Hierarchical simulation of monolithic CFETs using atomistic and continuum models

Woosung Choi[1†], Hyeon-Kyun Noh[1], Hong-Hyun Park[1], Anh-Tuan Pham[1], Seonghoon Jin[1], Byounghak Lee[1], Chihak Ahn[1], Hiroyuki Kubotera[1], Dae Sin Kim[2]
[1]*TCAD lab Samsung Semiconductor Inc. 3655 N 1st street, San Jose, CA 95134, United States*
[2]*Computational Science & Engineering Team, Samsung Electronics, 1-1 Samsungjeonja-ro, Hwaseong-si, Gyeonggi-do, Korea*
[†]*email:woosung.c@samsung.com*

2024 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD) | 979-8-3315-1635-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/SISPAD62626.2024.10733121

***Abstract*— We present a number of in-house TCAD capabilities that are required for the simulation and analysis of the future logic devices, e.g. the complementary FET (CFET) beyond 2nm node. A few selected simulation analyses of key process and device parameters of the monolithic CFET are presented. In particular, the atomistic simulation of the epitaxial growth of the source and drain of the stacked NMOS and PMOS, device performance evaluation using NEGF and MSBTE models, and the extraction of the parasitic resistance and capacitance of the 6T SRAM with CFETs are demonstrated.**

Keywords—CFET, TCAD, KMC, EPI, PEX

I. Introduction

For the current state-of-the-art 3nm logic technology node, FinFET is still used, however MBCFET™ is adopted as the better option than FinFET thanks to its superior gate controllability and larger effective channel width [1]. Going forward, due to the limitation of device scaling knobs more innovative ideas on device architecture are required. In addition, the congestion of the interconnect lines including middle-of-line (MOL) makes it more challenging to reduce the standard cell size [2]. To solve those problems, the idea of the 3D stacked FET (3DSFET) structure has been proposed in both monolithic and sequential fabrication methods[2-8]. The pros and cons of the monolithic 3DSFET (a.k.a. CFET) in comparison to the sequential counterpart are explained in [4,8]. The former has the advantage of self-aligned patterning and lower wafer costs in comparison to the latter, however the high aspect ratio in the vertical stacked direction pauses a number of process challenges, one of which is the source and drain (SD) epitaxial growth process (EPI). We will focus on the simulation of the monolithic CFET (m-CFET) with nanosheet (NS) NFET on top of NS PFET, and show atomistic SD EPI simulation, device simulation using non-equilibrium Green's function (NEGF) [10], multi-subband Boltzmann transport equation (MSBTE), and drift-diffusion (DD) models [11,12], and the parameter extraction (PEX) of the parasitic resistance and capacitance (RC) components from the 6T SRAM cell utilizing GPGPU computing [13]. Potential issues identified from the EPI simulation of m-CFET will be discussed. From the device simulation using NEGF and MSBTE models, we will discuss on a few device parameters affecting its performance. Finally, we demonstrate the parasitic RC PEX from a SRAM cell comprising of four m-CFET's of stacked NMOS and PMOS with two of the bottom PMOS inactive.

II. Simulation Setup

The SRAM cell with m-CFET is created by the in-house process emulation based on the architecture shown in [14]. The final structure and the abridged process flow are shown in Fig. 1. We chose the silicon substrate instead of the SOI so that the EPI region can grow from the bottom silicon seeds. It allows for the larger compressive PMOS channel stress from the EPI grown SiGe SD regions. Note that we chose the signal and ground lines (VSS) laid on top of the transistors and the power lines (VDD) in the buried power rail (BPR) in this study.

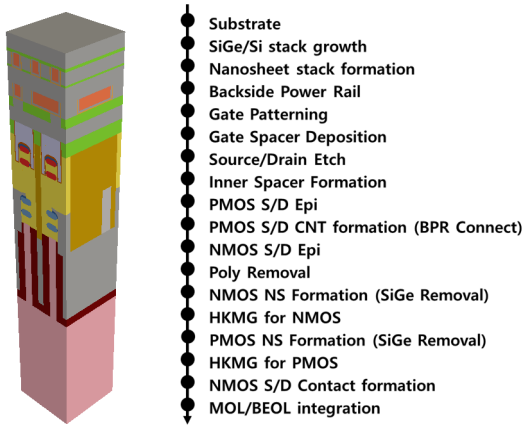

Fig. 1. CFET SRAM structure (a) from process flow (b).

