From Atoms to Reactors: Multi-Scale Modeling for Semiconductor Fabrication

Lado Filipovic
CDL for Multi-Scale Process Modeling of Semiconductor Devices and Sensors at the
Institute for Microelectronics, TU Wien, 1040 Vienna, Austria
Email: filipovic@iue.tuwien.ac.at

***Abstract*—Accurately predicting surface topography evolution is essential for manufacturing complex three-dimensional (3D) semiconductor devices. To overcome the limitations of slow and costly experimental development cycles, a predictive, physics-based simulation approach that bridges multiple scales is required. This paper presents comprehensive multi-scale modeling approaches that integrate simulations from the atomistic to the reactor level. At the lowest scale, we use Density Functional Theory (DFT) and Molecular Dynamics (MD), including machine learning-enhanced methods, to derive fundamental parameters like sticking coefficients and sputtering yields. These inform robust feature-scale models implemented in the open-source ViennaPS simulator, which captures complex surface kinetics by tracking local properties such as species coverages, passivation layer thickness, and physical damage. At the highest scale, reactor simulations provide physically-based boundary conditions, while surrogate models are developed to efficiently link equipment settings to process outcomes. The predictive power of the methods is demonstrated through a calibrated model for selective SiGe etching that accurately predicts etch profiles for different device geometries and process conditions. This holistic approach provides an end-to-end simulation capability, enabling predictive process optimization and establishing a foundation for future applications in automated process control and inverse design.**

***Index Terms*—Multi-scale modeling, Semiconductor fabrication, Process TCAD, Topography simulation, Molecular Dynamics, Plasma etching, PECVD, DTCO.**

I. Introduction

The semiconductor industry is driven by continuous miniaturization and the development of increasingly complex 3D device architectures, such as gate-all-around (GAA) transistors, complementary field-effect transistors (CFETs), and 3D NAND memories [1]. This relentless scaling demands advanced fabrication techniques capable of achieving high precision and extreme aspect ratios. Accurately predicting and controlling surface topography evolution during these processes is crucial for optimizing manufacturing yield, device performance, and overall manufacturability [2]. This optimization, known as Process/Design Technology Co-Optimization (PTCO/DTCO), combines process development and circuit design, ensuring that all aspects are considered together for optimal outcomes.

Traditional process development often relies on costly and time-consuming experimental iterations, which limit the full exploration of the design space. To overcome these challenges, process and device simulations, commonly referred to as Technology Computer-Aided Design (TCAD), have become indispensable tools. Process topography TCAD enables the prediction of etch and deposition profiles, guiding process development and significantly reducing the time and cost associated with in-silicon experiments, potentially shortening development cycles from months to days [3].

The overall TCAD workflow, from equipment and process simulation to device characterization and circuit-level analysis, is depicted in Fig. 1. This integrated approach allows for comprehensive optimization across different design scales and stages. However, the complexity of modern semiconductor processes, involving multiple plasma species and highly sensitive surface interactions, poses significant challenges for developing generalized and predictive models. Many existing models depend on empirical fitting against experimental data, which can limit their applicability to new plasma chemistries and process conditions due to the difficulty in fully knowing or reproducing exact experimental conditions.

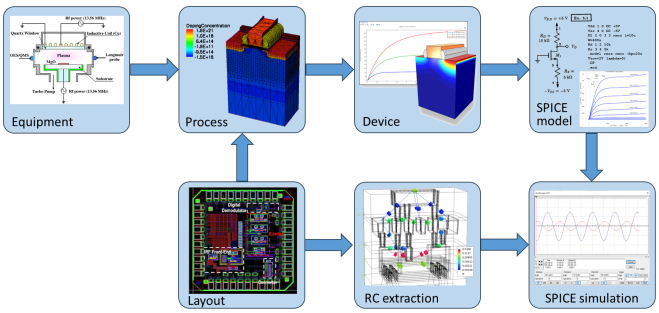


Fig. 1. Integrated TCAD workflow spanning from equipment and process simulations to device modeling and SPICE simulations for circuit design.

