2024 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD) | 979-8-3315-1635-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/SISPAD62626.2024.10733099

Equipment-Informed Machine Learning-Assisted Feature-Scale Plasma Etching Model

Lado Filipovic[1,2]*, Tobias Reiter[1,2] Julius Piso[2], and Roman Kostal[2]
[1]Christian Doppler Laboratory for Multi-Scale Process Modeling of Semiconductor Devices and Sensors
[2]Institute for Microelectronics, TU Wien, 1040 Vienna, Austria
*Email: filipovic@iue.tuwien.ac.at

***Abstract*—We investigate means to merge feature-scale and reactor-scale models during plasma etching using Machine Learning (ML) and interpolative approaches. First, we test the SF_6 plasma etching models based on a small dataset from literature. We find that Gaussian Process Regression (GPR) leads to significant over-fitting, resulting in waviness in the predicted values in the range where no data is available. A Neural Network (NN) model was likewise implemented with rectified linear unit activation. This model provides linear prediction between known values, resulting in a better qualitative fit to experimental observations. Finally, we perform 18 750 chamber simulations of a Cl_2/Ar plasma while varying relevant input parameters. The data is used to build a six-dimensional spline-based inerpolative model of the chamber and provide a means to quickly extract relevant fluxes for the feature-scale model.**

***Index Terms*—process simulation, plasma etching, machine learning, interpolation, Gaussian process regression, neural networks**

I. Introduction

Plasma etching is a fundamental step during complementary semiconductor-metal-oxide (CMOS) fabrication [1] and performing technology computer aided design (TCAD) simulations for this process provides insights into plasma behavior, enables optimization of process parameters, predicts etch profiles, and reduces cost and time during development [2]. However, the limitation of process TCAD is that it concentrates on the feature scale and the link to equipment conditions is often missing or is difficult to introduce [3]. We test several approaches to merge equipment information into the feature-scale model so that an effective digital twin of the plasma chamber is the input to the feature-scale model. The implementation is based on an established and calibrated model for silicon etching in an inductively-coupled plasma (ICP) with SF_6/O_2 gas, as described by Belen et al. [4], [5] and Bobinac et al. [6]. For these, the main parameters to the physical model are the fluxes of neutral radicals and ions. Experiments are performed to find how these fluxes change while the plasma chamber inputs are varied with respect to the pressure p in mTorr, O_2 fraction in the feed (y_{O_2}), and applied bias V_b in V. The inductive coil power (800 W), temperature (5 °C), and total gas flow rate (80 sccm) are kept constant.

II. Methods

The feature-scale plasma etching model assumes ballistic transport of particles and Langmuir-type adsorption [7]. In the chamber, the particle motion is primarily governed by inter-particle interactions, as visualized in Fig. 1. The blue circles represent neutral radicals which diffuse in the chamber and whose direction of motion is represented by a cosine distribution at the feature-scale. The red circles represent ions, which are accelerated in the sheath potential and whose motion is represented by a more directional power cosine distribution.

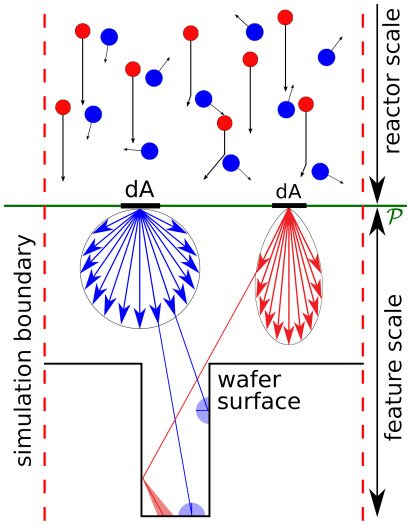


Fig. 1. The typical simulations scales when modeling plasma etching. The reactor scale tracks particle-particle interactions, while the feature scale treats particle motion as rays and only wafer surface intersections are recorded.

