.

With the energy results, the partial ionization under different doping concentrations and temperatures was calculated according to the electronic structure of *c*-BN. Then the carrier concentrations under various conditions were obtained, as shown in Fig. 1 (a)-(e) for holes and Fig. 2 (a)-(e) for electrons. Under light doping, doped atoms can be full ionized. After a critical doping concentration, which depends on ionization energy and temperature, they can only be partially ionized.

III. Carrier Mobility and Conductivity under Different Doping Conditions and Temperatures

The electronic structures and phonon dispersions were calculated using Quantum ESPRESSO [9]–[11]. The exchange-correlation interaction was described by the PBEsol functional [12]. These results were adopted to calculations for electron-phonon interaction, as implemented in the EPW code [13]. By solving the Boltzmann transport equation iteratively, we first calculated the phonon-limited electron and hole mobilities at different temperatures, and then the mobilities with ionized impurity scattering under different doping concentrations. Partial ionization was considered as well.

At room temperature, the calculated phonon-limited drift and Hall mobilities of holes are 292 and 217 $cm^2V^{-1}s^{-1}$, respectively, and those of electrons are 985 and 998 $cm^2V^{-1}s^{-1}$, respectively. As temperature increases, the carrier mobilities decreases significantly, because the lattice vibration becomes more violent, inducing stronger electron-phonon scattering.

After doping, the interaction between ionized impurities and carriers becomes another scattering mechanism that reduces the carrier mobility, as shown in Fig. 1(f)−(o) for holes and Fig. 2(f)−(o) for electrons. The only exception among these dopants is Si (substituting N), whose mobility remains almost constant within the doping concentration range considered in this work [see Fig. 1(j)(o) and Fig. 2(j)(o)]. Its ionization energy is so large that only a very small portion of doped atoms can be ionized and contribute to carrier scattering.

For other dopants, to be specific, under light doping (e.g., doping concentration $< 10^{16}$ cm^{-3}), the carrier mobilities are almost identical to their intrinsic counterparts. Under moderate doping (e.g., $10^{16} - 10^{19}$ cm^{-3}), the mobilities starts to decrease, indicating non-negligible effect of impurity scattering. And under heavy doping (e.g., larger than 10^{19} cm^{-3}), the ionized impurity scattering can become the predominant scattering mechanism, and the carrier mobility may be several times lower than the undoped case. Therefore, it is necessary to exactly consider the dependence of carrier mobilities on both temperature and doping, especially for shallow acceptors or donors, and when the doping concentration is relatively high.

As mentioned before, both doping and temperature have two opposite effects on the electrical conductivity, by increasing the carrier concentration whilst decreasing the carrier mobility. Hence, to determine the conductivity (or resistivity) under various conditions, all these effects should be evaluated simultaneously at the quantitative level. Using the *ab initio* data we have obtained, we calculated the conductivity and resistivity under different doping concentrations and temperatures, as shown in Fig. 1(p)−(y) Fig. 2(p)−(y) for *p*-type *n*-type doping, respectively. Under light doping, the conductivity is higher at lower temperature. The reason is that, on the one hand temperature has little influence on carrier concentration in this case, as doped atoms can be fully ionized at any temperature (at least within the range considered here); on the other high temperature reduces the carrier mobility. But when the doping concentration is very high, higher temperature can promote the ionization dramatically. This effect is more significant than the reduction in carrier mobility. Consequently, the conductivity is lower at lower temperature under heavy doping. In addition, at constant temperature, the conductivity always increases with doping concentration, in all conditions considered here.

IV. Conclusions

We have studied five possible acceptors [Be, Mg, Zn, C (on N), and Si (on N)] and five possible donors [O, S, Se, C (on B), and Si (on B)] in *c*-BN. Using DFT calculations with the HSE06 exchange-correlation functional to overcome the band gap problem, we obtained the ionization energy of each species. All of them can serve as shallow acceptors (donors), expect Si (on N), which is a deep acceptor. By adopting the energy results to fully *ab initio* electron-phonon calculations, we obtained the carrier mobilities accurately under a series of doping concentrations and temperatures. Finally, the conductivity and resistivity under various doping conditions and temperatures were calculated quantitatively, with all effects considered precisely. This work elucidates the doping effect on the electrical properties of *c*-BN, and provides helpful guidance for its practical application.

