

Unified Lattice Kinetic Monte Carlo Model for Epitaxy, Etch and Reflow Processes for Advanced Semiconductor Device Architectures

Alexander Schmidt
CSE Team,
Samsung Electronics
Hwaseong, Korea
alexander.shmidt@samsung.com

Kyungmi Yeom
CSE Team,
Samsung Electronics
Hwaseong, Korea
k.yeom@samsung.com

Byoungnak Lee
TCAD-lab,
Samsung Semiconductor Inc.
San Jose, United States
byoungnak.l@samsung.com

Marton Voeroes
TCAD-lab,
Samsung Semiconductor Inc.
San Jose, United States

Hiroyuki Kubotera
TCAD-lab,
Samsung Semiconductor Inc.
San Jose, United States

Woosung Choi
TCAD-lab,
Samsung Semiconductor Inc.
San Jose, United States

Anthony Payet
Process TCAD Lab,
Samsung Device Solutions R&D Japan,
Yokohama-shi, Japan

Takeshi Nishimatsu
Process TCAD Lab,
Samsung Device Solutions R&D Japan,
Yokohama-shi, Japan

Hiroo Koshimoto
Process TCAD Lab,
Samsung Device Solutions R&D Japan,
Yokohama-shi, Japan

Yasuyuki Kayama
Process TCAD Lab,
Samsung Device Solutions R&D Japan,
Yokohama-shi, Japan

Joohyun Jeon
CSE Team,
Samsung Electronics
Hwaseong, Korea

Seungmin Lee
CSE Team,
Samsung Electronics
Hwaseong, Korea

Jachoon Jeong
CSE Team,
Samsung Electronics
Hwaseong, Korea

Abstract—We have developed a unified framework for chemical vapor deposition (CVD) simulation that enables the simultaneous treatment of deposition, etching, and reflow processes. The framework is based on a lattice kinetic Monte Carlo (LKMC) model coupled with a precursor molecule trajectory tracing algorithm, which accounts for dynamic changes in the structure's shape during the epitaxy process and for local layout effects. Despite the empirical nature of physical models included in the framework, they can be flexibly calibrated to reflect various process conditions and materials. Simulations were used to demonstrate differences in the epitaxy of semiconductor transistor source/drain regions during the transition from FinFET technology to advanced gate-all-around (GAA) transistors.

Keywords— Chemical Vapor Deposition epitaxy, kinetic Monte Carlo, bond-counting model, particle tracing, deposition, etching, reflow, FinFET, Gate-All-Around

I. INTRODUCTION

Since the advent of non-planar semiconductor Logic devices with the introduction of FinFET technology and, especially, after the transition to GAA Multi Bridge Channel (MBCFET) devices [1], Chemical Vapor Deposition (CVD) epitaxial growth of Silicon and SiGe structures has become one of the key factors determining device performance and yield.

Conventional continuum models, which rely on reaction-diffusion equations [2] or the level-set method to simulate epitaxial surface evolution [3], are inadequate for capturing the atomistic events that dominate nanoscale

epitaxy: precursor molecule decomposition, surface step propagation, local facet dependent kinetics, etc. Fully atomistic Molecular Dynamics (MD) simulations can accurately predict atomic interactions, but can cover time scales in the nanosecond range, far shorter than typical epitaxy times (seconds and longer). This huge scale gap can be bridged by methods like Lattice Kinetic Monte Carlo (LKMC), which, by forfeiting explicit treatment of lattice vibrations and by employing empirical assumptions about surface reaction kinetics, can simulate real process time scales while keeping track of the growing surface morphology at the atomic scale [4].

One of the advantages of the CVD process is the possibility of an in-chamber transition from epitaxy to the etching process, depending on the relative flux rates of silicon precursors, etchant gas (typically HCl) and carrier gas (typically H₂). This provides an opportunity to use cyclic processes, when deposition and etch steps are interleaved to generate uniform and high quality epitaxy layers of the desired shape. Therefore a model developed for epitaxy simulation should seamlessly include the etching process.

