Combined Process Simulation and Emulation of an SRAM Cell of the 5nm Technology Node

Xaver Klemenschits, Siegfried Selberherr, and Lado Filipovic
Institute for Microelectronics, TU Wien, Gußhausstraße 27-29/E360, 1040 Wien, Austria
Email: klemenschits@iue.tuwien.ac.at

***Abstract*—Fast geometric process emulation models were combined with sophisticated physical simulations on a common simulation platform. In order to show the capabilities of this approach, an entire SRAM cell was generated with process specific features in the source and drain regions, which can be used to study the process-aware resistance and capacitance characteristics of the circuit. Therefore, we have shown that using this approach, even sophisticated physical models can be used to simulate the fabrication of complex devices within a matter of minutes. The simulation flow presented here was carried out in less than 16 minutes on a consumer desktop computer.**

I. Introduction

Ongoing miniaturisation of metal oxide semiconductor field effect transistors (MOSFETs), has led to considerable improvements in computing performance, a reduction in the power dissipation and the total chip area of integrated circuits (ICs). Modern circuits include a variety of different materials in complex three-dimensional (3D) arrangements, with tight design rules and little room for process variations. To save on production time and costs, it is essential to understand process variations and their effect on device performance in detail. Process technology computer aided design (TCAD) is an important component in the design technology co-optimisation (DTCO) cycle used to understand the origin of process variations and to find potential means for their mitigation. However, sophisticated physical simulations require substantial computational effort, making it often unfeasible to simulate an entire integrated circuit process flow. Especially for variations across entire device cells the computational requirements of physical models lead to unacceptable runtimes. Nonetheless, many fabrication steps can be emulated using computationally efficient empirical models, while intricate physical simulations need only be carried out for critical process steps. We present the seamless integration of sophisticated physical process models with simple emulation models in the in-house simulation framework of ViennaLS [1].

II. Process Simulation and Emulation

Vertical-gate 3D transistor designs have been employed for several years and the FinFET has become the standard design for modern ICs up to the most recent technology nodes [2]. However, increased parasitic resistance and capacitance due to the 3D nature of the design, have become a concern for process engineers [3]. These properties depend heavily on the exact geometry of the source and drain (S/D) regions, so to simulate this effect, these regions must be modelled with high physical accuracy. Using typical process TCAD approaches, either the entire device must be physically simulated, which is unfeasible due to the long runtimes of such simulations, or the entire device must be emulated based only on empirical measurements and thus without proper physical properties. Due to simulation and emulation frameworks being based on different data structures, moving the geometries between them introduces non-negligible overhead and potential loss of surface information. In order to show the capabilities of the combination of process simulation and emulation on a single simulation platform, a full SRAM cell of the 5nm technology node was generated using this approach. The physical simulation models, encompassing Monte Carlo (MC) ray tracing and modelling of the chemical reactions on the material interfaces, are used to model processing steps which shape the S/D regions in order to achieve the most physical description possible. The other numerous processes, required to generate the SRAM cell, are only emulated using

978-1-6654-0685-7/21/$31.00 ©2021 IEEE

the recently presented geometric level set advection algorithm [4], keeping simulation runtime small. Therefore, highly physically accurate simulation models can be applied for one part of interest of a device, while highly computationally efficient emulation models are applied for all other parts. Thereby, this approach provides the fast generation of process-aware structures and ultimately enables the efficient investigation of process-specific device properties for a large range of different process parameters with reasonable simulation effort.

III. Process Flow for a 5nm SRAM cell

A detailed process flow for the generation of an SRAM cell first described in [5] is provided in Table I. We generate this structure using our combined emulation and simulation framework, only simulating the steps in bold text in Table I. These steps are discussed in more detail in the following.

A. Fin Patterning

The first important step for an accurate description of the S/D regions is the fin creation, performed here with a SF_6/CH_2F_2 ion-enhanced plasma etch model [6]. The process model is based on the physical model described in [7] using two different particle types: directional CF_x^+ ions, and neutral SF_x radicals. A thin protective CF_x polymer is formed on the sidewall by line-of-sight deposition. In order to properly describe this type of polymer deposition, reflections and re-emissions must be modelled accurately, requiring sophisticated MC ray-tracing methods [8], because line-of-sight deposition is highly dependent on the particle transport through the feature scale region. This type of sidewall passivation leads to a highly controlled, vertical etch profile. However, as can be seen in Fig. 1a, some tapering does occur, especially at the etch front.