This paper presents comprehensive multi-scale modeling approaches that combine atomistic, feature, and reactor scales to enable predictive process design and optimization in semiconductor fabrication. We demonstrate how applying these approaches, while utilizing tools like the in-house feature-scale simulator ViennaPS [4], reduces reliance on costly experimental iterations, thereby establishing a foundation for inverse design workflows and real-time equipment analysis during fabrication.

II. Multi-Scale Process Modeling

Our multi-scale modeling approach integrates different levels of physical phenomena, ranging from atomic interactions to

macroscopic reactor behavior, to provide a comprehensive and accurate simulation of semiconductor fabrication processes, as illustrated in Fig. 2 and discussed in [5].

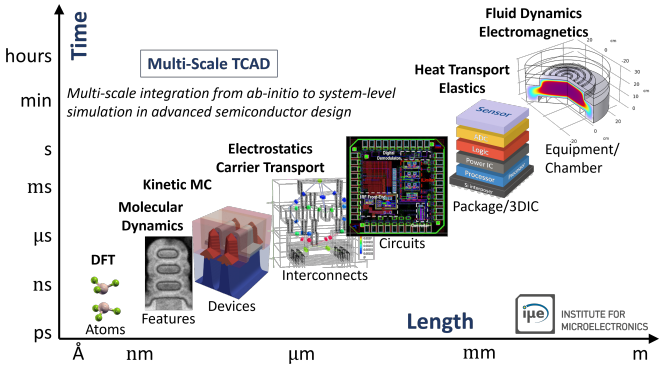


Fig. 2. The multi-scale modeling paradigm in semiconductor TCAD. Simulations range from ab-initio methods at the atomistic level up to continuum models for heat transport and fluid dynamics at the equipment level.

A. Atomistic Scale: Molecular Dynamics and First-Principles

At the most fundamental level, the interaction of plasma species with material surfaces is governed by atomic-scale phenomena. Molecular dynamics (MD) simulations and Density Functional Theory (DFT) provide crucial insights into these interactions, which are often inaccessible through experiments. These methods help in understanding surface reactions such as adsorption, desorption, and sputtering [6].

In reactive ion etching (RIE), MD simulations can be used to derive angle- and energy-dependent sputtering yields [7], [8]. Recent work has focused on determining the sputtering yield for CF_3^+ ions on SiO_2 substrates [9]. By simulating ion impacts at various energies and angles, a threshold energy for sputtering (e.g., 8.4 eV for CF_3^+ on SiO_2) and an angle-dependent yield function can be extracted. Atomistic simulations also reveal the formation of a fluorocarbon-rich layer on the substrate surface, which effectively changes the target material and physically justifies a lower energy requirement for sputtering than pure SiO_2. These simulations eliminate the need for empirically fitting critical parameters to experiments, enhancing the predictive power and robustness of the models.

DFT calculations are instrumental in determining fundamental parameters for classical simulations, such as the adsorption energies and sticking coefficients of reactive species. By mapping the potential energy surface for an etchant precursor interacting with a substrate, DFT can reveal the activation energy barriers for key reaction steps like dissociative chemisorption. For example, recent DFT studies have investigated the chemisorption of methyl fluoride on silicon nitride surfaces for atomic layer etching [10] and to calculate the sticking coefficients for silicon- and carbon-containing species during the growth of silicon carbide (SiC) [11].

Furthermore, large-scale atomistic modeling based on machine learning molecular dynamics (MLMD) frameworks are emerging to simulate etching processes, such as the lateral isotropic selective etching of SiO_2/Si stack structures in GAA FETs using F^* radicals. Recent MLMD models have shown to handle over 600,000 atoms, achieving significant speedups and improved accuracy compared to classical molecular dynamics (CMD) [12]. This approach leverages active learning strategies to generate datasets for machine learning potential functions, continuously correcting errors with DFT during the etching process. The general concept of this MLMD approach, which links quantum accuracy to large-scale simulations, is illustrated in Fig. 3.