Another important factor from the technological point of view is the reflow process: high temperature annealing of the surface in a hydrogen atmosphere. While often used for surface cleaning, this process also leads to significant surface diffusion and morphological changes. As the reflow process is often used for initial surface preparation before epitaxy and can significantly affect the resulting epitaxial structure shape [5], incorporating a reflow process model into the KMC

simulation framework is also necessary to build a predictive and quantitatively accurate epitaxy simulation flow.

II. SIMULATION FRAMEWORK

A. Lattice Kinetic Monte Carlo Model and the Precursor Molecule Tracing Module

Previously, we developed a framework for lattice KMC simulation of the epitaxy process taking into account local bonding to neighboring atoms up to the third-nearest neighbors in the silicon lattice [6]. To apply this framework to real transistor structures, we also had to extend it with a model describing the interaction of deposited particles with non-silicon materials (mostly amorphous), e.g. oxide spacers, nitride liners and other elements of semiconductor devices. Since lattice KMC, by definition, requires predefined lattice sites, a quasi-lattice approximation was used to model these non-silicon materials: a virtual silicon lattice is extended into them, and special energy parameters are defined to describe the interaction of silicon atoms with virtual atoms of non-silicon materials. We assume that the interaction is mostly repulsive, reflecting energy barriers at the silicon-oxide and silicon-nitride interfaces, and leading to facet formation at the silicon/non-silicon interface as often observed in experimental data [7]. Equation (6) from ref. [6] was modified as follows:

$$E_b = E_b^{1NN}(n_1) + n_2 E_b^{2NN} + n_3 E_b^{3NN} - \sum_{i=1}^3 n_i^{SP} E_{sp}^{iNN} \quad (1)$$

In this equation, n_i is the number of i -th nearest silicon neighbors of a given lattice site, and E_b^{2NN} and E_b^{3NN} are the binding energy values for next-nearest and next-next-nearest neighbors, respectively. For the nearest neighbor interaction $E_b^{1NN}(n_1)$ a non-linear model is used for a binding energy depending on the number of nearest silicon neighbors (n_1). Interactions with spacer materials are modeled as repulsive with a linear dependence on the neighbor count for all coordination shells. Here, n_i^{SP} is the number of i -th nearest neighbors belonging to non-silicon materials and E_{sp}^{iNN} denotes the corresponding repulsion energy for the respective i -th nearest neighbor atom (Fig. 1).

Another important improvement of the model is the inclusion of the non-uniformity of the incoming precursor flux. Since typical CVD conditions applied in silicon-based semiconductor technology use relatively low pressure and temperature ranges, the typical mean free path of precursor molecules is much larger than the contemporary semiconductor device or pattern size, leading to a Knudsen number much greater than 1. Under these conditions, we can assume that the precursor flux near the surface can be approximated by molecular flow and therefore we can apply ray tracing algorithms for particle trajectory calculations considering only interactions with structure surfaces and ignoring all gas-phase scattering. The ray tracing module (KMCRay) employs the Bounding Volume Hierarchy (BVH) algorithm for extremely fast ray tracing in the continuum structure and uses Alpha-Wrapping meshing [8] which allows efficient extraction of the polygonal surface from the atomistic KMC data. Using particle tracing, local precursor density statistics in the vicinity of the epitaxial surface are first accumulated, and then the probabilities of related reactions (e.g. deposition reactions) are scaled depending on the local incoming precursor densities. After a thin layer is deposited, the new surface configuration is extracted, and the procedure is repeated to update the flux statistics.

The main advantage of this approach is that highly efficient parallelization can be used for particle tracing, since it can be performed simultaneously for a fixed structure. Furthermore, the flow is compatible with rejection-free KMC algorithms, such as [9], since at every KMC step all the scattering statistics are available and fixed. The trade-off is the potential non-uniformity of the flux due to local “depletion” of precursor scattering information if the KMCRay update happens too rarely. We have developed a procedure that ensures that the statistical information is available and automatically initiates its update if needed.