B. Low-k Spacer Etch

Once the Silicon fins have been patterned, the shallow trench isolation (STI) and a dummy gate with gate spacer are formed, all encapsulated in an interlayer dielectric (ILD). The ILD is then patterned to free the NMOS or PMOS regions in order to make the spacer and fin accessible. To form low-resistance contacts at the S/D regions, the spacer must be etched away without damaging the Silicon fin. This is usually achieved in a selective plasma etching process using a CH_3F based chemistry [9]. This fabrication step is modelled using two particle types, ions and neutrals, as described above. However, sidewall passivation is achieved through polymer growth of neutral particles stabilising the surface. In our model, the selectivity to Silicon is achieved via the deposition of a thick protective polymer above the Silicon surface through the adsorption of neutral particles. Nevertheless, etching of the fin is also observed, as the reactive layer penetrates into the Silicon [10]. This may lead to some damage at the top of the Silicon fin, as can be seen in Fig. 1b and Fig. 1e, respectively.

C. Fin Recess

A clean Silicon surface for the S/D epitaxy is then generated by etching the S/D regions of the fin isotropically. This step is modelled using a constant rates approach, meaning there is one neutral particle type, which etches the Silicon at the same rate in every direction. If the spacer was not removed completely, the Silicon fin is only etched effectively in the exposed regions, leading to the confined fins shown in Fig. 1c and Fig. 1f, respectively.

D. S/D Epitaxy

The last step heavily influencing the S/D regions of a device is the formation of the S/D contacts via epitaxial growth. The growth rate varies strongly for different crystal orientations and thus produces the characteristic diamond shapes shown in Fig. 1h. The epitaxial growth is modelled using the Stencil-Local-Lax-Friedrichs (SLLF) numerical level set advection scheme [11], taking into account the nature of the strongly varying speed function. The directional rates are interpolated from experimental data [12] to all necessary directions, depending on the surface normal.

IV. Results

Due to the application of sophisticated physical models, process specific effects can be observed in the S/D regions. Due to the insufficient removal of the fin spacer, a so-called confined fin is created. Subsequently, the fin recess step leads to a particularly shaped fin, characterised by a peak in its centre, embedded inside the fin spacer. Hence, the epitaxially grown silicon wraps around the spacer

Fabrication step	Applied model
Fin Mask	Mask
Fin Patterning	**SF_6/CH_2F_2 Plasma Etching**
STI Deposition, CMP	Geometric, CMP
STI Etching	Geometric
Dummy Gate Depo, CMP	Geometric, CMP
Dummy Gate Mask	Mask
Gate Patterning	Geometric
Spacer Deposition	Geometric
ILD Deposition, CMP	Geometric, CMP
Once for NMOS and PMOS each	
Mask NMOS/PMOS	Mask
PMOS/NMOS ILD Etch	Geometric
Spacer Etch	**CH_3F Plasma Etching**
Fin Recess	**Selective Dry Etch**
S/D Epitaxy	**SLLF Epitaxial Growth**
ILD Deposition, CMP	Geometric, CMP
Dummy Gate Removal	Geometric
HKMG Deposition, CMP	Geometric, CMP

TABLE I: Process steps used to generate the final FinFET structure and how they are modelled. The bold text shows which steps were applied using physical models. The Mask models add the respective mask from a two-dimensional mask layout, the Geometric models perform a geometric advection [4] with empirical parameters, and CMP simply cuts materials at a certain height.

Fabrication step	Simulation Runtime
Fin Mask	0.1s
Fin Patterning	**186.8s**
STI Deposition, CMP	0.3s
STI Etching	13.5s
Dummy Gate Deposition, CMP	0.3s
Dummy Gate Mask	0.1s
Gate Patterning	24.4s
Spacer Deposition	2.8s
ILD Deposition, CMP	0.3s
Once for NMOS and PMOS each	
Mask NMOS/PMOS regions	0.2s / 0.1s
PMOS/NMOS ILD Etching	60.7s / 76.6s
Spacer Etching	**204.7s / 225.9s**
Fin Recess	**60.6s / 63.6**
S/D Epitaxy	**28.8s / 28.2s**
ILD Deposition, CMP	0.3s / 0.4s
Dummy Gate Removal	0.1s
HKMG Deposition, CMP	6.7s
Emulation Models	2m 22.8s
Physical Models	**13m 18.6s**
Total Runtime	15m 41.4s

TABLE II: Runtime for the simulation of each modelled process step. Physical models are shown in bold text. The simulation was carried out by an AMD Ryzen3950X processor and took less than 16 minutes to complete, where more than 85% of the simulation time was consumed for the evaluation of the physical models.

after it has grown out from the top of the fin in its characteristic diamond shape, as shown in Fig. 1d. Therefore, a realistic geometry including process-induced effects is created, adequately representing features of the S/D region. This implementation allows to study the impact of varying physical parameters on the full SRAM cell, with minimal computational effort consumed for the generation of the non-S/D relevant structures.

