

Autonomous Hybrid Optimization of a SiO₂ Plasma Etching Mechanism

Florian Krüger¹, Du Zhang², Pingshan Luan², Minjoon Park², Andrew Metz² and Mark J. Kushner^{1,3}

- ¹ Electrical Engineering and Computer Science Department, University of Michigan, 1301 Beal Ave., Ann Arbor, MI 48109-2122 USA fkrueger@umich.edu, mjkush@umich.edu
- ² TEL Technology Center, America, LLC, 255 Fuller Road, Suite 214, Albany, NY 12203 USA du.zhang@us.tel.com, pingshan.luan@us.tel.com, minjoon.park@us.tel.com, andrew.metz@us.tel.com
- ³ Author to whom correspondence should be addressed.

Abstract:

Computational modeling of plasma etching processes at the feature scale relevant to the fabrication of nanometer semiconductor devices is critically dependent on the reaction mechanism representing the physical processes occurring between plasma produced reactant fluxes and the surface, and the reaction probabilities, yields, rate coefficients and threshold energies that characterize those processes. The increasing complexity of the structures being fabricated, new materials and novel gas mixtures for plasma produced radical fluxes to the wafer also increase the complexity of the reaction mechanism used in feature scale models, and the difficulty in developing the fundamental data required for the mechanism. This challenge is further exacerbated by the fact that acquiring these fundamental data through more complex computational models or experiments is often limited by cost, technical complexity or inadequate models. In this paper, we discuss a method to automate the selection of fundamental data in a reduced reaction mechanism for feature scale plasma etching of SiO₂ using a fluorocarbon gas mixture by matching predictions of etch profiles to experimental data using a gradient descent / Nelder-Mead method hybrid optimization scheme. These methods produce a reaction mechanism that replicates the experimental training data as well as experimental data using related but different etch processes.

Keywords: plasma etching, plasma surface interaction, computational modeling, model optimization, machine learning, physics informed optimization

I. Introduction

The computational modeling of plasma etching processes is receiving new interest due, in part, to its ability to aid in the design and understanding of semiconductor device manufacturing and to provide the training data for machine learning based optimization of processes [1–5]. In simulation of plasma etching, the framework typically includes a reactor scale model and a feature scale model [6–8]. The reactor scale model provides reactive fluxes to the wafer, including their energy and angular distributions, that are used as input to the feature scale model [9][10]. There are several classes of feature scale models – level-set-methods [6,11–13], statistical voxel approaches [14–16], and molecular dynamics (MD) [17,18]. Although MD methods are the most fundamental, they are currently computationally limited in the size of the feature that can be simulated and by availability of inter-particle potentials for complex chemistries. As a result, most full-feature scale capable models employ level set or statistical voxel methods with, in most cases, less fundamental reaction mechanisms.

Due to limits resulting from computational complexity and sometimes incomplete understanding of the physical process involved in plasma etching, feature profile simulations typically make use of simplified or reduced reaction mechanisms. These mechanisms use rate coefficients, reaction probabilities, threshold energies and energy and angular scaling to represent reactions between gas phase and surface species. The outcome of these reactions are the addition of material (deposition), removal of material (etching or sputtering) and modification of the material (passivation, implantation) [15,19,20]. The reaction mechanism, and the coefficients and parameters that quantify the mechanism, are ideally based on either experiments or more complex computational models that do not allow for an efficient real-time evaluation.

Rapidly and accurately producing reaction mechanisms for profile simulation for novel materials, gas mixtures and temperature regimes, as in cryogenic etch processes [21–23], is challenging. This is particularly the case in data starved regimes, which is often the case when addressing new materials. So there is a need to robustly and efficiently adjust and improve reaction mechanisms. Classical optimization techniques as well as machine learning (ML) based approaches are areas of active research [4,5,24–26].

