Design-Technology Co-Optimization (DTCO) for Emerging Disruptive Logic & Embedded Memory Process Technologies

Jessie Xuhua Niu, Hasita Veluri, and Aaron Voon-Yew Thean

Department of Electrical and Computer Engineering, National University of Singapore,

4 Engineering Drive 3, Singapore 117583

Email contacts: jessie.niu@nus.edu.sg, Aaron.Thean@nus.edu.sg

Abstract

In this paper, we discuss the DTCO of two disruptive emerging device technologies; Vertical gate-all-around FETs and embedded monolithic 3D (1T1R) integration of 2D material-based resistive RAM and switch transistors. We showed that VFET has the potential for lower parasitics and improved SRAM performance. In the case of the 2D material-based 1T1R, we show that stacking of nanosheets and reduction of set current is necessary to scale the cell below 0.1μm^2 cell sizes.

Introduction

To enable continual circuit density and energy scaling for 3nm and beyond technologies, disruptive device technologies that lead to major changes to the circuit and computational system design may emerge. Beyond the FinFETs and the lateral stacked Nanowires and Nanosheets, the scalability of lateral thin-body FETs to sustain continual logic circuit area scaling will come into question. In addition, with the increasing machine-learning and deep-learning applications, where there is a demand for high-density integration of embedded non-volatile memories. The convergence of new devices and new architectures require early DTCO to understand new process-system dependencies due to the intersections. In this paper, we discuss the DTCO of two disruptive emerging device technologies; Vertical gate-all-around FETs and embedded monolithic 3D (1T1R) integration of resistive RAM and switch transistors.

Vertical GAA Logic FETs

Simulations have shown that GAA Nanowire (NW) FETs offer incremental improvement in short-channel electrostatics (SCEs) over FinFETs, allowing the gate-length (Lgate) to scale down to 15nm (Fig. 1) with NW critical diameter (CD) of 5-7nm. However, the transistor gate, sidewall spacers, and source/drain contacts of a lateral device (LFET) compete for space within the contacted gate pitch (CGP), making continual scaling unsustainable (Fig. 1) [1]. This is exacerbated by the growing non-scalability of gate and contact to manage SCE, resistance, yield, and reliability. Unless we find a solution that fundamentally changes the device-circuit physical design, the gate-contact non-scalability leads to a dead end for logic/SRAM cell scaling for 3nm (N3) and beyond (Fig. 1).

The objective of vertical device (VFET) is to orientate the device channel vertical with respect to the circuit wiring topology, this brings the opportunity to overcome the CGP-Gate-Contact non-scalability (Fig. 2). Physical design of basic standard cells highlights unique layout attributes (Fig. 3).

VFET has the potential for lower parasitic RCs (Fig. 4). Due to the intrinsically top-down asymmetry of the device with respect to the S/D wiring, VFET's bottom access resistance can be challenged by the increased wiring reach needed (Fig. 2). On the other hand, the SCE improvements afforded by Lgate relaxation in VFETs can translate significantly into switching performance. In one case, as much as 27% VFET Ieff gains at a Lgate= 18nm over a LFET at Lgate=10nm, due to improved subthreshold swing & DIBL with the longer gate.

Constrained by device mismatch and leakage, a read-stable while writable dense SRAM bitcell is a key challenge of technology scaling. Figure 5 shows the 3-D model for 6T-SRAM with vertical nanowires (VNWs) as well as vertical Fins (VFFINs). The Lgate relaxation offered by VFETs offers a mean to optimize for variability and leakage in a scaled SRAM footprint. As such, VFET-based bitcells can be 30% smaller than LNWFET SRAM, due to more compact pull-up NW placement. Comparing LNWFET VS VNWFETs, VFET SRAM bitcells can achieved lower leakage for a given read current (Fig. 6). By replacing VNW in the pull-down (PD) transistor, we can enhance the bitcell read stability for the same writability [1].

Monolithic-3D 1T1R Resistive RAM

The resistive random access memory (RRAM) crossbar array has been extensively studied as a promising candidate for future high-density nonvolatile memory technology. Emerging Non-Von Neumann architectures with intensive in-memory computing like next-generation deep learning and neuromorphic chips will demand high-density integration of embedded memory. 3-D monolithic

978-1-5386-6508-4/19/$31.00 ©2019 IEEE

(sequential) multi-layer stacking of transistors and memory among interconnects (Fig.7). Such architectures will not only overcome the 2-D die limitations, but also enable new 3-D system partitions where logic and memory elements are intimately co-located to significantly improve the memory access bandwidth and energy [2].

Beyond-Si devices that can be co-integrated additively on Si CMOS chips, like Carbon Nanotubes (CNT) FETs and 2-D semiconducting van der Waal-layered crystals (2DMat), can overcome the low thermal budget road block of monolithic 3-D integration. Shulaker et al. has already shown the feasibility of 3-D integration with CNTs [2]. 2DMat TFT with oxide RRAM has also been reported [3]. In this work, we investigate the full 2-D 1T1R with 2-D material TFT and RRAM.