Fig. 2 (a) shows the cross-section of the m-CFET stacks before the EPI process of PMOS SD. In this example, there are two NS channels for PMOS and NMOS, respectively. After the PMOS SD EPI growth (Fig. 2(b)), a dielectric layer is formed for the

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on November 07,2024 at 13:58:46 UTC from IEEE Xplore. Restrictions apply.

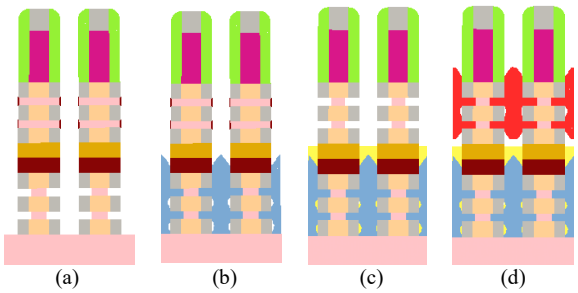

Fig. 2. Cross-section view of m-CFET (a) before PMOS EPI growth, (b) after PMOS SD EPI growth, (c) after dielectric isolation layer formation, and (d) after NMOS SD EPI growth.

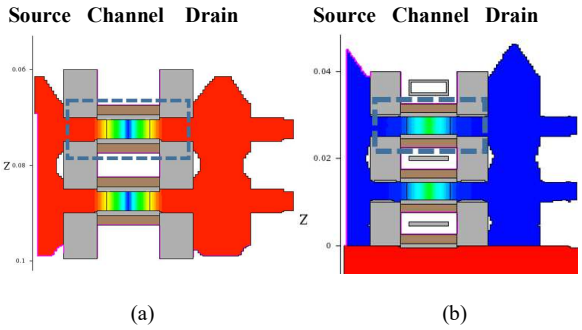

Fig. 3. 2-d cross-section view of device structure for (a) NMOS and (b) PMOS for device simulation. A simplified single channel device structure with uniform channel geometry is created from the dashed box area in the NMOS and PMOS, respectively. Analytic doping profile is applied to SD (N_{peak}=5×10^{20}cm^{-3}, Gaussian decay with λ_c=2nm, underlap of 2nm from gate edge) and the channel is undoped. The simplified device geometry is with channel thickness of 4nm, channel width of 9nm, gate length of 13nm, SD length of 13nm, gate oxide thickness of 1nm, and gate high-k dielectric thickness of 2nm. The NMOS channel is unstrained and the PMOS channel is with uniform uniaxial stress of -1 GPa. Wafer orientation and channel direction are set as (100) and <110>, respectively.

isolation of top and bottom SD (Fig. 2(c)) followed by the NMOS SD EPI growth shown in Fig. 2(d) [5]. Simplified single channel NMOS and PMOS structures for the device simulation using NEGF and MSBTE simulation are created from the final CFET structure (Fig. 3). Phonon scattering and surface roughness scattering are included in both NEGF and MSBTE simulation for the Id-Vg and Id-Vd simulations. Then on the SRAM cell in Fig. 1, PEX is performed using the in-house field solver on GPGPU computing server.

III. Results and Discussion

The kinetic Monte-Carlo (KMC) model [9] is applied to the atomistic simulation of the SD EPI growth coupled with the raytracing of downstream atoms as an emulation of gas flow to provide the boundary condition of the surface reaction. With the given NS width of 9nm and fin pitch of 38nm, it is shown that the bottom SD of the PMOS is not merged yet while the top one is (Fig. 4(a)). If we increase the fin pitch by 6nm, the bottom SD can be merged (Fig. 4(b)). It implies that the SD formation by EPI process can be significantly affected by the fin pitch and NS width of the CFET depending how much space is allowed for the depositing elements to reach the growing surface. A similar effect is observed in the NMOS EPI growth.

