First Demonstration of Ultra-low D_{it} Top-Gated Ferroelectric Oxide-Semiconductor Memtransistor with Record Performance by Channel Defect Self-Compensation Effect for BEOL-Compatible Non-Volatile Logic Switch

Chun-Kuei Chen†, Zihang Fang†, Sonu Hooda†, Manohar Lal, Umesh Chand, Zefeng Xu, Jieming Pan, Shih-Hao Tsai, Evgeny Zamburg, and Aaron Voon-Yew Thean*
Department of Electrical and Computer Engineering, National University of Singapore (NUS), Singapore 117583
†Equal contribution; Email: Aaron.Thean@nus.edu.sg

2022 IEEE International Electron Devices Meeting (IEDM) | 978-1-6654-8959-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEDM45625.2022.10019440

***Abstract*—**We demonstrate, for the first time, a short-channel (L_G:40nm) back-end-of-line (BEOL) compatible top-gated (TG) self-aligned FeFETs with the ultra-low interface/bulk trap density (D_{it}/D_{bulk}) down to $10^{11}cm^{-2}eV^{-1}$, a 100x improvement over conventional amorphous Indium-Gallium-Zinc-Oxide (IGZO) devices. High memory and drive performance are both achieved, exhibiting a large and stable memory window of 2.1V, excellent endurance exceeding 10^7 cycles, close-to-ideal subthreshold swing (*S.S.*) of 62mV/dec., and the record-low read-after-write delay of 200ns. This is accomplished by utilizing the defect self-compensation effect in the ITO-IGZO heterojunction channels for ferroelectric top-gate stack stabilization. We leverage these advantages and proposed a novel Monolithic 3D (M3D) FPGA architecture with the demonstrated short-channel (L_G:40nm) BEOL dual-gated (DG) merged memory-logic FeFETs with excellent drive performance as a non-volatile reconfigurable interconnect switch. Our BEOL-compatible DG FeFET switch enables a compact interconnect switch fabrics with a V/2 bias scheme, featuring excellent G_{on}/G_{off} of 10^6, ultra-low sub-pA leakage, disturb-free, and sneak-current-free read-write operation. This work sets new oxide-semiconductor FeFET performance records useful for future BEOL non-volatile logic applications.

I. Introduction

Low-thermal budget oxide-semiconductor-based logic and memory FETs, such as amorphous Indium-Gallium-Zinc-Oxide (IGZO) [1], Indium Oxide (In_2O_3) [2], and W-doped Indium Oxide (IWO) [3] are promising candidates as technology enablers of future embedded non-volatile memory for monolithic 3D (M3D)-integrated systems-on-chip and data-abundant IoT-edge applications. This is motivated in part by FeFETs' low-cost process, fast write/read operation speed, and low standby power [4]. To date, most of performant back-end-of-line (BEOL)-compatible oxide FeFETs are bottom-gated (BG) structures **(Fig. 1)**, in which the oversized gate design results in extra parasitic gate-channel-S/D overlap capacitance, and the non-self-aligned gate process can cause the problem of current drive loss due to source carrier injection energy barrier variation [5]. Therefore, there are still numerous significant technological advantages to develop low-thermal-budget top-gated (TG) oxide FeFETs, to leverage the self-aligned gate-S/D patterning capability for easing aggressive transistor feature-size and circuit density scaling [6]. Furthermore, top-gated FeFETs can further enable non-planar 3D transistor structures; boosting memory and driving performance for non-volatile logic applications [7]. Unlike Silicon (Si) technologies, stabilizing a performant low-thermal-budget TG oxide FeFET still faces great hindrances. This is due to the following fundamental material challenges: (1) The yet unclear ferroelectric poly-crystallization mechanism for metal-oxide (E.g. High-k) gate dielectric deposited on an amorphous oxide-semiconductor channel (I.e. top-gated) under low temperature; (2) The high sensitivity of excess interface/bulk trap formation in oxide-semiconductor as a function of gate dielectric process; (3) The notoriously difficult process window to optimize between trap passivation and defect-induced channel carrier concentration in oxide-semiconductor by post-deposition annealing **(Fig. 2)** [8].