Developing and applying ML techniques for process design is an expanding area of research with the goal of addressing the increasing number of process parameters, complexity of the processes and non-linear behavior. Model-based feedback has enabled precise, repeatable and

stable process control [27–29]. As the control models face some of the same issues with respect to data availability and consistency as in profile evolution, control methods are now also based on machine learning methods which efficiently allow for rapid real time evaluation [30–34]. Other applications of ML include the development of surrogate methods where statistical models are trained on experimental or computed data to simulate a process as a whole or a sub-process [35,36]. Such practices being applied to large scale simulations of fusion plasmas [37–43]. Ideally, these models applied in semiconductor manufacturing are capable of reproducing reactor and feature scale etch processes either directly as an outcome of the surrogate model or by interpolating between known results [44,45]. These types of approaches have underlying model structures that are typically valid in their training parameter space. However, the parameters that are the outcome of the underlying model, while providing a solution, do not necessarily provide physical insight. Many of these data driven ML approaches are themselves fundamentally limited by the availability or accuracy of the data used for training purposes. Producing this data is itself costly and time consuming, while the resulting trained ML based model may have limited applicability to process conditions outside the training set.

The challenge of needing large training sets of data has led to the use of physics informed ML and optimizations techniques. Using these techniques, lack of data is, to some degree, mitigated by the explicit use of physical formulations and models to reduce the degrees of freedom in the ML-derived model. Using this approach, a reaction mechanism can be built using known forms of probabilities of gas-surface interactions which may have poorly known coefficients. Using ML-methods to derive the physics-based coefficients should, in principle, produce a reaction mechanism that extends beyond the conditions of the training set.

These challenges specifically apply to the modeling of high aspect ratio (HAR) plasma etch processes [46–49]. The aspect ratio (AR) is the ratio of the etch depth and to the width of the feature, often given by the opening of the mask material. These processes are relevant to the fabrication of vias, contact holes and isolation trenches [48]. Features are now produced for 3-dimensional (3D) memory devices having AR approaching 100. Ideal HAR features have vertical side walls with widths replicating the mask with a minimum of bowing [50] or tapering [51]. There are many process-dependent challenges which make these outcomes difficult. For example, polymer deposition is a critical process in dielectric plasma etching in fluorocarbon gas mixtures [52,53]. A critical process challenge is the control polymer deposition in the upper regions of the

mask during the etch process, which may result in necking and clogging [44,54,55]. (*Necking* is the narrowing of the opening in the mask by polymer deposition. *Clogging* is closing of the mask opening by the deposition.) Both necking and clogging result from the same processes with different degrees of severity. Clogging will stop etching while necking can limit neutral gas transport further into the feature, trap etched surface species and shadow ions and photons entering into the feature. Modeling these processes, critical to feature optimization, is additionally challenged by lack of fundamental data.

Physics informed ML approaches to process development have also taken the form of compact or reduced models. For example, through ML methods, reduced reaction mechanisms can be generated [56–58]. In a reduced reaction mechanism, species or reactions for which the outcome is weakly dependent are removed from a comprehensive mechanism, resulting in a model that is computationally more efficient. Reduced reaction mechanisms, though more rapidly executing, are generally valid over a limited range of operation, such as power, gas mixture or pressure.

The optimization of the reaction mechanism itself has been investigated using Monte Carlo approaches and simulations based on cellular automata[59–63]. Optimization schemes such as evolutionary algorithms and particle swarm methods can require many evaluations of the model. This large number of evaluations can lead to a large computational expense or needing to reduce the domain over which the optimization is being performed, either spatially or in terms of interaction complexity (e.g. limiting the number of species).

In this paper, we discuss the development and application of a semi-autonomous optimization technique to derive physics-based coefficients for reactions mechanisms used in voxel-based feature simulation of plasma etching. The optimization technique couples a gradient descent method with a Nelder Mead approach. The demonstration system is SiO₂ HAR plasma etching in a fluorocarbon gas mixture using a capacitively coupled plasma. The reactor scale simulations that provide reactant fluxes were performed with the Hybrid Plasma Equipment Model (HPEM). The feature scale simulations were performed with the Monte Carlo Feature Profile Model (MCFPM).