The equations used to calculate the steady-state surface coverages of fluorine θ_F and oxygen θ_O for each time step are given by

$$\frac{1}{\theta_F} = 1 + \left(\frac{k\sigma_{Si} + 2Y_{ie}\Gamma_i}{\gamma_F\Gamma_F}\right)\left(1 + \frac{\gamma_O\Gamma_o}{\beta\sigma_{Si} + Y_O\Gamma_i}\right) \quad (1)$$

$$\frac{1}{\theta_O} = 1 + \left(\frac{\beta\sigma_{Si} + Y_{ie}\Gamma_i}{\gamma_O\Gamma_O}\right)\left(1 + \frac{\gamma_F\Gamma_F}{k\sigma_{Si} + 2Y_{ie}\Gamma_i}\right), \quad (2)$$

where $\Gamma_{F/O/i}$ are the fluxes of F, O, and ion species, k is the chemical etch rate constant, β is the oxygen recombination rate, σ_{Si} is the surface site density of Si, $Y_{O/ie}$ are the O and ion-enhanced etch yields, and $\gamma_{F/O}$ are the F and O sticking coefficients. The surface etch rate is then calculated using

$$\mathrm{ER} = \frac{1}{\rho_{Si}}\left(\frac{k\sigma_{Si}\theta_F}{4} + Y_p\Gamma_i + Y_{ie}\Gamma_i\theta_F\right), \quad (3)$$

where ρ_{Si} is the density of the Si substrate. This feature-scale model for SF_6/O_2 plasma etching is implemented in ViennaPS, an in-house level set (LS) based process simulator [8], while the accumulation of particles along the different locations of the surface is solved using Monte Carlo ray tracing with ViennaRay [9]. Therefore, the feature scale model takes the particle fluxes as input, along with an oxygen proportionality constant A_O. These values are obtained from a calibrated plasma chamber model, where the equipment inputs, such as pressure, oxygen proportionality, and bias voltage are varied [4], [5]. The data is provided in Table I.

TABLE I
SUMMARY OF THE CALIBRATED DATA FROM BELEN ET AL. [4], [5] FOR THE FLUXES AND THE OXYGEN PROPORTIONALITY CONSTANT A_O, USED TO TRAIN THE MACHINE LEARNING (ML) MODEL FOR SF_6/O_2 PLASMA ETCHING. Y_O IS DIRECTLY PROPORTIONAL TO A_O.

p	y_{O_2}	V_b	Γ_F	Γ_O	Γ_i	A_O
25	0.4375	120	5500	1500	10	3
25	0.5000	120	5000	2500	10	3
25	0.5625	120	4500	6000	10	2
25	0.6250	120	4000	10000	10	1
10	0.5000	20	2000	1000	20	3
25	0.5000	20	5000	2500	10	3
40	0.5000	20	7500	6000	5	3
5	0.0000	120	1500	10	20	5
25	0.0000	120	6000	20	8.5	4
75	0.0000	120	12000	80	2.5	3

The biggest challenge in providing a feature-scale model which is linked to the equipment conditions is ensuring that the fluxes of the different species can be represented as a function of the chamber inputs. This can be achieved using chamber-level simulations [10] which would likewise need to be calibrated to the physical behavior. Instead, we attempt to use the calibrated model and apply a ML approach to link the chamber inputs (p, V_b, y_{O_2}) and outputs ($\Gamma_{F/O/i}$, A_O), cf. Fig 2. We test Gaussian Process Regression (GPR) and Neural Network (NN) implementations to assess the benefits and limitations of each approach on the relatively small data set.

A. Gaussian Process Regression

The benefit of GPR is that it provides flexibility, small data efficiency, and uncertainty estimation [11]. This should allow us to assess, based on the uncertainty, where more data is required to improve the model. With GPR, the prior distribution of model parameters is updated using observed data while Gaussian distributions are used for the parameters and training error. Bayes' rule is used to calculate the subsequent distribution of parameters, resulting in a Gaussian predictive outcome function. While the method comes with a computational complexity of the order N^3, with N being the number of training samples, this is not a significant concern when running small data sets, as we have here. The model uses a squared exponential kernel $k(x, x') = \sigma^2 \exp\left(-\gamma \|x - x'\|^2\right)$ with training parameters γ and σ^2.