Here, we overcome the above challenges by applying engineered defect self-compensation effect of oxide-channel [9] that significantly passivates the interface/bulk traps for top-interface FeFETs by a tuned ultra-thin 7nm $InSnO_x/InGaZnO_x$ (ITO-IGZO) heterojunction channel. With such a method, we demonstrate for the first time that top-interface ferroelectricity memory and transistor drive performance can be achieved in single-gated TG and double-gated (DG) structures under a low thermal budget ($< 400^\circ$ C). Table I benchmarks our TG and DG devices with the most competitive ferroelectric memory devices reported to date. Our devices show excellent memory properties and a significant enhancement in drive, and electrostatic performance, setting a new record.

II. Fabrication of Top-Interface TG/DG FeFETs

Fig. 3(a) depicts the process flow for TG and DG device fabrication. To study the role of ITO film in the heterojunction channel, the control group of top-interface FeFETs with an IGZO-only channel is also fabricated. Our short-channel TG/DG device (L_G:40nm) with a thin-body hetero-structure channel (~7nm) is confirmed by top-view SEM and cross-section STEM images **(Fig. 3(b) and (c))**. Our DG ferroelectric switch is uniquely designed for memory and high-performance transistor drive. Both top and bottom channels are enhanced by heterojunction channel (ITO-IGZO) engineering. The HZO-based top gate stack is designated to be the memory element, while the bottom gate stack is HfO_2, meant for field-effect high-k gate modulation of the bottom channel conduction **(Fig. 3(d))**.

III. Defects Self-compensation on TG FeFETs

We start by investigating the defect self-compensation effect in TG IGZO and TG ITO-IGZO FeFETs. High-resolution TEM images reveal ferroelectric HZO crystallization on both channel types **(Fig. 5a))**. GI-XRD spectroscopy confirms ferroelectric o(111) and o(200) phases can be formed in both devices **(Fig. 5(b))**. Nevertheless, it is found that the TG ITO-IGZO FeFETs appear to have a more stable MW while TG IGZO FeFETs show a hysteresis-free I_D-V_G transfer curve **(Fig. 5(c))**. It is found that significant interfacial electron trapping/de-trapping undesirably offset the desired ferroelectric anticlockwise hysteretic I-V response in the TG IGZO devices, leading to complete loss of memory window [10]. We further extract the trap densities for both TG devices, using the high-low-frequency capacitance method **(Fig. 5(d))**. The flat band voltage for IGZO is defined based on the suggested method in ref. 11. The extracted

results show that the deep-level trap density is dramatically reduced by nearly 100x for TG ITO-IGZO devices **(Fig. 5(e))**.

The XPS O_{1s} spectra give the information that the peak intensity referring to oxygen vacancy at the binding energy of ~530.5 eV reduces in the as-deposited ITO-IGZO heterojunction **(Fig. 5(f))**. This has been found to be a predominately defect self-compensation effect, in which the interstitial oxygen defects can be effectively suppressed by introducing metal ions with higher bond dissociation energy (BDE) [9]. In the ITO-IGZO heterojunction system, the BDE of Sn-O in ITO is at 528kJ/mol, which is prominently higher than those of the In-O (320kJ/mol) [12]. Therefore, the Sn^{2+} ions act as an efficient oxygen vacancy and the interstitial O-defects suppressor that subsequently improves *S.S.* and electron mobility of top-interface FeFETs **(Fig. 11)**.

IV. High-performance BEOL TG/DG FeFETs

Our TG ITO-IGZO FeFETs achieve excellent electrical characteristics including a highly stabilized MW of 0.57V, a high μ_{eff} of 57cm^2/Vs, a near-ideal *S.S.* of 62mV/dec., and I_{on}/I_{off} ratio of 10^7 **(Fig. 6(a) and (b))**. Our TG devices also show excellent endurance exceeding 10^7 cycles and memory retention over 10 years **(Fig. 6(c) and (d))**. Notably, the high I_{on}/I_{off} ratio and μ_{eff} can be sustained in AC operation over 10^6 cycles **(Fig. 6(f))**.

We further demonstrated ITO-IGZO short channel BEOL DG FeFETs. The top-channel FeFETs exhibit a large MW of 2.1V with minimal device-to-device variation, a high I_{on}/I_{off} ratio of 10^6, and an I_{on} of 1.2μA/μm **(Fig. 7(a))**. The bottom-channel FETs exhibit a hysteresis-free transfer curve and a higher I_{on} of 10μA/μm **(Fig. 7(b))**. The I_{on}/I_{off} ratio as a function of the memory writing pulse amplitude and pulse width is shown in **(Fig. 7(d))**. The drive current of the device is highly boosted over 125μA/μm **(Fig. 7(e))**.