References

[1] X. Zhang, *Thin Solid Films*, vol. 544, pp. 2–12, 2013.
[2] J. Y. Tsao *et al.*, *Adv. Electron. Mater.*, vol. 4, no. 1, p. 1600501, 2018.
[3] H.-P. Komsa *et al.*, *Phys. Rev. B*, vol. 86, p. 045112, 2012.
[4] G. Kresse *et al.*, *Comput. Mater. Sci.*, vol. 6, no. 1, pp. 15–50, 1996.
[5] ——, *Phys. Rev. B*, vol. 54, pp. 11 169–11 186, 1996.
[6] A. V. Krukau *et al.*, *J. Chem. Phys.*, vol. 125, no. 22, p. 224106, 2006.
[7] C. Freysoldt *et al.*, *Phys. Rev. Lett.*, vol. 102, p. 016402, 2009.
[8] ——, *Phys. Status Solidi B*, vol. 248, no. 5, pp. 1067–1076, 2011.
[9] P. Giannozzi *et al.*, *J. Phys.: Condens. Matter*, vol. 21, no. 39, p. 395502, 2009.
[10] ——, *J. Phys.: Condens. Matter*, vol. 29, no. 46, p. 465901, 2017.
[11] ——, *J. Chem. Phys.*, vol. 152, no. 15, p. 154105, 2020.
[12] J. P. Perdew *et al.*, *Phys. Rev. Lett.*, vol. 100, p. 136406, 2008.
[13] H. Lee *et al.*, *npj Comput. Mater.*, vol. 9, no. 1, p. 156, 2023.

Dopant (p-type)	Acceptor ionization energy (eV)	Dopant (n-type)	Donor ionization energy (eV)
Be (on B)	0.23	O (on N)	0.15
Mg (on B)	0.29	S (on N)	0.22
Zn (on B)	0.33	Se (on N)	0.24
C (on N)	0.26	C (on B)	0.12
Si (on N)	0.74	Si (on B)	0.26

TABLE I

THE CALCULATED IONIZATION ENERGIES OF DIFFERENT DOPANTS.

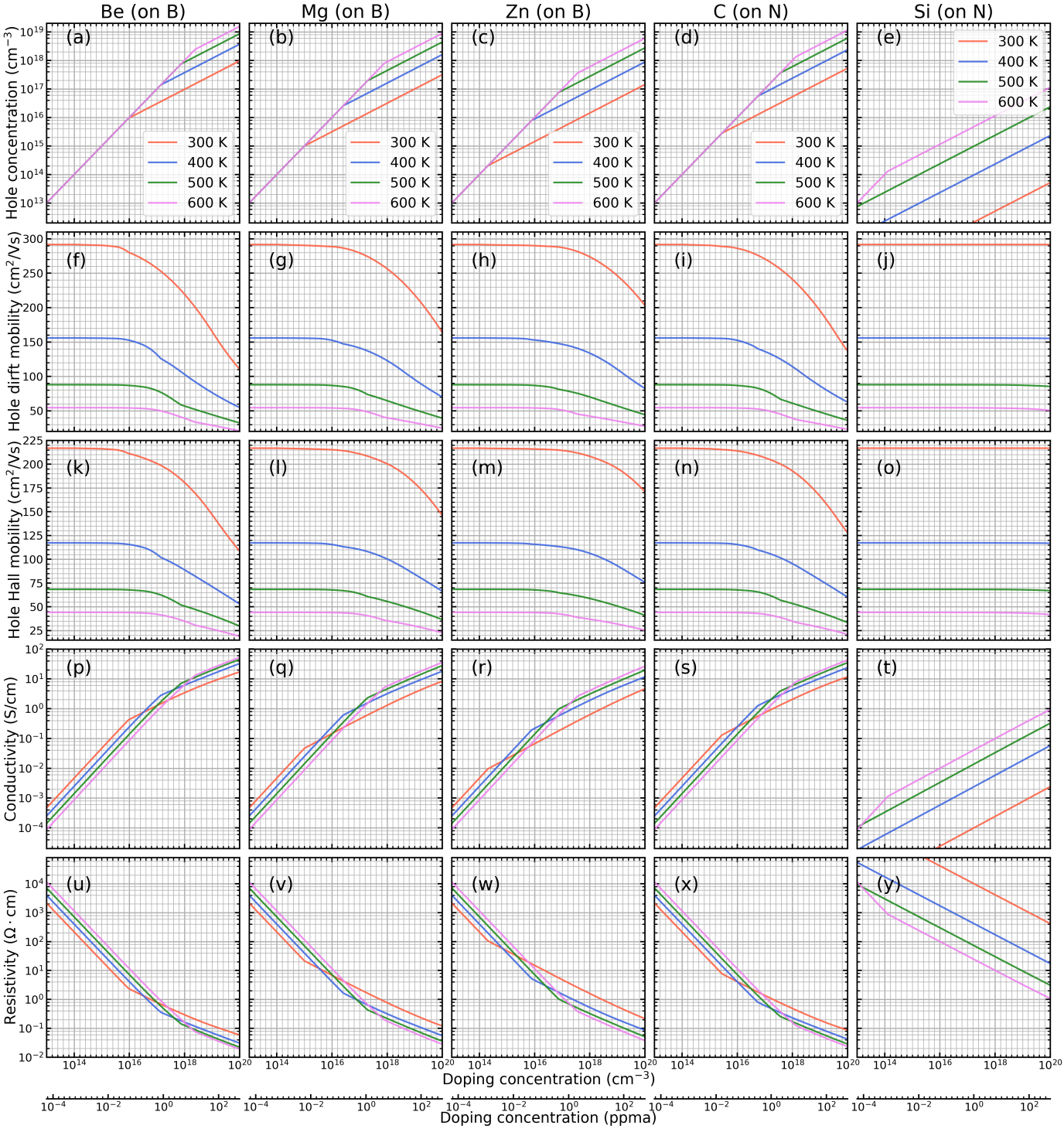


Fig. 1. Electrical properties of c-BN with different p-type dopants, under various doping concentrations ($10^{13} - 10^{20}$ cm^{-3}) and temperatures (300−600 K). (a)-(e): hole concentration. (f)-(j): hole drift mobility. (k)-(o): hole Hall mobility. (p)-(t): conductivity. (u)-(y): resistivity.