In order to at least roughly assess the quality of the model in predicting values between those in the trained range, we plot the results of a system where the pressure p and bias voltage V_b are kept constant at 25 mTorr and 120 V, respectively, and only the ratio between the O_2 and SF_6 flows is varied y_{O_2}. The plots are shown in Fig. 3(a). Intuitively, it should be clear that increasing the oxygen content while keeping all other parameters the same, including the total gas flow rate, would result in an increased oxygen flux and reduced fluorine flux at the wafer surface. However, due to the Gaussian nature of the fitting and oversampling, we note waviness and an increased oxygen flux when the oxygen fraction is at about 0.25. This is unrealistic and requires that a new measurement is carried out at this point. A new measurement provides significantly improved results and uncertainty estimates, as shown in Fig. 3(b).

B. Neural Network

We then also implement a NN model [12] on the same data set using an input layer with 3 neurons corresponding to p, V_b, and y_{O_2} and a hidden layers with 12 nodes employing rectified linear unit activation. The output layer consists of 4 neurons with no activation corresponding to the F, O, and ion fluxes and A_O. Mean Squared Error (MSE) loss is used to measure the difference between the predicted and trained values. The results of the training for the fluoride flux are given in Fig. 4. When we visualize the data points where the pressure and bias voltage are fixed ($p = 25$ mTorr, $V_b = 120$ V) while y_{O_2} is varied, it is clear that there is a linear prediction between the training points. This prediction is a more realistic one than what we observed with the GPR since it qualitatively provides an increasing oxygen flux at the wafer surface as a consequence of an increased oxygen flow in the chamber.

C. Spline Interpolation

We also simulate a multi-scale Cl_2/Ar plasma etch process, from the reactor- to the feature-scale. A wide range of input conditions is modeled at the chamber-scale, while varying the chamber parameters such as coil power P (W), gas flow rate Q_f (sccm), pressure, Cl_2 fraction in the feed, temperature T, and bias voltage V_b, cf. Fig. 5. In total, 18 750 different combinations of the chamber input parameters were simulated, shown in Table II. The plasma is generated by first applying electric power to the coils, creating a strong magnetic field inside the chamber, thereby inducing high-energy electrons. The gases, consisting of Cl_2 and Ar are introduced into the chamber through a nozzle at a specified flow rate. Collisions between the gases and high-energy electrons result in several possible reactions. The plasma chemistry for this study includes electron impact reactions derived from the electron momentum transfer cross sections provided by the SIGLO database [13], released as part of the Plasma Data Exchange Project and the LXCat website [14].

The etch rate and Cl surface coverage were calculated using a model taking into consideration chemical etching, sputtering, and ion-enhanced etching

$$\mathrm{ER} = \frac{1}{\rho_{TiN}} \left(k \sigma_{TiN} \theta_{Cl} + Y_p \Gamma_i + Y_{ie} \Gamma_i \theta_F\right) \tag{4}$$