The read-after-write delay test is conducted to investigate the carrier trapping/de-trapping in the DG top-interface FeFETs. **Fig. 8(a)** illustrates the applied pulse scheme during write operation with the ultra-fast I-V measurement tool, in which the pulse time step can be reduced to 10ns. There is no observation of programming read-after-write delay in ITO-IGZO FeFETs with pulse width from 100μs to 1ms, and pulse amplitude from 2 to 4.5V **(Fig. 8(b) and (c))**. With the continuous inspection of the channel current change right after the programming operation **(Fig. 9(a))**, the record-low read-after-write delay of 200ns is observed **(Fig. 9(b))**. On the other hand, the top-interface IGZO FeFETs exhibit a serious delay ≥ 10s **(Fig. 10(a))**, agreeing with the extracted result of higher interface/bulk trap density. For the erasing operation, the delay is not significant for both channel configurations **(Fig. 10(c))**. This is explained by the extremely low hole mobility in IGZO due to a large amount of near-valence band trap states [8].

It should be noted that oxygen vacancies and interstitial oxygen atoms are the main contributors of traps in HfO_2-based materials, which cause the poor read-after-write performance for FeFETs [13]. This work shows that the self-compensation effect is a very efficient passivation approach for the deep-level traps; mitigating the serious read-after-write delay issues in the ultra-scaled top-interface FeFETs **(Fig. 11)**.

V. BEOL DG Reconfigurable Interconnects

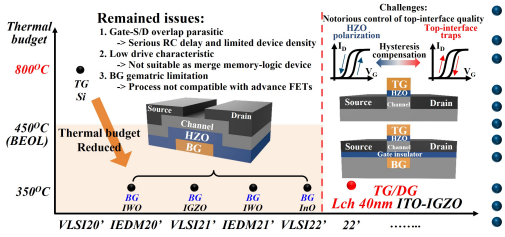
To illustrate the novel system application possibility with BEOL TG/DG FeFETs, we investigate, by computer modeling, a compact M3D-enabled FPGA architecture using non-volatile reconfigurable interconnect switch fabric, replacing conventional SRAM+MUX switches. Compared to conventional 2D FPGA architecture, in which 50% of its area is occupied by global routing resources, our proposed M3D FPGA's footprint is mainly reserved for configurable logic blocks (CLB) with monolithically stacked connection boxes (CB) and switch boxes (SB) **(Fig. 12)**. With the merger of memory and logic functionality in DG structure, our DG FeFETs function as non-volatile logic switches can be configured as: 1) Memory-on/BG-on, 2). Memory-on/BG-off, 3). Memory-off/BG-on, 4). Memory-off/BG-off **(Fig. 4)**. The memory state can be programmed through top-channel FeFET while the bottom-channel FET is responsible for the main control over the switch. This allows the ferroelectric-encoded memory state to stay disturbance free during FET switching operation.

Our DG interconnect shows excellent reconfigurability with G_{on}/G_{off} of 10^6, in which the memory-off and BG-off states achieve G_{off} of 10^{-12}S **(Fig. 14 (c))**. We next characterize our DG interconnects in the switch array based on the V/2 bias scheme **(Fig. 14 (d))**. The selected DG switch can be perfectly programmed/erased to on/off-state while half-selected and other switches are free of write disturb **(Fig. 14 (e))**. Such excellent performance also applies to the erase operation. In contrast, two-terminal unselected RRAMs without a selector or select transitors tend to contribute to sneak current paths in array circuits. This results from RRAM's significantly lower on-off current ratio as compared to a 3-terminal transistors. Hence, RRAM system designs need to choose to prioritize the minimization of off-state leakage or minimization of on-resistance. Overcoming such limitation, our DG transistor can eliminate sneak path leakages due to the significantly higher on-off current ratio. Moreover, gate leakage between bottom-top electrodes and the semiconductor channel, during memory read/write operation, is negligible because of the high-quality gate dielectric barriers **(Fig. 15)**. With an independent bottom gate switch operation, our DG switches eliminated the need for the select transitors, further reducing the cell area by 80% as compared with the 1T-1FeFET scheme [14], empowering our architecture with higher routing capability.