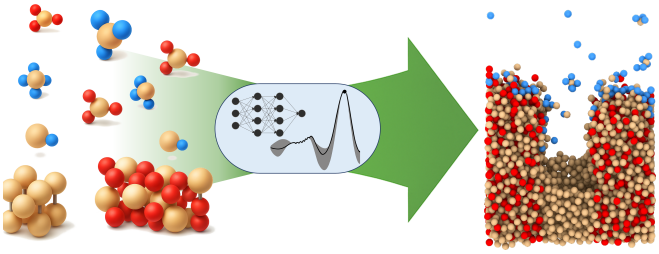


Fig. 3. Conceptual overview of the machine learning molecular dynamics (MLMD) workflow. High-accuracy data from small-scale quantum mechanical calculations (DFT) on various atomic configurations (left) is used to train a machine learning potential (center). This computationally efficient potential then enables large-scale, classically-run MD simulations of complex processes like plasma etching (right) with near-DFT accuracy.

B. Feature Scale: ViennaPS Process Simulator

The insights gained from atomistic simulations are directly integrated into feature-scale models that simulate the 3D topography evolution of semiconductor devices. ViennaPS is an open-source process simulator and emulator developed at the Institute for Microelectronics, TU Wien, designed for modeling semiconductor fabrication and operation [4]. The modular software architecture of ViennaPS is shown in Fig. 4.

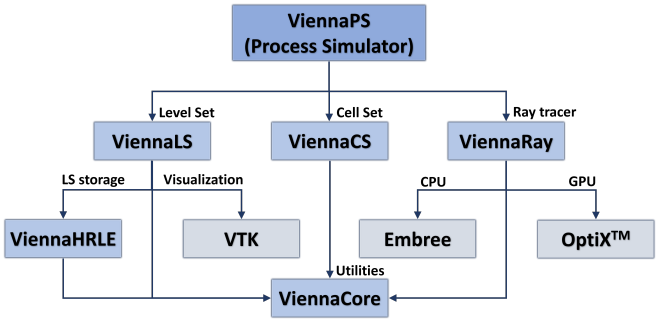
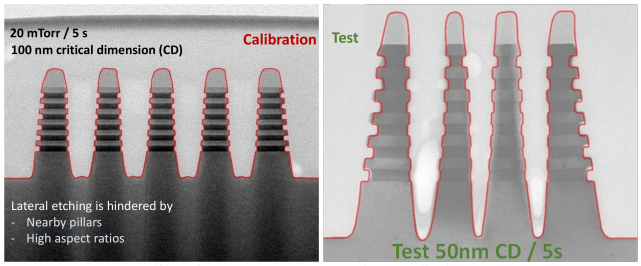


Fig. 4. The software architecture of ViennaPS. It is built on a modular framework including components for Level Set (ViennaLS) and Cell Set (ViennaCS) methods, and a ray tracer (ViennaRay) with support for both CPU (Embree) and GPU (NVIDIA Optix™) acceleration.

ViennaPS utilizes the Level-Set (LS) and Cell-Set (CS) methods for surface and volume representation and evolution [13]. The LS represents surfaces as the zero-level set of a higher-dimensional function, allowing for seamless handling of complex topological changes such as merging and splitting [14]. For computational efficiency, a narrow-band or sparse-field level set is stored instead of the full domain. The CS, on the other hand, discretizes the domain into voxels, which can store material information and interface properties.

The physical models in ViennaPS calculate local surface velocities by modeling the interactions of incoming particles

with the surface [14]. This is typically achieved by solving surface-site balance equations that track one or more properties stored locally on the surface. These properties can range from simple species coverages, as used in models for SF_6/O_2 or CF_4/O_2 plasma etching, to the thickness of passivation films [15] or the extent of material damage [16]. The predictive power of this physics-based approach is demonstrated in Fig. 5 for the selective etching of SiGe. A model for CF_4/O_2 plasma is calibrated against a specific experimental condition. The same model is then used to accurately predict the etch profile for a different geometry with a smaller critical dimension, validating its physical basis for capturing complex microloading effects [17].