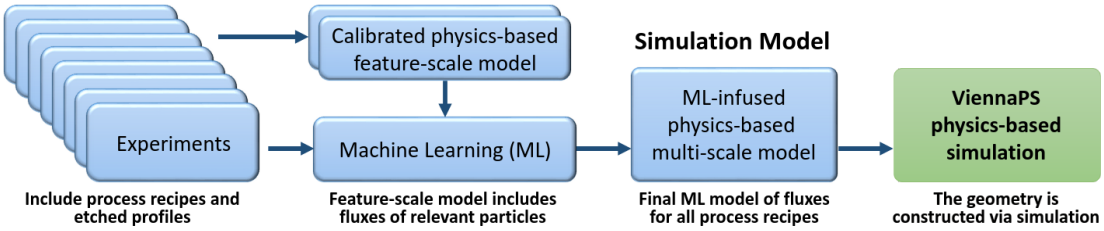
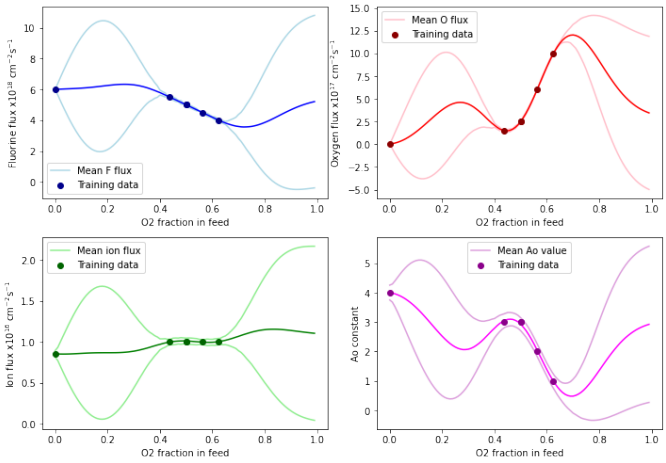
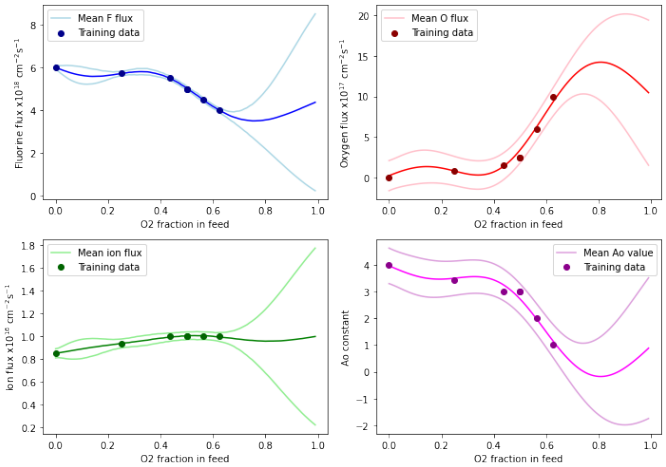


Fig. 2. Workflow used for generating the relevant particle fluxes from a combination of experiments and a calibrated feature-scale model.



(a) Results of a GPR model when varying only y_{O_2}

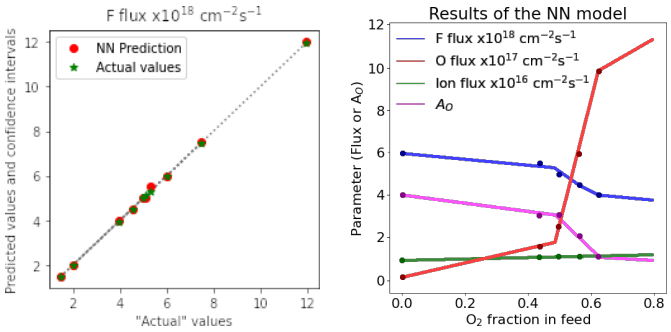


(b) Results after adding a data point at $y_{O_2} = 0.25$

Fig. 3. Results of a trained GPR model to predict feature-scale parameters from a calibrated SF_6/O_2 ICP. The training values are shown as dots, the model output as dark lines, and the uncertainty estimate ($\pm 2\sigma$) as thin light-colored lines.

$$\theta_{Cl} = \frac{\gamma_{Cl}\Gamma_{Cl}}{\gamma_{Cl}\Gamma_{Cl} + k\sigma_{TiN} + 2Y_{ie}\Gamma_i} \tag{5}$$

The results of the simulation were compiled into a hexeract - a six dimensional hypercube - where each dimension represents one varied chamber input. This data is then used for the spline interpolation, cf. Fig. 6(gray arrows). When using the model at the feature-scale, performing chamber simulations is not necessary. Instead, the inputs are directly used by



(a) NN model training fit (b) Flux predictions varying y_{O_2}

Fig. 4. NN model results after training to calibrated SF_6/O_2 ICP fluxes.