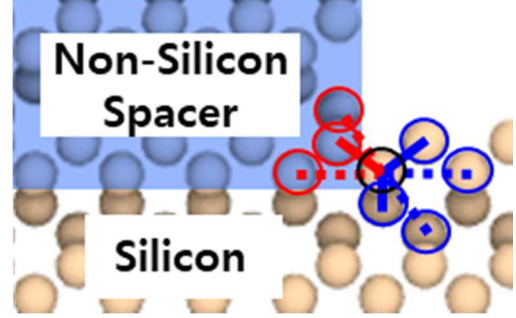


Fig. 1. Example of local bonding for a target lattice site (shown in black) near the spacer (non-silicon material) region: bonds to nearest silicon neighbors are shown with blue solid lines and next-nearest silicon neighbors are shown with dashed blue lines, while repulsive interaction with spacer material neighbors is shown in red (solid and dashed lines for nearest and next-nearest neighbors, respectively). Third-nearest neighbors are not shown for simplicity.

Compared to the conventional homogeneous gas flux assumption that fails to predict void formation in complex topographies, our simulation scheme can provide insight into how local topography affects epitaxial structures.

The combined KMC and KMCRay methodology was applied to simulate source/drain epitaxy in a simplified FinFET-like structure (Fig. 2). A wide Fin device (Fig. 2d) in this case can serve as a representative of a GAA device [10].

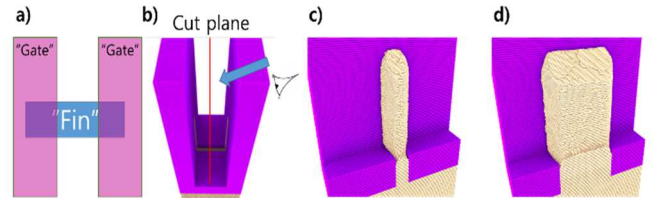


Fig. 2. a) Simplified layout of the FinFET-like device at the source/drain epitaxy step; b) Fin recess cavity is shown along with a cut plane in the middle of growing S/D and a view direction. Intermediate epitaxy stages for narrow (c) and wide (d) fins are presented. LKMC model parameters were calibrated to literature and internal data.

One of the important results of the incorporation of the precursor distribution into the simulation was the ability to account for not only transistor-scale non-uniformities of the epitaxy rate, but also local layout effects, since our algorithm can perform particle tracing in relatively large structures up to several micrometers in size. It was shown that the layout density not only affects the chip-scale growth time [11] but also contributes to variations in the shape of the epitaxially grown structure. The simulation results for the epitaxial growth of the simplified FinFET-like device's S/D region including the effects of precursor reactivity are presented in Fig. 3.

Due to the increased reactivity of the precursors, the flux at the top of the structure increases, leading to faster effective growth. This can lead to premature merging at the top of the structure and therefore to a defect formation. KMC simulations with parameters calibrated to specific process conditions and precursor types allow us to optimize the epitaxy rate, maximizing throughput while keeping the defect formation probability relatively low (Fig. 4)

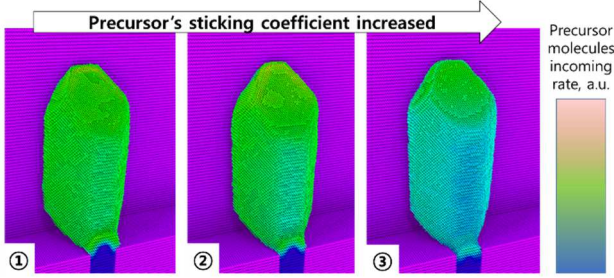


Fig. 3. Three epitaxial structures during early stages of S/D growth: normalized incoming precursor rates are shown with color. Relative growth at the top of the structure increases with the increase of the chemical reactivities (sticking coefficients) of precursors.