A high-level overview of the optimization scheme is given in Sec. II. The process to be optimized as well as the simulation tools used for the gas phase and feature scale simulation are discussed in Secs. III and IV. Metrics for the optimization process are defined in Sec. V. The optimization schemes are described in Secs. VI and VII where the hybrid approach is motivated.

Results of the combined optimization method are discussed in Sec. VIII. To test the transferability of the derived mechanism, etch processes outside the original training regime were simulated, with those results being presented in Sec. IX for a single feature and for an array of features in Sec. X. Concluding remarks are in Sec. XI.

II. Description of the optimization scheme

The goal of this work is to optimize a reaction mechanism that describes plasma-surface interactions for SiO₂ etching that best replicates experimental data. The experimental data was provided in the form of scanning electron microscopy (SEM) images. The optimization is centered on adjusting physics parameters in the reaction mechanism, represented by scalar values p_i , each representing a physical quantity, contained in a parameter set of dimension n ,

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}. \quad (1)$$

The size of the parameter space scales like $O(c^n)$, where c is a constant, making exhaustive search algorithms impractical for even moderately large numbers of parameters. Enforcing constraints and relations based on prior physical knowledge of the likely range of p_i can help to reduce the total parameter space and enable the use of conventional optimization techniques. This physical knowledge is, in our case, implemented by using a well-established surface reaction mechanism having poorly known reaction probabilities, and applying physically reasonable bounds to those poorly known coefficients.

The optimization scheme is based on a hybrid approach, combining a gradient descent method [5,64,65] with a Nelder-Mead optimization scheme [66–68]. A top-level overview of the method is shown in Fig. 1. The optimization setup requires external input (located on the left-hand side of Fig. 1) which in this case are the reactant fluxes of radicals and ions and their energy and angular distributions (EADs) that are incident onto the wafer being processed. These fluxes and EADs were provided by the HPEM using reactor conditions that produced the features imaged by the SEMS. The fluxes and EADs provided by the HPEM are here considered as being *ground truth*. A legitimate issue is the sensitivity of the optimized feature scale reaction mechanism to the accuracy of the fluxes and EADs provided by the HPEM. The HPEM, as for all modeling and

simulation approaches, is subject to approximations and simplifications. This applies to its fluid approach and finite grid size as well the input data in terms of reaction rates, cross sections and material coefficients. Although the quantitative validity of the HPEM and related approaches in general is an important area of discussion, it is

beyond the scope of this particular investigation which is focused on developing the feature scale reaction mechanism. As such, the fluxes and EADs from the HPEM are treated as being *true* for the purposes of this discussion which is intended to be a proof of concept of the optimization process itself. That said, the results presented below will highlight the ability of HPEM and MCFPM in conjunction to replicate relevant process trends. Details of the experiment and gas phase simulation are provided in Sec. III.

The gas phase fluxes and EADS are used as input to the MCFPM which performs the etch process simulations outlined in Sec. IV. The results of these simulations are compared to the features in the SEM images. Based on a set of geometric metrics, discussed in Sec. V, a model error is determined. The data produced by the MCFPM and their differences to experimental data effectively act as the loss function to be minimized. The model error is used as the basis for the optimizer which iteratively adjusts the model parameters $\mathbf{p}$ to minimize the loss. This loop is repeated until the system converges.

III. Reactor Scale Plasma Simulation

The reactor scale, gas phase simulations were performed using the Hybrid Plasma Equipment Model (HPEM), which combines fluid and kinetic concepts to model a diverse set of plasma processes. The HPEM is described in detail in Refs. [69–71] and so will be only briefly described here.