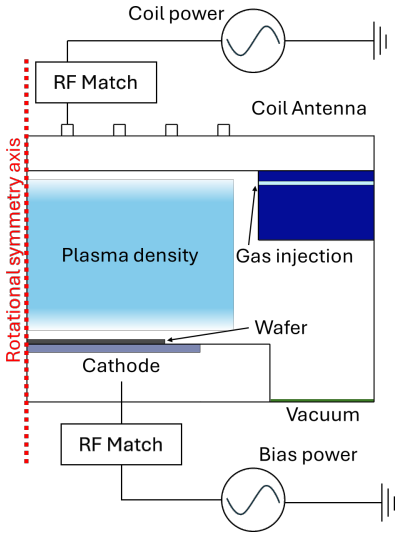


Fig. 5. Schematic diagram of the ICP system.

ViennaPS to apply the interpolated fluxes from the splines, cf. Fig. 6(green arrows). This large data set will also allow us to investigate the optimal ML algorithm without performing costly and time-intensive experiments.

III. Results and Conclusion

The provided implementation allows to perform feature-scale simulations on any generic geometry while using equipment parameters as inputs without running chamber-level simulation. As an example, the results of a cylindrical hole etch using SF_6 plasma with low and high O_2 concentration ($y_{O_2} = 0.3$ and $y_{O_2} = 0.7$, respectively) are shown in

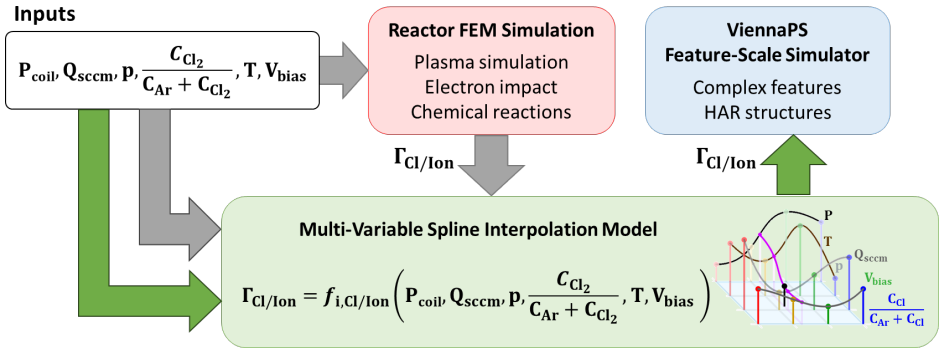


Fig. 6. Workflow of a multi-scale plasma etch model using multi-variable spline interpolation. Gray arrows show data generation and green arrows show the application of the model at feature scale.

TABLE II
RANGE OF CHAMBER INPUTS FOR THE CHLORINE/ARGON PLASMA ETCHING SIMULATION.

Parameter:	**Simulated values**				
Coil power P_{coil} **(W)**:	200	300	400	500	600
Gas flow rate Q_f **(sccm)**:	50	75	100	125	150
Pressure p **(mTorr)**:	7	10	13	16	19
Cl_2 ratio in Cl_2/Ar y_{Cl_2}:	0.55	0.65	0.75	0.85	0.95
Temperature T **(°C)**:	25	35	45	55	65
Bias voltage V_b **(V)**:	0	30	60	90	120	150

Fig 7. We show the results of a 20 s etching simulation using both GPR- and NN-trained SF_6 plasma. It is clear that the GPR and NN models predict and extrapolate the results quite differently, albeit with the same general trend. The GPR model underestimates the O flux - $44.6\times10^{15}\text{cm}^{-2}\text{s}^{-1}$ - as a consequence of the oversampling noted in Fig. 3b. The NN model provides a linear prediction for the flux, which it calculates to be $103.3\times10^{15}\text{cm}^{-2}\text{s}^{-1}$. While these values need to be tested with experiments or chamber simulations, the polynomial fitting, which is part of the GPR model, might not be well suited for such problems, when only a small sample size is available.