One of the key design considerations is the mitigation of sneak path current. Sneak current limits array size and degrade the array performance significantly [4]. A selector circuit results in increase in set voltage due to its inherent turn-on voltage. This problem can be alleviated by using a transistor as gating element. However, the TFT drive strength and set/reset current of the memory element needs to be co-designed, taking in consideration of the material properties.

We perform simulation of 1T1R structure integrating 2-D WSe2 RRAM and WSe2 TFT. Figure 8 shows the calibration of BSIM-IMG compact model [5] description of the TFT to measurements and known WSe_2 material parameters. A hysteron-based compact model, as reported by Patterson et al. [6] has been calibrated to our WSe_2 RRAM (Fig. 9). To allow for greater than 100 times LRS-HRS read margin in anticipation of significant device-to-device variability, RRAM with a nominal set current compliance (CC) of 20 μA and set voltage of 0.6 V has been analyzed (Fig. 9).

Two 1T1R configurations with respect to the bit line (BL) were studied. Compared to circuit 1, we find in circuit-2, a reduced portion of the BL voltage (VBL) appears across the RRAM element (Fig.9). Hence, circuit 2 requires higher VBL to achieve the CC required to set the RRAM. Circuit 1 reached the CC of 20 μA with VBL ~ 0.65 V while circuit 2 attains CC at VBL = 0.96 V, leading to circuit 1 being preferred.

We analyze the 1T1R cell scaling base on MOSIS based design rule description, where F =3λ =minimum metal ½ pitch [7]. 1T1R cell size is limited by the select transistor size (min. cell area = $35F^2$). While select transistor drive is strongly dependent on scaling due to its width (W~4/3F), RRAM switching current is largely insensitive to the cell size. Reducing cell size and transistor width negatively impacts switching CC to set (write) the RRAM, which in turn reduces the High-Resistance State – Low-Resistance State (HRS-LRS) read margin. To mitigate this, we propose increasing CC by stacking 2DMat nanosheet-TFT channels to recover the geometrical W reduction due to cell scaling (Fig. 11). For 3nm/5nm technology rules (F~12nm), we find that we have to stack more than 20 TFT channels (NStack > 20) to support CC of 10-20 μA, for all cell sizes below 0.02 μm2 (4x $35F^2$) (Fig.12). Only when we reduce the RRAM set current down to 5 μA, we find NStack < 10, for the smallest cells, this comes with a compromise of HRS-LRS read window by ~10x (Fig. 14). Therefore, both TFT stacking and RRAM set current reduction are necessary to enable dense 1T1R cells.

Conclusion

We have discussed the DTCO of Vertical gate-GAA FETs and embedded monolithic 3D 1T1R integration of 2D material-based devices. We showed that VFET has the potential for lower parasitics and improved SRAM performance over lateral devices. However, there are compromises on drive strength. In the case of the 2D material-based 1T1R, we showed that stacking of nanosheets and reduction of RRAM set current is necessary to scale the cell below 0.1μm^2 cell sizes.

Acknowledgments

The authors gratefully acknowledge prior work from IMEC, Belgium. The work has also been supported in part by Singapore's National Research Foundation.

References

[1] Thean et. al. Symp. VLSI Tech, 2015, pp.T26-T27
[2] Shulaker et. al., DATE, 2015, pp. 1197 – 1202.
[3] Yang et. al IEDM, 2017, pp. 19.5.1 - 19.5.4
[4] Kim et. al. IEEE Trans. Electron Devices, v. 61, No. 8, Aug. 2014.
[5] BSIM-IMG model (http://bsim.berkeley.edu/models/bsimimg/)
[6] Patterson et.al, SCED 2017.
[7] Yeh et.al, IEEE J.S.S.C, v.50, Issue:5 p.1309, 2015.

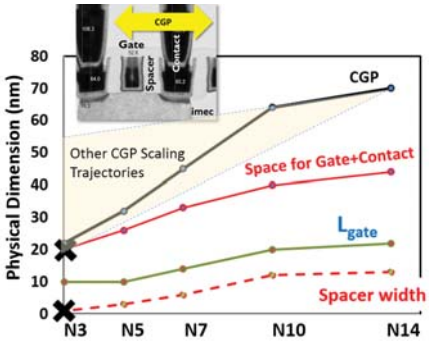

Fig. 1 Expected 1st-order scaling trend of CGP, gate, contact, & spacer size for upcoming technology nodes.

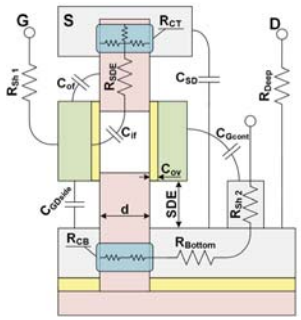

Fig. 2 Illustration of vertical Gate-All-Around (GAA) devices. Channel can be NW or Fin in structure.