We benchmark the proposed M3D FPGA against a baseline 2D FPGA architecture [15] utilizing the same logic block design and 45nm CMOS technology. The P&R tool is customized for routing of the M3D layer [16]. Across the listed MCNC benchmark tests, our method shows a 62% area reduction compared to the baseline model. Reduced tile area and routing track length can save energy consumption by 38% and reduce critical path delay by 31% **(Fig. 16)**. By benchmarking other reported switches, we show that the combination of a high-performance DG memtransistor architecture and system solution, can lead to a significant system performance-area advantage over more aggressive technology nodes (**Benchmarking Table II**).

VI. Conclusion

By exploring the defect self-compensation effect, we demonstrate for the first time TG BEOL FeFETs with an ultra-low D_{it} of $10^{11}cm^{-2}eV^{-1}$, achieving a highly stabilized large MW of 2.1V, a near-ideal *S.S.* of 62mV/dec., and a record-low read-after-write delay of 200ns. Such excellent transistor and memory performances exceeded the best-in-class FeFETs, reported to date. we demonstrate here, different from post-passivation treatment, a novel approach effectively and intrinsically passivate the Dits for an ultra-scaled BEOL oxide FeFET, making it compatible with the essential top-gated process for future M3D integration of dense memory and non-volatile logic applications.



- Ultra-low Dits~10^{11}cm^{-2}eV^{-1} in low-thermal-budget TG oxide FeFETs
- Record-low read-after-write delay <200ns
- Nearly-ideal *S.S.*~ 62mV/dec. @ L_G:10μm Average low *S.S.*~ 68mV/dec. @ L_G:40nm
- Highest μ_{eff} of 57cm^2/Vs
- High memory drive current >125 μA/μm
- Endurance >10^7cycles; retention >10years
- Sub-pA gate/substrate leakage
- High-performance DG FeFETs as BEOL reconfigurable non-volatile switch
- Excellent switch ratio with G_{on}/G_{off} >10^6
- Free of read/write disturbance
- Free of sneak path current (<pA)

Fig.1. The evolution of BEOL FeFETs with BG geometric limitation and highlight of this work. We demonstrate the first top-interface BEOL TG/DG FeFETs by addressing the notorious control of top-interface quality challenges. High drive and memory performance are achieved by intrinsically passivating the bulk/interface D_{it} down to 10^{11}cm^{-2}eV^{-1}.

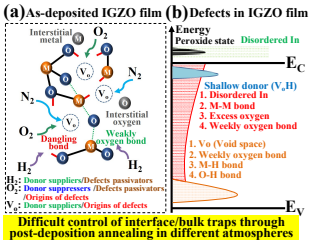


Fig. 2. (a) The roles of different atmospheres during post-annealing treatments, and (b) the various types of intrinsic defects in IGZO film [8].

	This work	2021 VLSI [1]	2022 VLSI [2]	2021 IEDM [3]	2021 VLSI [17]	2021 IEDM [18]
Structure	TG/DG	BG	BG	BG	1T1FeRAM	TG
Channel	ITO-IGZO	IGZO	In_2O_3	IWO	IGZTO	Si
Mobility (cm^2/Vs)	57	N.A.	N.A.	N.A.	20	N.A.
S.S. (mV/dec.)	62	85	N.A.	N.A.	72	N.A.
Normally off	Yes	Yes	No	N.A.	Yes	Yes
Retention	100s (I_{on}/I_{off})	1.7V	N.A.	0.3V	N.A.	N.A.
Endurance (#)	10^7	10^8	10^9	10^{10}	10^7	N.A.
MW (V)	2.1V	2.8V	2.5V	0.7V	10μC/cm^2	0.4V
Read-after-write delay (s)	200ns	N.A.	100ms	500ns	N.A.	400ns

Benchmarking Table I: Enhanced drive performance of TG/DG FeFETs as compared with reported Si-based and BEOL FeFETs.