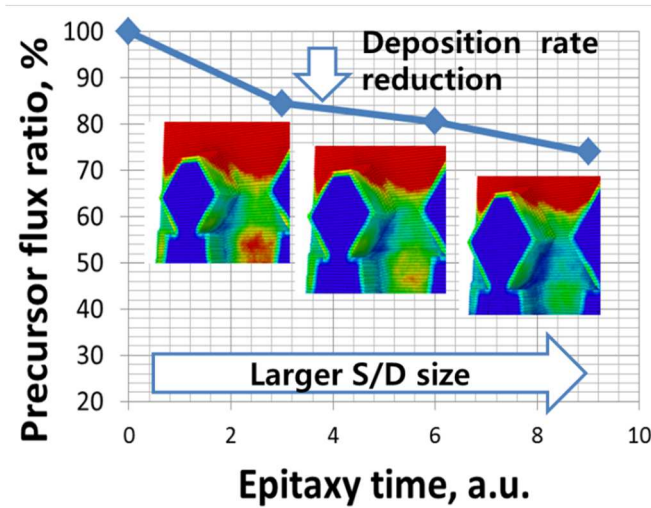


Fig. 4. Due to the dynamic change of the precursors distribution during epitaxy, neighboring devices can influence each other's growth rates and therefore co-optimization of the process depending on local layout is necessary.

B. Efficient Reflow Process Model

At first glance, the reflow process can be easily simulated by adding surface diffusion processes to the LKMC model, because high-rate surface diffusion naturally leads to a surface energy minimization. However, within the KMC method, this approach poses a significant simulation performance issue: as shown in Fig. 5, diffusion events (1) and (2) would basically compensate for each other and do not lead to any significant surface energy changes. This leads to severe simulation inefficiency due to frequently repeating self-compensating events, especially when the surface diffusion rate is high. This is a well-known issue of the KMC method and several attempts have been made to improve simulation performance for such systems, for instance, by the introduction of long range jumps [12].

To achieve real surface energy relaxation, the redistribution of surface atoms near kinks and steps is needed (reaction (3) in Fig. 5). To emulate long-range fast diffusion without jeopardizing the highly efficient local bond-counting

model we have developed a two-stage reaction: first, an atom “evaporates” (4) from the surface, transitioning to a virtual state, and then it is deposited based on the neighbor counting binding energy of the final site (5). This approach allows us to retain the efficient rejection-free KMC algorithm (see, e.g. [13]), while improving the system evolution rate by orders of magnitude.

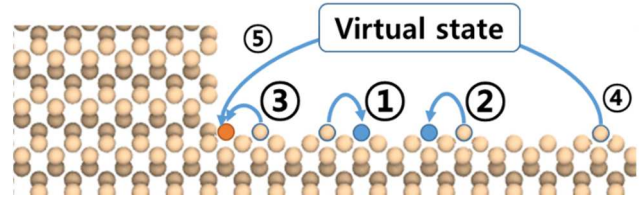


Fig. 5. Reflow simulation illustrating a chain of surface diffusion events for nearly flat surfaces: reactions (1) and (2) compensate for each other, leading to negligible surface changes at the expense of two KMC simulation steps. Only the rare reaction (3) is efficient. Fast surface diffusion can be mimicked by virtual “desorption” (4) – “adsorption” (5) reaction pairs simulated as a single KMC step, which are not limited to close neighbor sites.

An example of a simulation for a structure with rectangular defects at the sides of a trench is shown in Fig. 6. The initial continuum input of the KMC simulation (Fig. 6 left) is followed by the views of the central slice of the structure at the early, intermediate and final stages of the reflow process. Once the structure reaches an equilibrium surface shape (defined by the model parameters) it stops changing after some time (except for atomic noise). To improve the performance of the simulation in such cases we have developed a large number of functions that can extract structure features on-the-fly during the simulation and stop it, trigger the structure output routines, etc.