As shown in Table II, the total simulation is kept very small, due to the geometric emulation models which are applied for fabrication processes not affecting the S/D regions directly. Physical models are again shown in bold text, highlighting how much of the total simulation time they require, although they only form a small set of the simulated fabrication steps.

For large geometries, even emulation models may require significant computational effort, as is the case for the etching of the PMOS/NMOS interlayer dielectric. Even though the changes are applied using only geometric considerations, the large volume of the material which must be removed, in combination with the complex structure underneath the material, results in runtimes of over one minute to perform these emulation steps. Still, the presented model and combined simulation and emulation framework is far more efficient than physics-based models and thus results in a substantial decrease in simulation time. As shown in Table II, the emulation models only require around 15% of the total runtime, although more processing steps with larger effects on the geometry were performed with these models. On the other hand, the physics-based models could be focused on a few processing steps of interest in order to generate the most accurate picture possible and to study physically-induced variations in these steps, as shown in Fig. 1h. Due to the computational effort required to simulate particle transport and surface chemistry these models required more than 85% of the simulation time.

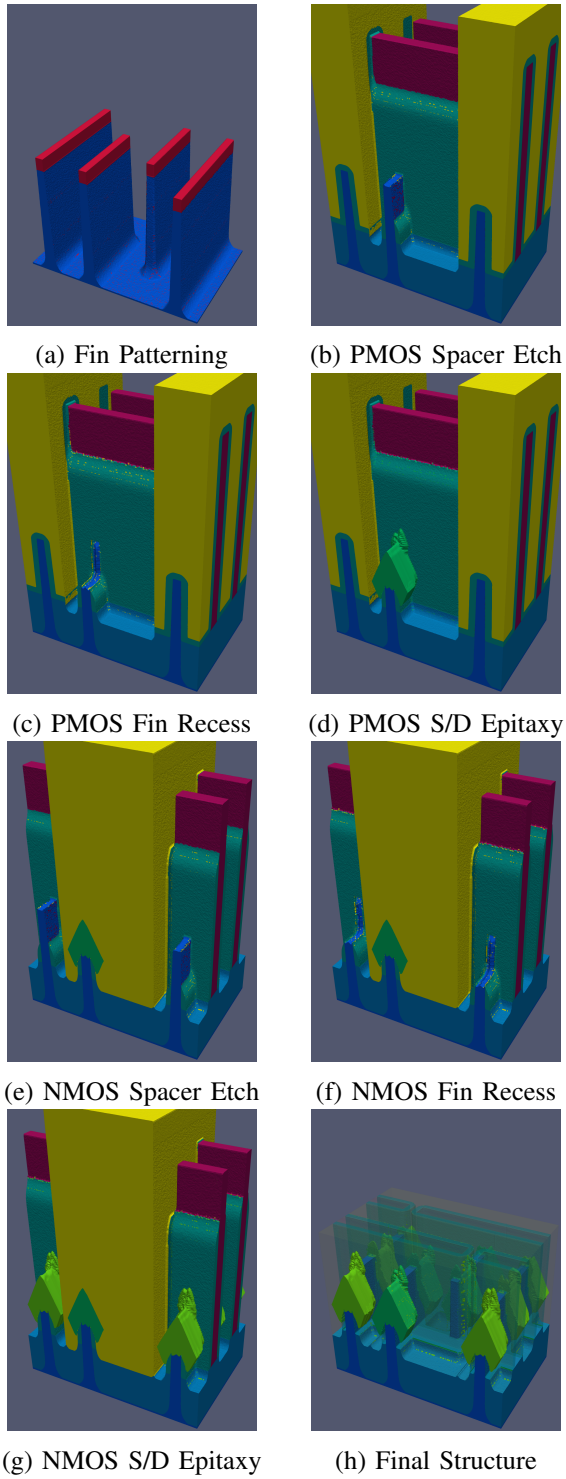
V. Conclusion

Process simulation and emulation capabilities were combined on a single level set based software platform, enabling the alternating use of physics-based simulation models and empirical emulation models. By focusing on a single device region of interest, computationally costly simulations have to be carried out only there. All other regions can be formed quickly and efficiently with empirical emulation models, as these regions do not affect the final device properties of interest, nonetheless leading to a full device description.