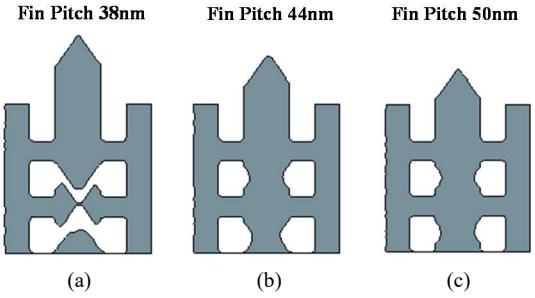

Fig. 4. PMOS SD structure from KMC EPI simulation: (a) unmerged with fin pitch of 38 nm, (b) and (c) merged with fin pitch of 44 nm and 50 nm, respectively.

With the EPI grown SD as shown in Fig. 4(b), stress simulation is performed for the PMOS. Fig. 5 shows the stress profile in the PMOS SD and channel with the uniform germanium mole fraction of 30% which generates about -1 GPa uniaxial compressive stress in the channel.

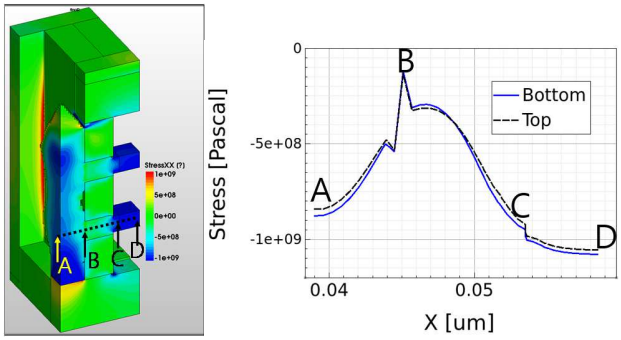

Fig. 5. Stress profile of the PMOS device: (a) 3-d profile of the stress, (b) 1-d cut profile of the stress. The germanium mole-fraction is assumed as 30% and uniform in the SD.

Next the device simulation is performed using the NEGF and MSBTE models on the simplified single channel NS NMOS and PMOS with the given feature sizes shown in Fig. 3(a) and Fig. 3(b), respectively. The NMOS channel is assumed unstrained and the PMOS channel is with the uniform uniaxial stress of -1 GPa. The SD contacts are assumed ohmic for further simplification by applying a high SD active doping concentration of 5×10^{20}cm^{-3}. Fig. 6 shows the Id-Vg and Id-Vd curves after V_T adjustment. Using the NEGF or MSBTE simulation data, the density gradient (DG) model and DD model can be calibrated for the simulation of realistic device structure of Fig. 7 [12]. For the DD simulation with DG model as the quantum correction, a novel model of tunneling quantum correction potential at the Schottky contact is proposed [15]. In this ultra-small feature scale with tight room for improvement of intrinsic transistor performance, the SD and contact engineering is a very critical knob to reduce the extrinsic resistance of the transistors. In that respect, another physical phenomenon called quantum access resistance (QAR) is identified in the NS devices as the QAR can affect the on-current significantly by the constriction of carrier transport path from SD to channel

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on November 07,2024 at 13:58:46 UTC from IEEE Xplore. Restrictions apply.