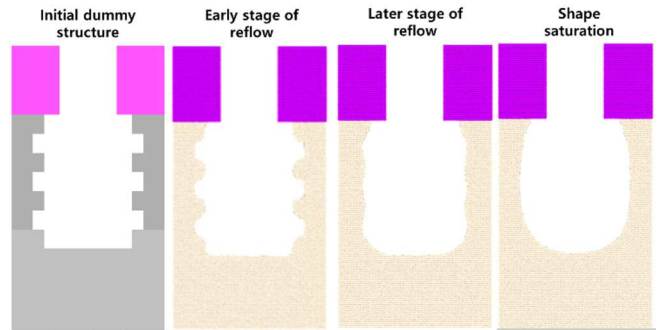


Fig. 6. Several stages of the reflow process for reduction of side wall artifacts. Depending on the model parameters, the final shape of the structure can have a different aspect ratio.

III. EPITAXY FOR NON-PLANAR DEVICES

Effective growth rates for the structures presented in Fig. 2 were analyzed using typical model parameters and epitaxial process conditions. The growth rate (defined for a complex geometry structure as the number of atoms deposited per unit of time) shows complex temporal evolution (Fig. 7) due to strong faceting during growth. Initially, the wide device shows a much higher rate due to its large open area and reduced {111} facet formation, which dominates the early stages of narrow Fin growth (Fig. 7, point 1).

Spikes in the growth rate (points 2 and 3 for wide and narrow devices, respectively) are attributed to S/D merging, which can lead to extremely fast growth, while the final saturation of the shape (point 4) reflects the formation of a

{111} facet-dominated diamond structure. Small fluctuations in the growth rate have several sources: first, they correspond to an increased deposition rate when steps propagate along the {111} facets; second, they arise from a suboptimal KMCRay flux update frequency that leads to a small reduction in the overall epitaxy rate if the layer grown between updates is too thick. The latter can be mitigated by more frequent flux statistics updates.



Fig. 7. Effective deposition rate for narrow and wide Fin devices. Key phases of epitaxy are indicated by numbers: 1 – initial stage limited by {111} faceting, 2, 3 – growth rate spikes during structure sidewalls merging for wide and narrow fin devices, respectively; 4 – saturated growth after the formation of a diamond-like shape.

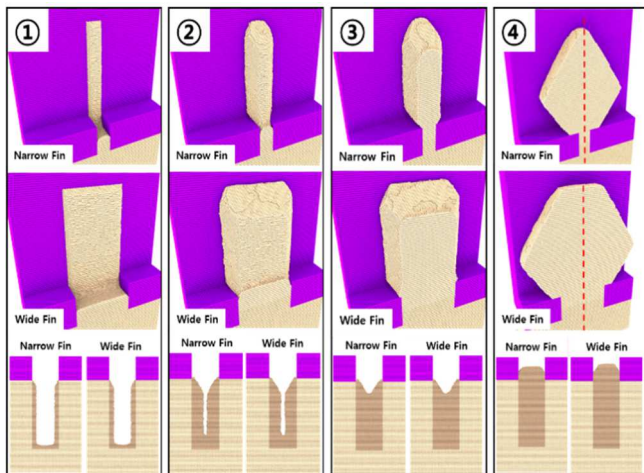


Fig. 8. Snapshots of the simulation corresponding to points 1, 2, 3, and 4 in Fig. 7. At the early stage of growth, the narrow FinFET device becomes limited by facets (1) and the overall growth rate is much lower than that of the wide device. This trend persists until just before the merging of the sidewalls (2). After merging (3), the growth rate trends converge, as both are dominated by {111} facets.

IV. CONCLUSIONS

A unified LKMC framework, combined with a high-efficiency precursor-molecule tracing module, successfully captures coupled deposition, etching, and reflow process phenomena in FinFET- and GAA-type structures. The reflow process, modeled as a two-stage “evaporation-deposition” reaction, enables realistic surface-energy relaxation without compromising the rejection-free KMC efficiency.