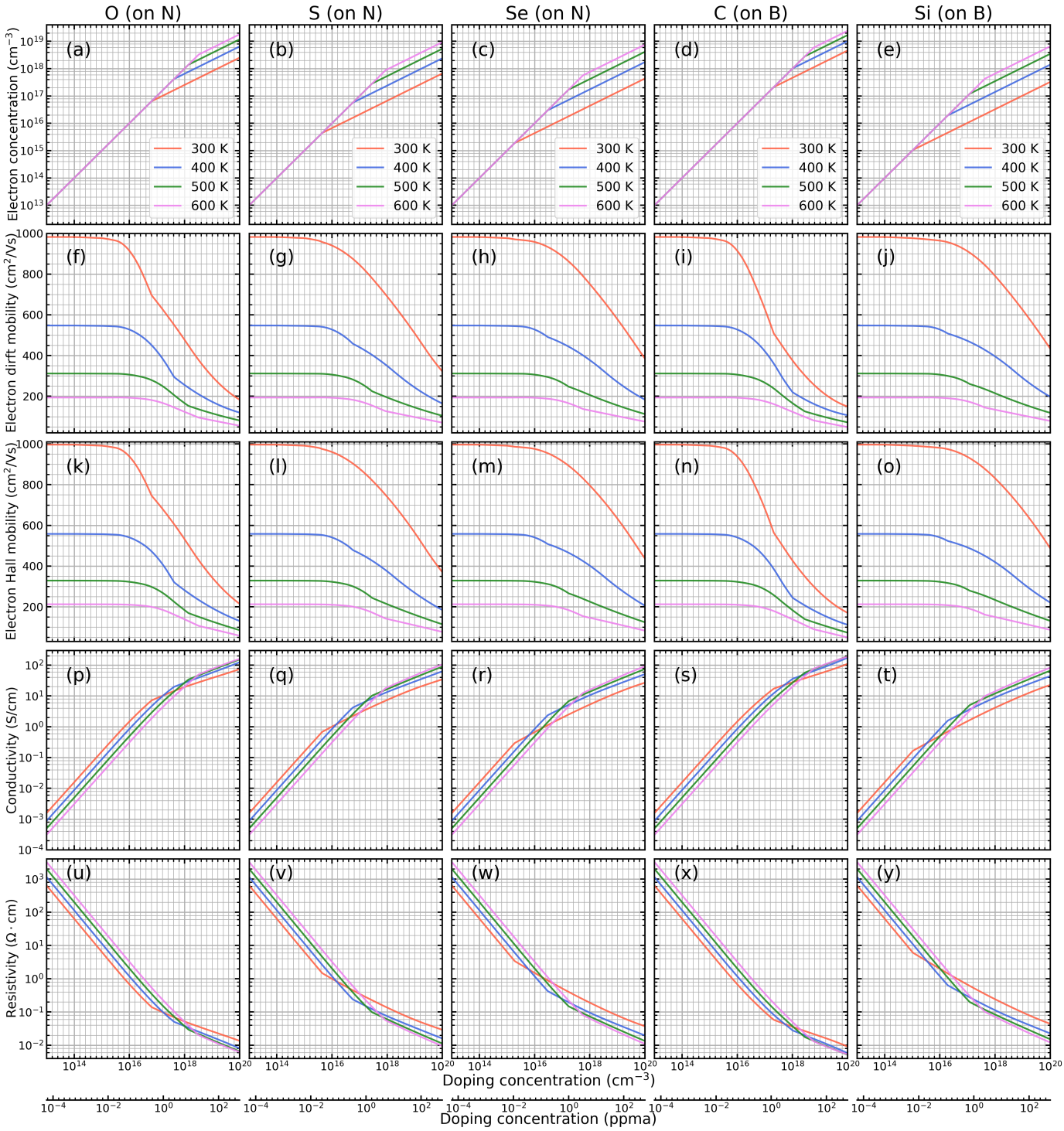


Fig. 2. Electrical properties of c-BN with different n-type dopants, under various doping concentrations ($10^{13} - 10^{20}$ cm^{-3}) and temperatures (300−600 K). (a)-(e): electron concentration. (f)-(j): electron drift mobility. (k)-(o): electron Hall mobility. (p)-(t): conductivity. (u)-(y): resistivity.