(a) Calibration (b) Predictive Test (Geometry)

Fig. 5. Demonstration of the predictive capability of the calibrated CF_4/O_2 model for selective SiGe etching. The simulation results (red lines) are overlaid on experimental TEM images. (a) The model is first calibrated to a multi-pillar structure with a 100 nm critical dimension (CD) at 20 mTorr. The same calibrated model is then used to accurately predict the etch profile for (b) a different geometry with a 50 nm CD [17].

Beyond species coverages, the framework can store other physical properties to link sequential process steps. One example is the modeling of advanced deep reactive ion etching (DRIE) techniques like the polymer-free CORE process [18]. The CORE process avoids the environmental concerns and process drift associated with fluorocarbon-based Bosch etching by using only SF_6/O_2 plasmas in four cyclic steps: Clear, Oxidize, Remove, and Etch. The process uses an O_2 plasma in the "Oxidize" step to grow a thin, self-limiting oxide layer for sidewall passivation. Our model handles this efficiently by tracking the oxide thickness as a scalar surface property rather than as a computationally expensive geometric layer [15]. This stored oxide is then directionally removed from the feature bottom during the high-bias "Remove" step, allowing the subsequent "Etch" step to proceed anisotropically. This modeling approach accurately reproduces characteristic features like sidewall scalloping and allows for the optimization of such advanced, environmentally friendly etching processes.

To improve computational efficiency, especially for complex geometries and large domains, ViennaPS leverages GPU-accelerated ray tracing using NVIDIA Optix™. This significantly speeds up flux calculations, which are a major component of the simulation [19].

C. Reactor Scale: Plasma Simulations and Surrogate Models

The feature-scale models require accurate boundary conditions from the plasma environment. Reactor-scale plasma simulations, often performed using tools like the Hybrid Plasma Equipment Model (HPEM) from the University of Michigan, provide these spatially resolved ion and neutral flux distributions, as well as ion energy and angular distributions (IEADs/NEADs) [20]. The HPEM framework contains several interacting modules to solve for the plasma chemistry, electromagnetics, and fluid kinetics, allowing it to simulate various reactor types such as Inductively Coupled Plasmas (ICP).

Fig. 6 shows key outputs from an HPEM simulation of an Ar/O_2 inductively coupled plasma. The simulated process conditions, a 13.56 MHz coil power of 700 W, an RF substrate bias of 500 W, a total gas flow of 30 sccm, and a gas pressure of 45 mTorr with an Ar/O_2 ratio of 0.1/0.9, are designed to generate a highly energetic plasma. The ion energy and angular distribution (IEAD) for O^+ ions (left) is highly directional, arriving at the wafer at near-normal incidence. The corresponding ion energy distribution function (IEDF) (right) further reveals the distinct effect of the high substrate bias, displaying a broad and complex structure with significant ion populations extending to energies above 200 eV. This detailed, physically-based data on ion directionality and energy is precisely what is required as input for high-fidelity feature-scale simulations of ion-driven processes.

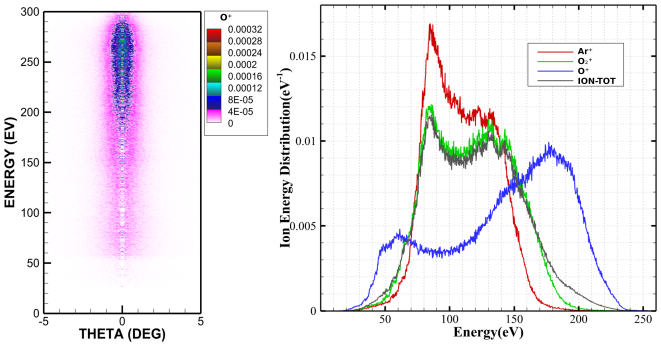


Fig. 6. Example simulation results from HPEM for an Ar/O_2 plasma under high substrate bias. Left: The Ion Energy and Angular Distribution (IEAD) for O^+ shows a highly directional flux. Right: The Ion Energy Distribution Function (IEDF) reveals a broad energy range with a bimodal structure, characteristic of RF-biased plasmas.