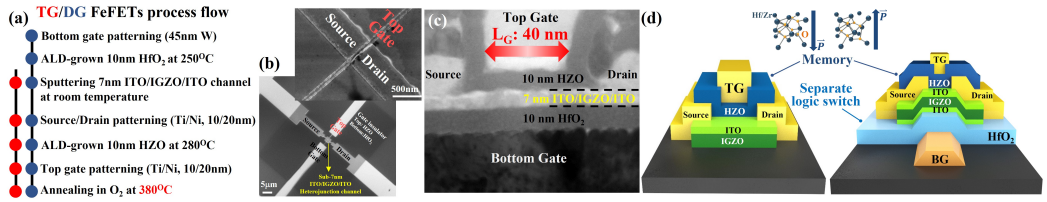


Fig. 3. (a) Key process flow of the fabricated TG/DG memtransistor. (b) Top-view SEM and (c) cross-sectional TEM image of our 40nm L_G TG/DG memtransistor integrated with ultra-thin 7nm heterojunction channel. (d) Device structures of TG/DG FeFETs. The HZO-based ferroelectric memory components are designed on the top-interface in both devices. Bottom-channel of DG device with HfO_2 meant for field-effect high-k gate modulation of channel conduction, as for a separated logic switch in merge memory-logic device.

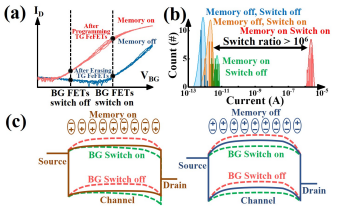


Fig. 4. (a) The 4 operation modes with (b) memory conductance distributions. (c) Illustrated 4-modes source-to-drain band diagram of the DG FeFETs as a non-volatile reconfigurable switch.

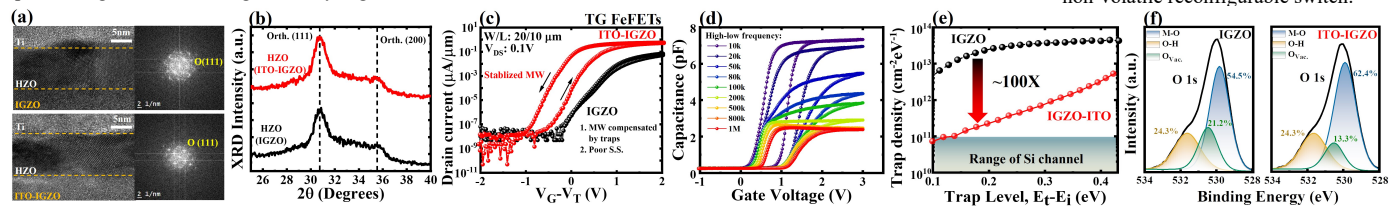


Fig. 5. (a) High-resolution TEM images of low thermal budget TG ITO-IGZO and IGZO FeFETs. (b) GI-XRD analysis shows HZO ferroelectric crystalline peaks for both channel configurations, indicating a successful HZO ferroelectric crystallization in both devices. (c) Measured I_D-V_G of TG ITO-IGZO and IGZO FeFETs. A stabler MW is only observed within the ITO-IGZO heterojunction channel device. (d) The measure high-low frequency C-V curve for D_{it} extraction, and (d) the extracted results. The deep-level interface/bulk traps can be effectively passivated down ~100x in the ITO-IGZO channel. (f) XPS analysis for O_{1s} binding energy of ITO-IGZO and IGZO channel. The intensity of oxygen vacancy is efficiently suppressed in the as-deposited ITO-IGZO channel.