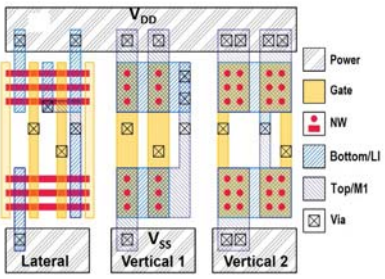

Fig. 3 Lateral device Vs Vertical device physical layout of NAND2 logic gate. Two Versions of Vertical1: Parasitic Optimized, Vertical2: Drive optimized.

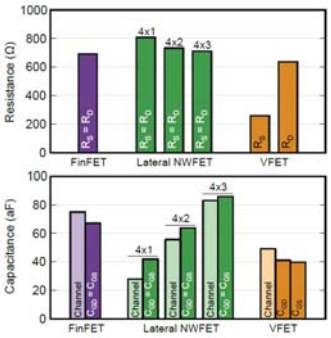

Fig. 4 Parasitic RC comparison between 3 different device architecture. Lateral NWFET is according to nfin x nstack.

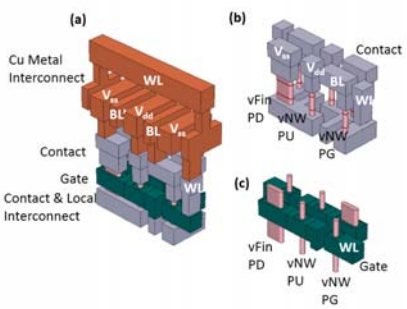

Fig. 5 Model of the SRAM bitcell with vertical devices, showing the layers (a) consisting of metal interconnect & contact layers, (b) contact, local-interconnect, & exposed transistor bodies, and (c) gate layer with transistor bodies.

	VNWFET			LNWFET		
Cell type	HD	LV	HP	HD	LV	HP
Configurations	111	112	122	111	112	122
Area [μm²]	0.010	0.013	0.013	0.014	0.016	0.016

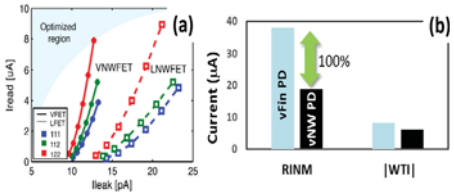

Fig. 6 Comparison between (a) LNWFET SRAM & VNWFET SRAM, (b) Performance of VFET SRAM with vFin PD Vs. vNW PD SRAM cell.

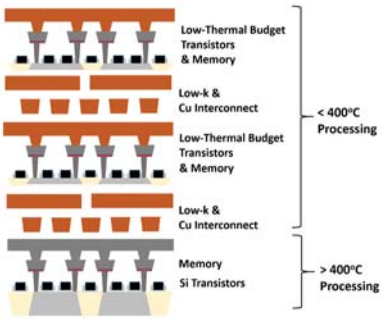

Fig 7. (a) Illustration of monolithic 3-D Integration of low-thermal budget transistor & memory.

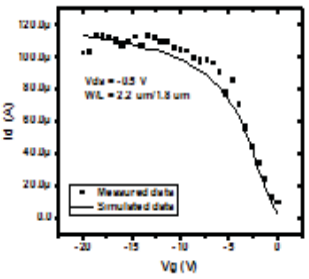

Fig 8: Id-Vg of SPICE TFT model Vs. measured 2-D WSe2 PFET

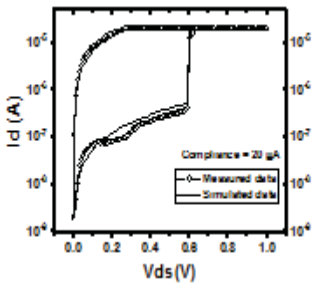

Fig 9: I-V of SPICE RRAM model Vs. measured for positive switching bias

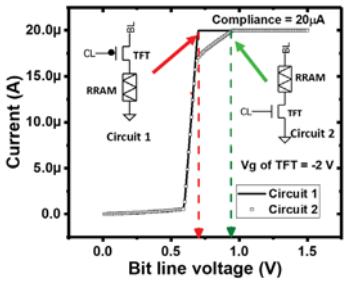

Fig 10: Two 1T1R circuits investigated for R/W the RRAM and switching current through RRAM for the two circuits. configurations.

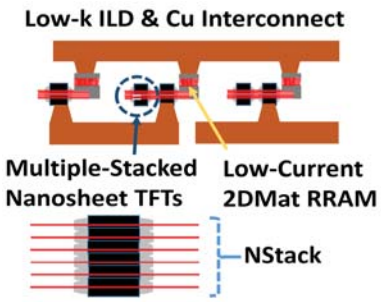

Fig. 11: Conceptual illustration of multiple-stacked TFTs with RRAM.

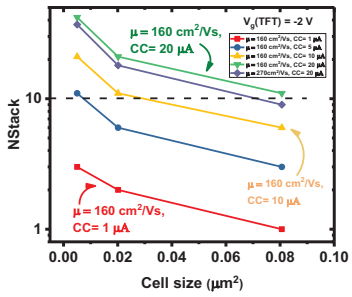

Fig 12: NStack (no. of nanosheet-TFT stacked) for different feature sizes.