To further improve efficiency, especially for DTCO applications where rapid feedback is crucial, machine learning-based surrogate models can be developed [21]. These surrogate models are trained on data from detailed reactor simulations and/or experimental measurements to predict plasma parameters (e.g., particle fluxes, IEADs) as a function of equipment inputs (e.g., coil power, gas flow rate, pressure, bias voltage). This allows to capture complex reactor behavior without the computational cost of running full-scale plasma simulations for every design iteration. Multi-variable spline interpolation models, for example, have been demonstrated to generate these surrogate models effectively, as depicted in Fig. 7.

For plasma-enhanced chemical vapor deposition (PECVD), we work with hybrid datasets consisting of both real production data (e.g., 39 points) and simulation data (e.g., 1000 points from a validated physical model) to train ML models [22]. The input features for these models typically include gas

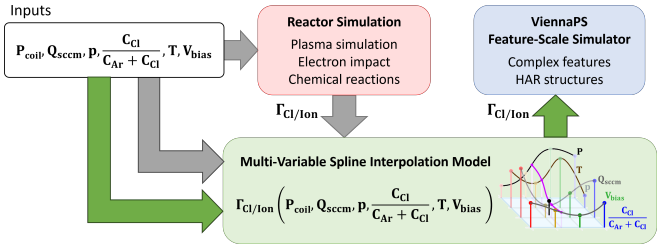


Fig. 7. Workflow for creating (gray arrows) and using (green arrows) a surrogate model to link equipment and feature scales. Reactor simulations are performed for a range of equipment input parameters to generate a database of resulting particle fluxes. A fast surrogate is trained on this data to predict the fluxes from the equipment settings, bypassing the need for computationally expensive reactor simulations.

flow rates (SiH_4, NH_3, N_2, H_2), process conditions (RF power, chamber pressure, deposition time), and equipment parameters (electrode spacing). The surrogate model, when built using SHAP (Shapley Additive exPlanations), can provide interpretability by quantifying the impact of various input features on deposition rates [22]. This not only accelerates process tuning but also offers insights into the underlying physics, which can be valuable for real-time DTCO. For example, SHAP analysis on a CatBoost model for PECVD deposition rate prediction showed that gas flow is the most influential factor (60% of total contribution), with SiH_4 and NH_3 flows being dominant. Higher chamber pressure and RF power also positively contribute to the deposition rate due to increased plasma density.

D. Automated Calibration and Optimization

To ensure model fidelity and reduce reliance on manual tuning, experimental SEM/TEM images are used for automated model calibration. This involves comparing simulated profiles with experimental images and iteratively adjusting model parameters (e.g., reaction rates, material-specific parameters) to minimize the difference between the two [23]. When using ViennaPS, the signed distance function (SDF) can be used to quantify these differences, and an optimizer can then adjust parameters to reduce the errors [24]. This automated tuning allows for rapid model calibration and continuous improvement of predictive capabilities. Fig. 8 illustrates this automated workflow, where the difference between a target experimental profile and a simulated profile is minimized by an optimizer that iteratively adjusts model parameters.

III. Conclusion

We have presented a comprehensive multi-scale modeling approach for semiconductor fabrication that spans from atomistic to reactor scales, providing an invaluable tool for predictive process design and optimization. By integrating molecular dynamics, feature-scale simulations with ViennaPS, and reactor-scale plasma models, we enable a holistic understanding and control over complex manufacturing processes. This integrated modeling strategy is critical for overcoming the challenges of continued miniaturization and increased use of the vertical dimension in semiconductor manufacturing.