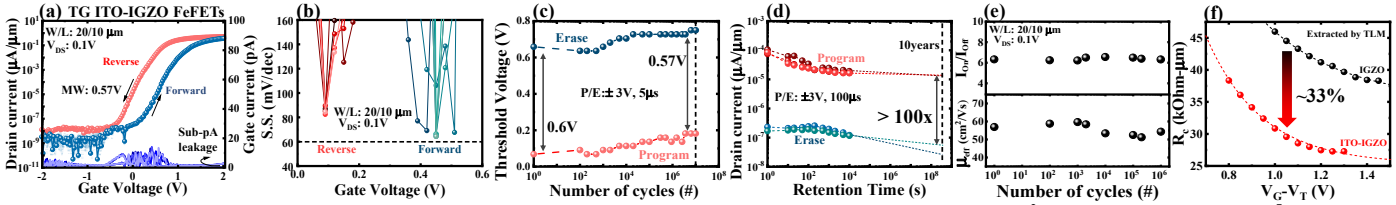


Fig. 6. (a) Measured I_D-V_G of TG ITO-IGZO FeFETs exhibiting extremely stable MW of 0.57V, high μ_{eff} of 57cm^2/Vs, high I_{on}/I_{off} ratio of 10^7, and sub-pA gate/substrate leakage. (b) The near-ideal *S.S.* distribution of TG ITO-IGZO FeFETs. (c) High memory read-write endurance exceeding 10^7 cycles and (d) memory retention > 10 years of TG ITO-IGZO FeFETs. (e) Extremely stable I_{on}/I_{off} ratio and μ_{eff} during AC operation exceeding 10^6 cycles. (f) The extracted contact resistance (R_C) for both channel FeFETs by TLM. R_C is reduced by 33% in the ITO-IGZO channel.

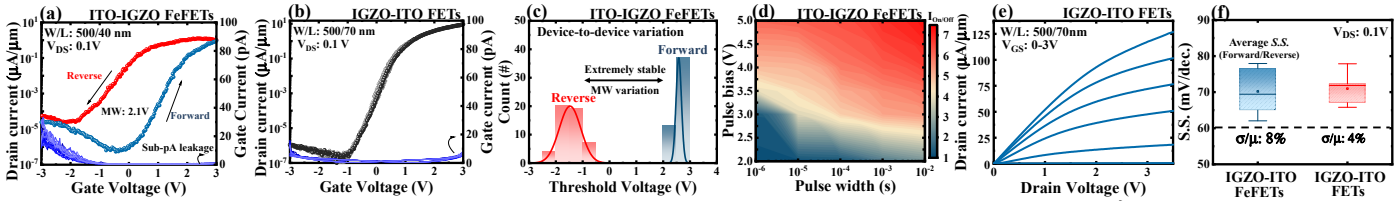


Fig. 7. (a) Measured I_D-V_G of DG top-channel FeFETs (memory) showing a stabilized large MW of 2.1V, a high I_{on}/I_{off} ratio of 10^6. and (b) measured I_D-V_G of DG bottom-channel FETs (switch) showing hysteresis-free loops, high I_{on} of 10μA/μm, and a higher I_{on}/I_{off} ratio of 10^7. (c) Measured small device-to-device MW variation, and (d) the I_{on}/I_{off} ratio as a function of memory programming pulse amplitude and pulse width of the DG top-channel FeFETs. (e) Measured I_D-V_D curve and (f) the S.S. distribution of DG devices, showing highly boosted drive current over 125μA/μm, and near-ideal *S.S.* value.

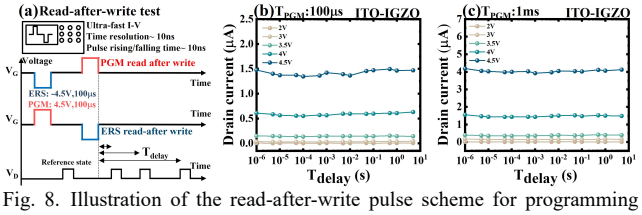


Fig. 8. Illustration of the read-after-write pulse scheme for programming and erase operation by using the ultra-fast I-V measurement tool. (b) The measured I_{PGM} versus T_{delay} with the programming pulse width of 100μs and (c) 1ms for top-interface ITO-IGZO FeFETs. No observation of after-after-read delay with the pulse amplitude from 2 to 4.5V.

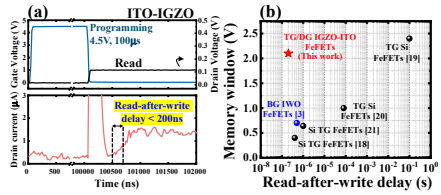
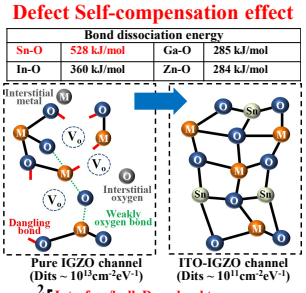


Fig. 9. (a) On-time inspection of the channel current change right after programming operation. (b) Record-low delay time <200ns with the largest MW was achieved as compared with the most competitive FeFETs reported to date.