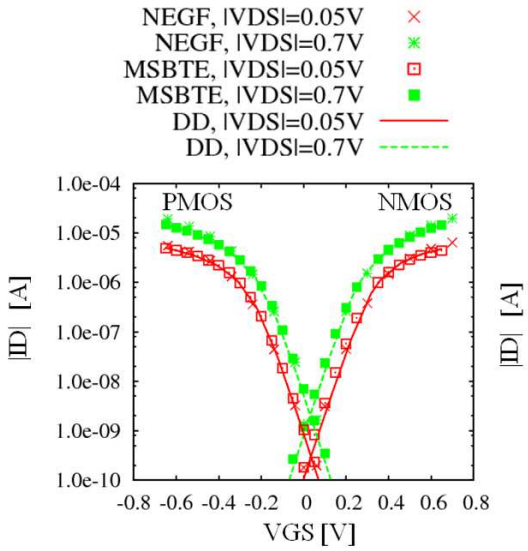

Fig. 6. Simulated Id-Vg and Id-Vd characteristics after V_T adjustment from NEGF, MSBTE, and the calibrated DD simulation. The sub-threshold slope of NMOS and PMOS are 73 and 72 mV/decade, respectively, and the DIBL are 105 and 86 mV/V for NMOS and PMOS, respectively. The DIBL is relatively large due to the undoped channel.

combined with relatively low SD doping in the channel access [16]. A proper optimization of SD doping may be required to minimize the negative effect. For the simulated devices in this paper, however, the QAR effect is suppressed due to the relatively high SD doping near the channel access. On the other hand, the variation of the on-current due to the atomistic random dopants can be as significant as 6% as shown in Fig. 8. With the calibrated DD model, the NMOS and PMOS in Fig. 7 are simulated and the IV characteristics are shown in Fig. 9. The parasitic RC components as tabulated in Table 1 are computed by the PEX solution of the 6T SRAM cell of which the 3-d schematic and cross-section view are shown in Fig. 10. For example, the large parasitic resistance from VSS to the NMOS source is shown as well as the coupling capacitances of word line (WL) to bit line (BL) and WL to VSS that are detrimental to the AC performance of the cell. An in-house PEX solver is developed for the application to cell-level and block-level simulations. For this SRAM example the solution time is quite short as it takes only two minutes with four V100 GPUs. It should be noted that the use of GPGPUs for PEX is more advantageous for larger cells as shown in Fig. 11 as the GPGPU solver enables the PEX on them and with one order of magnitude faster speed compared to the solver that uses only CPUs.

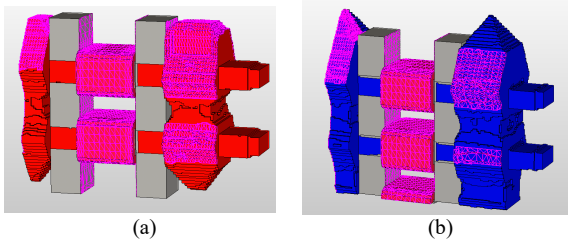

Fig. 7. Device structure with mesh for the device simulation using drift-diffusion model. The SD doping is analytic and the channel is undoped as described in Fig. 3. The silicide contacts are shown in the magenta colored grids in the SD.

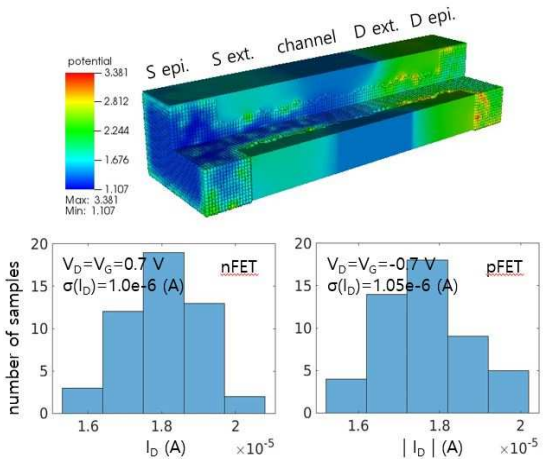

Fig. 8. I_{ON} variation due to atomistic random dopants and surface roughness.