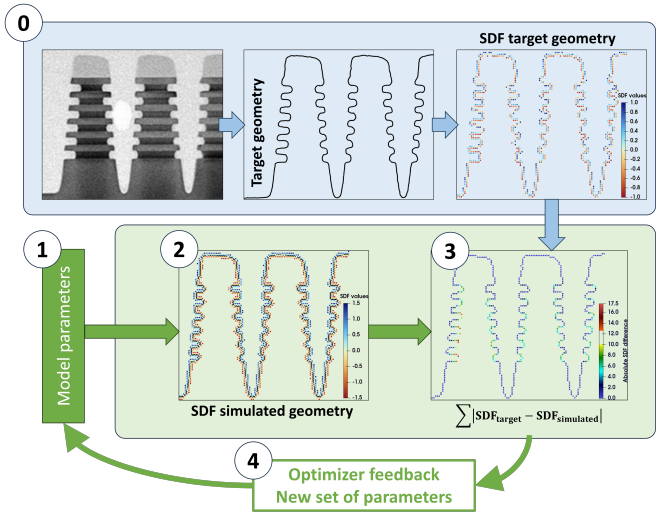


Fig. 8. The automated calibration loop. (0) A 2D SEM image is converted to a contour and then to an SDF representation of the target geometry. (1) A set of model parameters is selected for simulation. (2) ViennaPS generates a corresponding simulated geometry and its SDF. (3) The difference between the target and simulated SDFs is calculated to quantify geometric mismatch. (4) This error metric is fed back into a gradient-free optimizer, which proposes a new set of parameters for the next iteration. This loop continues until optimal agreement is achieved between simulation and experiment.

Acknowledgment

Financial support by the Federal Ministry of Labour and Economy, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged. The author wishes to thank Tobias Reiter, Robert Stella, Panpan Lai, and Roman Kostal from the Institute for Microelectronics, TU Wien; Ziyi Hu, Hua Shao, and Rui Chen from IMECAS; Andreas Hössinger from Silvaco; Manuel Kleinbichler from Infineon; and Prof. Mark Kushner from the University of Michigan for their valuable contributions.

References

[1] M. M. Shulaker *et al.*, *Nature*, 547(7661), pp. 74–78, 2017.
[2] M. Shen *et al.*, *Jpn. J. Appl. Phys.*, 62(SI), p. SI0801, 2023.
[3] A. Asenov *et al.*, in *Proc. VLSI-DAT*. 2017, pp. 1–2.
[4] T. Reiter *et al.*, ViennaPS 3.5.0, 2025.
[5] N. Xu, "GPU-Accelerated semiconductor process simulations: From atom to transistor," TSMC, NVIDIA, presented at NVIDIA GTC 2025.
[6] D. B. Graves and P. Brault, *J. Phys. D Appl. Phys.*, 42(19), p. 194011, 2009.
[7] N. Nakazaki *et al.*, *Jpn. J. Appl. Phys.*, 53(5), p. 056201, 2014.
[8] K. Ono *et al.*, *Thin Solid Films*, 518(13), pp. 3461–3468, 2010.
[9] R. Stella *et al.*, in *Proc. SISPAD*. 2025, pp. 01–04.
[10] E. Cheng and G. S. Hwang, *Appl. Surf. Sci.*, vol. 543, p. 148557, 2021.
[11] A. A. Alao *et al.*, *Vacuum*, vol. 205, p. 111414, 2022.
[12] Z. Feng *et al.*, in *Proc. IEDM*. 2024, pp. 1–4.
[13] T. Reiter *et al.*, in *Proc. SISPAD*. 2023, pp. 85–88.
[14] J. Bobinac *et al.*, *Micromachines*, 14(3), p. 665, 2023.
[15] T. Reiter *et al.*, in *Proc. SISPAD*. 2025, pp. 01–04.
[16] T. Reiter *et al.*, *Solid-State Electron.*, vol. 192, p. 108261, 2022.
[17] H. Shao *et al.*, *ACS Appl. Electron. Mater.*, 6(11), pp. 8124–8133, 2024.
[18] V. Thi Hoang Nguyen *et al.*, *ECS J. Solid State Sci. Technol.*, 9(2), p. 024002, 2020.
[19] T. Reiter *et al.*, in *Book of Abstracts (IWCN*, 2025, pp. 122–123.
[20] M. J. Kushner, *J. Phys. D Appl. Phys.*, 42(19), p. 194013, 2009.
[21] N. Kuboi, "*J. Micro-Nanolith. Mem.*, 22(04), 2023.
[22] Y. Zhai *et al.*, *Advanced Intelligent Discovery*, 2025, in Production.
[23] F. Krüger *et al.*, *J. Vac. Sci. Technol. A*, 42(4), 2024.
[24] R. Kostal *et al.*, in *Proc. SISPAD*. 2025, pp. 01–04.