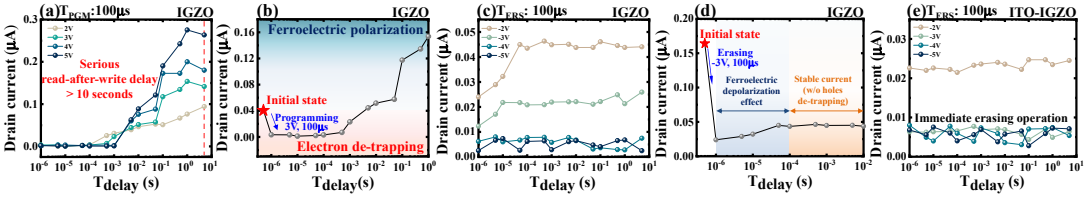


Fig. 10. (a) The I_{PGM} read-after-delay test on top-interface IGZO FeFETs, and (b) the change of I_{PGM} after programming operation. A serious delay >10s is observed. The initial I_{PGM} dropping is due to the electrons de-trapping. After 10ms, the HZO polarization dominates the increased channel current. (c) Measured I_{ERS} on top-interface IGZO FeFETs. (d) After the erasing operation, I_{ERS} initially increases due to the depolarization effect. Thereafter, I_{ERS} reaches a stable state. (e) The measured I_{ERS} on top-interface ITO-IGZO FeFETs. Holes trapping behavior is not observed due to the intrinsic low hole mobility in the IGZO channel.

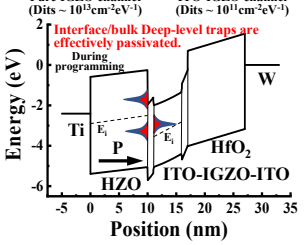


Fig. 11. The mechanism of self-compensation effect and simulated band diagram, showing passivated deep-level interface/bulk trap for top-interface FeFETs.

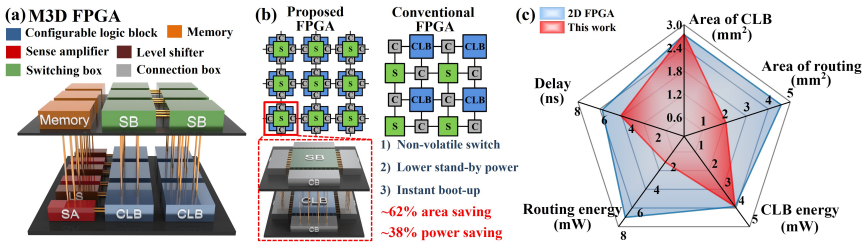


Fig. 12. (a) Proposed novel M3D FPGA with (b) BEOL non-volatile SBs/CBs. stacked on FEOL CLBs for higher area efficiency. (c) Performance comparison of the proposed M3D FPGA architecture against conventional 2D FPGA [15]. Monolithic stacked routing resources significantly reduce delay and routing energy consumption.

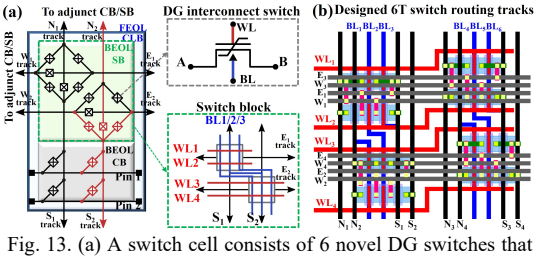


Fig. 13. (a) A switch cell consists of 6 novel DG switches that are connected to CLB IO pins 1 and 2 in CBs. Here, 2 WLs/3 BLs are shared within the cell to program while BLs are also shared among the cells for global switch control. (b) Switch cell array layout based on our DG FeFETs.