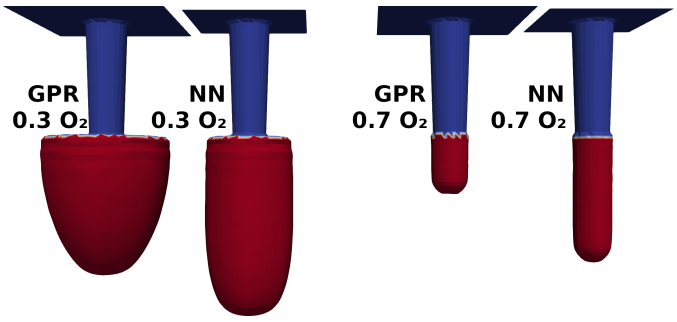


Fig. 7. SF_6 plasma etching through a cylindrical mask (blue) of diameter 35 μm, depth 120μm, and 2° taper angle. P_{coil}, Q_{sccm}, p, T, and V_b are set to 800 W, 80 sccm, 25 mTorr, 5°C, and -120 V, respectively, while y_{O_2} is set to 0.3 and 0.7.

ACKNOWLEDGMENT

The financial support by the Austrian Federal Ministry of Labour and Economy, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association, Austria is gratefully acknowledged.

REFERENCES

[1] M. Huff, "Recent Advances in Reactive Ion Etching and Applications of High-Aspect-Ratio Microfabrication," *Micromachines*, vol. 12, no. 8, p. 991, Aug. 2021.
[2] F. Krüger *et al.*, "Voltage Waveform Tailoring for High Aspect Ratio Plasma Etching of SiO_2 Using Ar/CF_4/O_2 Mixtures: Consequences of Low Fundamental Frequency Biases," *Physics of Plasmas*, vol. 31, no. 3, p. 033508, Mar. 2024.
[3] Y. G. Yook *et al.*, "Fast and Realistic 3D Feature Profile Simulation Platform for Plasma Etching Process," *Journal of Physics D: Applied Physics*, vol. 55, no. 25, p. 255202, Mar. 2022.
[4] R. J. Belen *et al.*, "Feature-Scale Model of Si Etching in SF_6 Plasma and Comparison With Experiments," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, vol. 23, no. 1, pp. 99–113, Dec. 2004.
[5] R. J. Belen *et al.*, "Feature-Scale Model of Si Etching in SF_6/O_2 Plasma and Comparison with Experiments," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, vol. 23, no. 5, pp. 1430–1439, Aug. 2005.
[6] J. Bobinac *et al.*, "Effect of Mask Geometry Variation on Plasma Etching Profiles," *Micromachines*, vol. 14, no. 3, p. 665, Mar. 2023.
[7] O. Ertl, "Numerical Methods for Topography Simulation," Doctoral dissertation, Technische Universitä t Wien, 2010.
[8] Reiter, Tobias and Karnel, Noah and Filipovic, Lado, "ViennaPS 2.0.0," 2024, [Online]. Available: https://github.com/ViennaTools/ViennaPS/releases/tag/2.0.0 (Accessed 10.06.2024). [Online]. Available: https://github.com/ViennaTools/ViennaPS
[9] Reiter, Tobias and Klemenschits, Xaver and Karnel, Noah and Filipovic, Lado, "ViennaRay 2.1.0," 2024, [Online]. Available: https://github.com/ViennaTools/ViennaRay/releases/tag/v2.0.1 (Accessed 10.06.2024). [Online]. Available: https://github.com/ViennaTools/ViennaRay
[10] B. Gul *et al.*, "Numerical Study of SF_6/O_2 Plasma Discharge for Etching Applications," *Plasma Chemistry and Plasma Processing*, vol. 41, no. 4, pp. 1223–1238, Mar. 2021.
[11] J. Wang, "An Intuitive Tutorial to Gaussian Process Regression," *Computing in Science & Engineering*, vol. 25, no. 4, pp. 4–11, Jul. 2023.
[12] A. Paszke *et al.*, "Pytorch: An Imperative Style, High-Performance Deep Learning Library," in *Proc. 33rd Conference on Neural Information Processing Systems (NeurIPS)*, 2019, pp. 1–12, (Accessed 10.06.2024). [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf
[13] SIGLO database, "LXCat," 2023, [Online]. Available: https://fr.lxcat.net/home/ (Accessed 10.06.2024).
[14] E. Carbone *et al.*, "Data Needs for Modeling Low-Temperature Non-Equilibrium Plasmas: The LXCat Project, History, Perspectives and a Tutorial," *Atoms*, vol. 9, no. 1, p. 16, Feb. 2021.