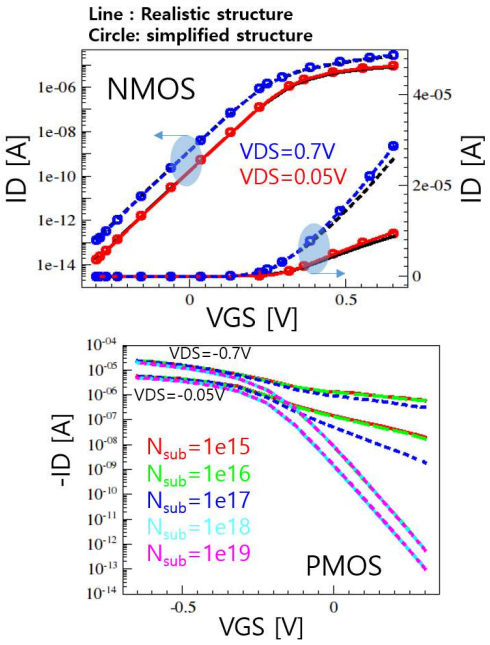

Fig. 9. Top: IV characteristic of the realistic NMOS from the calibrated DD simulation matches that of the simplified structure. Bottom: Sub-threshold characteristic of PMOS as a function of the uniform substrate doping concentration.

TABLE I. RC PARAMETERS FROM PEX

From	To	C [aF]	From	To	R [Ohm]
WL	$\overline{BL}$	23.1	$\overline{BL}$	N4 Drain	16.36
WL	VSS	29.2	VSS	N2 Source	23.32
WL	BL	23.1	VSS	N1 Source	23.86
WL	VDD	9.8	BL	N3 Drain	16.40
$\overline{BL}$	VSS	8.3	WL	N4 Gate	3.13
$\overline{BL}$	VDD	0.7	WL	N3 Gate	3.12
VSS	BL	8.3	N2 Gate	P2 Gate	3.54
VSS	VDD	5.4	N1 Gate	P1 Gate	3.46

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on November 07,2024 at 13:58:46 UTC from IEEE Xplore. Restrictions apply.

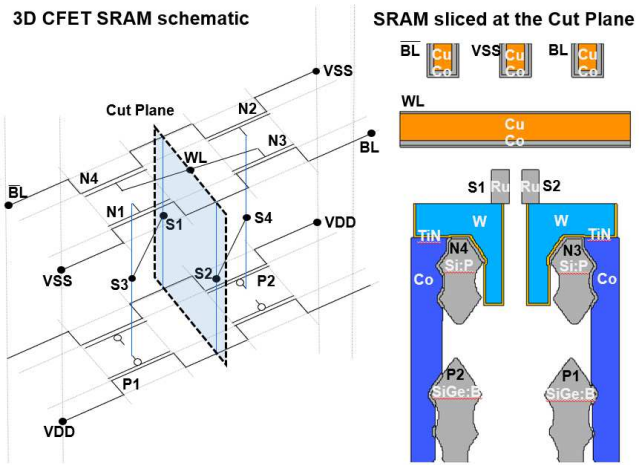

Fig. 10. 3-d schematic of the 6T SRAM (left) and the slice of the SRAM cell at the cut plane. N3 and N4 are the access NMOS's and P1 and P2 are the bottom PMOS's of the CMOS inverter pairs comprising of (N1,P1) and (N2,P2), respectively. The two unannotated PMOS's are inactive.

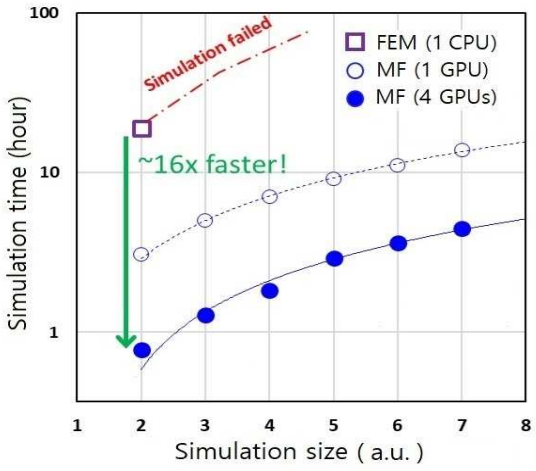

Fig. 11. Performance comparison of PEX simulations. The use of matrix-free (MF) method enables the simulation of larger cells than the regular finite element method (FEM) which requires the matrix assembly and factorization. The FEM simulation using CPU fails on the larger structures due to the memory limitation. Also, the use of GPGPUs combined with MF method significantly speeds up the PEX simulation.