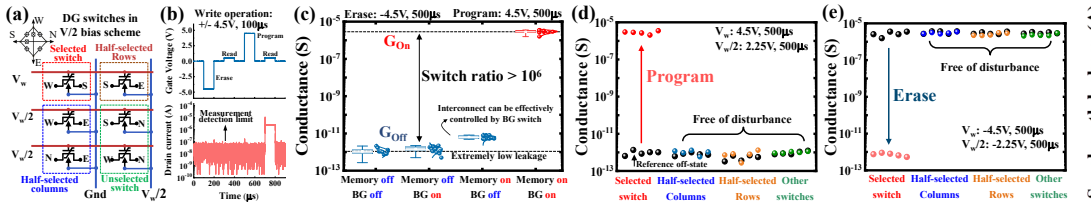


Fig. 14. (a) The illustrated write operation for the DG interconnect array based on the V/2 bias scheme. (b) The measured transient current of the DG interconnect showing effective write operation with pulse: +/- 4.5V, 100μs. (c) The measured channel conductance of DG switches based on four operation modes. The DG switch is turned on and off by applying V_{BG} at 0V, and 0.5V, respectively. (d) The write disturbance measurements for programming operation and (e) erasing operation.

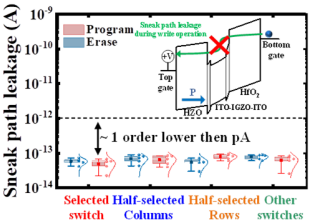


Fig. 15. Measured sneak path current of DG switches, which is intrinsically low thanks to very low electrons path through possibility.

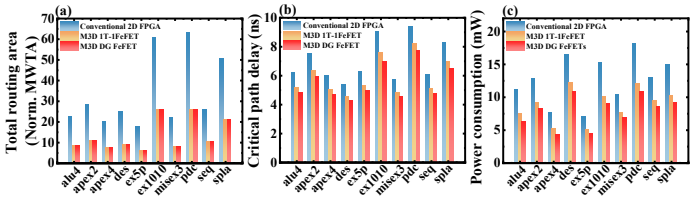


Fig. 16. Benchmark of the proposed M3D FPGA architecture using DG FeFETs as interconnect switch compared to conventional 2D FPGA [15] and M3D FPGA architecture using 1T1FeFET switch [14]. (a) The area, (b) critical path delay, and (c) power consumption are evaluated on the MCNC benchmark under 45nm technology node. (MWTA: minimum-width transistor area)

	M3D DG devices	1R [22]	1T1R [22]	1T1FeFET [14]	SRAM+NMOS [15]
Interconnect structure					
Design rule	45nm	32nm	32nm	22nm	7nm
Cell size (6 switches)	0.39 μm² (192F²)	0.01 μm² (12F²)	0.22 μm² (216F²)	0.46 μm² (950F²)	0.25 μm² (5000F²)
Sneak path leakage	No	Yes	No	No	No
Signal disturbance	No	Yes	Yes	No	No
Ron/Roff	10^6	10^6	10^6	<10^1	10^6

Benchmarking Table II: Comparison of our DG switches with other reported switches. Our DGs switches show significant technology advantages over other potential solutions.

Referances

[1] C. Sun et al., VLSI, 2021. [2] Z. Lin et al., VLSI. 2021. [3] S. Dutta et al., IEDM, 2021. [4] K. A. Aabrar et al., IEDM, 2021. [5] H-SP. Wong, IBM J. Res. Dev., 2002 [6] K. Han et al., IEEE EDTM 2021. [7] D. Matsubayashi et al., IEDM, 2015. [8] K. Ide et al., Phys. Status Solidi A, 2019 [9] J. He et al., Adv. Electron. Mater., 2019. [10] Y. C. Chien, , et al, IEEE EDL, 2020. [11] R. Winter et al, J. Vac. Sci. Technol. B, 2013. [12] R. N. Bukke et al., IEEE EDL, 2016. [13] E. Yurchuk et al., IEEE TED, 2016. [14] Y. Luo et al., IEDM, 2021. [15] V. Betz et al., Architecture and CAD for Deep-Submicron FPGAs 1999. [16] K. E. Murray et al., ACM TRETS, 2020. [17] J. Wu et al., VLSI, 2021. [18] Z. Wang et al., IEDM, 2021. [19] H. K. Peng et al., IEEE EDL, 2022. [20] A. J. Tan et al., IEEE EDL, 2021. [21] K. Ni et al., IEEE TED, 2021. [22] J. Cong, et al., IEEE T VLSI SYST, 2013.

Acknowledgement

This work is supported by Agency for Science, Technology and Research (A*STAR), Singapore under its AME Programmatic Funds (A1892b0026), National Research Foundation Grant RSS2015-003, and the Singapore Hybrid-Integrated Next-Generation μ-Electronics (SHINE) Centre hosted at the National University of Singapore (NUS).