The applicability of this approach is presented by generating an SRAM cell of the 5nm technology node, focusing on the exact geometry of the S/D regions. Sophisticated physical models are used to simulate all process steps shaping the S/D regions. Hence, the final geometry clearly shows process-specific properties, such as varying tapering, silicon fin damage, and the resulting crystal shape of the epitaxially grown S/D regions, while the overall simulation time was below 16 minutes. The vast majority of the total run time, more than 85%, is consumed by the simulation of fabrication steps shaping the S/D region. Due to the immense computational power required to conduct physical simulations, until now it has not been feasible to simulate large structures including sophisticated process-aware features. Only through the introduction of geometric modelling on the same simulation platform presented here, it became possible to reduce the computational effort enough to conduct such simulations on consumer desktop computers rather than large scale scientific clusters. Therefore, by combining highly sophisticated models with fast emulation models, we have shown that intricate process specific properties of entire device cells can be generated within within a short time, demanding only moderate computation resources.

Acknowledgements

This work was supported in part by the Austrian Research Promotion Agency FFG (Bridge Young Scientists) under Project 878662 ”Process-Aware Structure Emulation for Device-Technology Co-Optimization”.



(a) Fin Patterning (b) PMOS Spacer Etch (c) PMOS Fin Recess (d) PMOS S/D Epitaxy (e) NMOS Spacer Etch (f) NMOS Fin Recess (g) NMOS S/D Epitaxy (h) Final Structure

Fig. 1: (a)-(g) SRAM structure after the fabrication steps with physical models, highlighted in bold in Table I. (h) Final SRAM structure, with high-k dielectric and metal gate (HKMG), spacer, and ILD transparent to show the structure of the fins and S/D regions.

REFERENCES

[1] X. Klemenschits, O. Ertl, L. Filipovic, P. Manstetten, and J. Weinbub. ViennaLS. [Online]. Available: https://github.com/ViennaTools/ViennaLS

[2] A. Razavieh, P. Zeitzoff, and E. J. Nowak, "Challenges and limitations of CMOS scaling for FinFET and beyond architectures," *IEEE Transactions on Nanotechnology*, vol. 18, pp. 999–1004, 2019.

[3] B.-R. Huang, F.-H. Meng, Y.-C. King, and C. J. Lin, "Investigation of parasitic resistance and capacitance effects in nanoscaled FinFETs and their impact on static random-access memory cells," *Japanese Journal of Applied Physics*, vol. 56, no. 4S, pp. 04CD11–1–04CD11–5, 2017.

[4] X. Klemenschits, S. Selberherr, and L. Filipovic, "Geometric advection algorithm for process emulation," in *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2020, pp. 59–62.

[5] G. Yeap, X. Chen, B. R. Yang *et al.*, "5nm CMOS production technology platform featuring full-fledged EUV, and high mobility channel FinFETs with densest $0.021\mu m^2$ SRAM cells for mobile SoC and high performance computing applications," in *2019 IEEE International Electron Devices Meeting (IEDM)*, 2019, pp. 36.7.1–36.7.4.

[6] E. Altamirano-Sánchez, V. Paraschiv, M. Demand, and W. Boullart, "Dry etching fin process for SOI FinFET manufacturing: Transition from 32 to 22nm node on a 6T-SRAM cell," *Microelectronic Engineering*, vol. 88, no. 9, pp. 2871–2878, 2011.

[7] X. Klemenschits, S. Selberherr, and L. Filipovic, "Modeling of gate stack patterning for advanced technology nodes: A review," *Micromachines*, vol. 9, no. 12, pp. 631–1–631–31, 2018.

[8] O. Ertl and S. Selberherr, "Three-dimensional topography simulation using advanced level set and ray tracing methods," in *2008 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2008, pp. 325–328.

[9] H. Miyazoe, N. Marchack, R. L. Bruce *et al.*, "Nitride etching with hydrofluorocarbons III: Comparison of C_4H_9F and CH_3F for low-k nitride spacer etch processes," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 36, no. 3, pp. 032 201–1–032 201–7, 2018.

[10] N. Possémé, M. Garcia-Barros, C. Arvet, O. Pollet, F. Leverd, and S. Barnola, "Silicon nitride spacer etching selectively to silicon using $CH_3F/O_2/He/SiCl_4$ plasma," *Journal of Vacuum Science & Technology A*, vol. 38, no. 3, pp. 033 004–1–033 004–7, 2020.

[11] A. Toifl, M. Quell, X. Klemenschits, P. Manstetten, A. Hossinger, S. Selberherr, and J. Weinbub, "The level-set method for multi-material wet etching and non-planar selective epitaxy," *IEEE Access*, vol. 8, pp. 115 406–115 422, 2020.

[12] D. Dutartre, A. Talbot, and N. Loubet, "Facet propagation in Si and SiGe epitaxy or etching," *ECS Transactions*, vol. 3, no. 7, pp. 473–487, 2019.