As the modern semiconductor industry has shifted from geometric scaling to continuous device architecture innovation, it is essential to be able to predict and evaluate epitaxial shapes early in the process development cycle and

identify potential epitaxy process problems across different device architectures. The framework we have developed can pinpoint key differences in epitaxial growth behavior during the transition from FinFET to advanced GAA devices, thereby providing opportunities for the optimization of next-generation semiconductor nodes. Despite the empirical nature of KMC models, once basic calibration is performed, the framework can yield not only qualitative but also quantitatively accurate results that can guide process recipe development.

REFERENCES

- [1] G. Bae et al., “3nm GAA technology featuring multi-bridge-channel FET for low power and high performance applications,” in Proc. 2018 IEEE Int. Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2018, pp. 28.7.1–28.7.4, doi: 10.1109/IEDM.2018.8614629
- [2] M. Kolahdouz, L. Maresca, R. Ghandi, A. Khatibi, and H. H. Radamson, “Kinetic model of SiGe selective epitaxial growth using RPCVD technique,” J. Electrochem. Soc., vol. 158, no. 4, pp. H457–H464, 2011, doi: 10.1149/1.3548113.
- [3] S. Chen et al., “A level set method for thin film epitaxial growth,” J. Comput. Phys., vol. 167, no. 2, pp. 475–500, 2001, doi: 10.1006/jcph.2000.6689.
- [4] N. Zographos, C. Zechner, I. Martin-Bragado, K. Lee, and Y.-S. Oh, “Multiscale modeling of doping processes in advanced semiconductor devices,” Mater. Sci. Semicond. Process., vol. 62, pp. 49–61, 2017, doi: 10.1016/j.mssp.2016.10.037.
- [5] S. Chopra et al., “Epitaxial growth on high aspect ratio structures,” ECS Trans., vol. 33, no. 6, pp. 977–984, 2010.
- [6] R. Chen, W. Choi, A. Schmidt, and Y. Park, “A new kinetic lattice Monte Carlo modeling framework for the source-drain selective epitaxial growth process,” in Proc. Int. Conf. Simul. Semicond. Process. Devices (SISPAD), Glasgow, Scotland, 2013, pp. 9–12, doi: 10.1109/SISPAD.2013.6650561.
- [7] I. Martin-Bragado, R. Borges, J. P. Balbuena, and M. Jaraiz, “Kinetic Monte Carlo simulation for semiconductor processing: a review,” Prog. Mater. Sci., vol. 92, pp. 1–32, 2018, doi: 10.1016/j.pmatsci.2017.09.003.
- [8] H. Kubotera et al., “High-fidelity and efficient epitaxial growth simulations using hybrid meshing and BVH-based ray tracing,” in Proc. Int. Conf. Simul. Semicond. Process. Devices (SISPAD), Grenoble, France, 2025, doi: 10.1109/SISPAD66650.2025.11186355.
- [9] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, “A new algorithm for Monte Carlo simulation of Ising spin systems,” J. Comput. Phys., vol. 17, no. 1, pp. 10–18, 1975, doi: 10.1016/0021-9991(75)90060-1.
- [10] H. H. Yoon et al., “Enabling the Angstrom Era: 2D material-based multi-bridge-channel complementary field effect transistors,” npj 2D Mater. Appl., vol. 9, Art. no. 68, 2025, doi: 10.1038/s41699-025-00591-z.
- [11] A. Schmidt et al., “Simulation method and system,” U.S. Patent 11,010,532, May 18, 2021.
- [12] I. Martin-Bragado, N. Zographos, and M. Jaraiz, “Long and double hop kinetic Monte Carlo: techniques to speed up atomistic modeling without losing accuracy,” Mater. Sci. Eng. B, vols. 154–155, nos. 1–3, pp. 202–206, 2008, doi: 10.1016/j.mseb.2008.08.015.
- [13] T. P. Schulze, “Efficient kinetic Monte Carlo simulation,” J. Comput. Phys., vol. 227, no. 4, pp. 2455–2462, 2008, doi: 10.1016/j.jcp.2007.10.021