IV. Conclusion

We have demonstrated that a TCAD simulation flow using a monolithic CFET works seamlessly. It comprises the process emulation to construct the SRAM cell structure and the KMC EPI simulation to create more realistic source and drain shapes. Advanced carrier transport models such as the NEGF and MSBTE are applied to the calibration of the DD model that is used for the analysis of the CFET device characteristics. The PEX solver using GPU shows its potential for the large-scale cell-level or block-level applications.

Acknowledgment

We would like to express our sincere acknowledgment to Hisashi Kotakemori at Samsung Device Solutions Research Japan for his excellent technical support and Young-Seok Song at Samsung Device Solutions Korea for sharing the valuable benchmark data of the PEX solvers.

References

[1] S. Choi, "Redefining Innovation: A Journey forward in New Dimension Era," in 2023 IEEE International Electron Devices Meeting (IEDM), Dec 2023.

[2] J. Ryckaert et al., "The Complementary FET (CFET) for CMOS scaling beyond N3," in 2018 IEEE Symposium on VLSI Technology, June 2018.

[3] C.-Y. Huang et al., "3-D Self-aligned Stacked NMOS-on-PMOS Nanoribbon Transistors for Continued Moore's Law Scaling," in 2020 IEEE International Electron Devices Meeting (IEDM), Dec 2020.

[4] M. Radosavljević et al., "Opportinities in 3-D stacked CMOS transistors," in 2021 IEEE International Electron Devices Meeting (IEDM), Dec 2021.

[5] J. Park et al., "First demonstration of 3-dimensional stacked FET with top/bottom source-drain isolation and stacked n/p metal gate," in 2023 IEEE International Electron Devices Meeting (IEDM), Dec 2023.

[6] M. Radosavljević et al., "Demonstration of a Stacked CMOS Inverter at 60nm Gate Pitch with Power Via and Direct Backside Device Contacts," in 2023 IEEE International Electron Devices Meeting (IEDM), Dec 2023.

[7] S. Liao et al., "Complementary Field-Effect Transistor (CFET) Demonstration at 48nm Gate Pitch for Future Logic Technology Scaling," in 2023 IEEE International Electron Devices Meeting (IEDM), Dec 2023.

[8] N. Horiguchi et al., "3D Stacked Devices and MOL Innovations for Post-Nanosheet CMOS Scaling," in 2023 IEEE International Electron Devices Meeting (IEDM), Dec 2023.

[9] R. Chen, W. Choi, A. Schmidt, K.-H. Lee, and Y. Park, "A new kinetic lattice Monte Carlo modeling framework for the source-drain selective epotaxial growth process," in 2013 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Sept 2013.

[10] H.-H. Park, W. Choi, M. A. Pourghaderi, J. Kim, U. Kwon, and D.-S. Kim, "NEGF simulations of stacked silicon nanosheet FETs for performance optimization," in 2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Sept 2019.

[11] S. Jin et al., "Performance evaluation of InGaAs, Si, and Ge nFinFETs based on coupled 3D drift-diffusion/multi-subband Boltzmann transport equation solver," in 2014 IEEE International Electron Devices Meeting (IEDM), Dec 2014.

[12] S. Jin, A.-T. Pham, W. Choi, M. A. Pourghaderi, J. Kim, and K.-H. Lee, "Performance Evaluation of FinFETs: From Multi-Subband BTE to DD Calibration," in 2016 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Sept 2016.

[13] CUDA C++ programming guide, Nvidia.

[14] V. Moroz et al., "DTCO Launches Moore's Law Over the Feature Scaling Wall," in 2020 IEEE International Electron Devices Meeting (IEDM), Dec 2020.

[15] A.-T. Pham et al., "Tunneling quantum correction potential model for drift-diffusion simulations of Schottky contact resistance," in 2024 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Sept 2024.

[16] K.-Y. Kim et al., "Quantum Transport through a Constriction in Nanosheet Gate-all-around Transistor," submitted to Nature Electronics.

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on November 07,2024 at 13:58:46 UTC from IEEE Xplore. Restrictions apply.