The HPEM is a 2-dimensional plasma hydrodynamics model that employs both kinetic and fluid formulations to simulate low temperature plasmas using a time-slicing approach. A range of

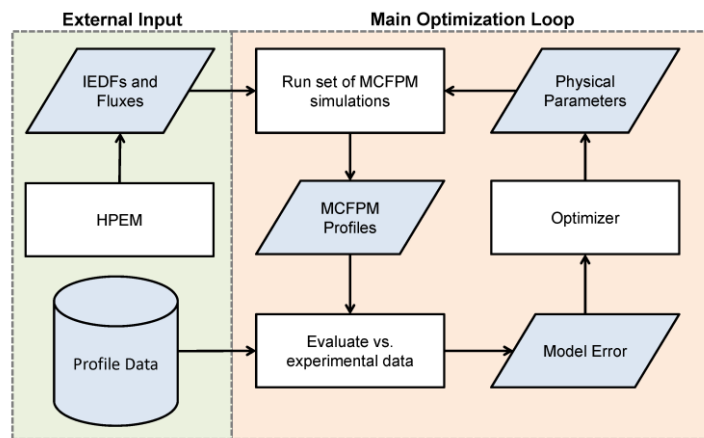


Figure 1: Global schematic of the optimization process.

physics regimes are addressed in individual modules. These modules are coupled through the exchange of physical quantities – such as electric and magnetic fields, species densities and rate coefficients. The modules used in this work are the Fluid Kinetics-Poisson Module (FKPM), the Electron Energy Transport Module (EETM), and the Plasma Chemistry Monte Carlo Module (PCMCM).

In the FKPM, the continuity, momentum, and energy equations of heavy particles are integrated in conjunction with Poisson's equation, yielding heavy particle densities, fluxes, temperatures and electrostatic potential. Electron transport is determined by solving the continuity and momentum conservation equations. In the EETM, electron energy distributions (EEDs) as a function of spatial position are obtained using a kinetic, Monte Carlo method based on the electric fields produced in the FKPM. The energy and angular distributions (EAD) of electrons striking the substrate are also recorded in the EETM. Based on the bulk plasma electron distributions, their transport and rate coefficients are produced and made available to the other modules.

After reaching a (quasi) steady-state, the PCMCM calculates the heavy particle EADs incident onto surfaces by tracking their trajectories kinetically and treating gas phase collision using a Monte Carlo approach. These distributions are recorded and then used as input to the MCFPM.

The HPEM was used to model a capacitively coupled plasma (CCP) sustained in a $C_4F_6/Ar/O_2$ gas mixture at 10 mTorr. In the base case the total gas flows were $C_4F_6/Ar/O_2 = 140/100/105$ sccm. The reactor configuration is shown in Fig. 2. Two electrodes with a radius of 15 cm are separated by a 4 cm gap. Dual radio frequency (RF) power V_{RF} is applied to the bottom electrode. The voltage waveform consists of low and high frequency components, $V_{RF}(t) = V_{LF}$

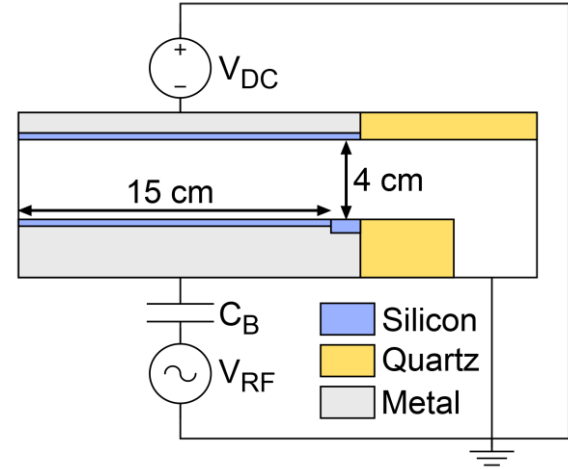


Figure 2: CCP reactor geometry and circuit.

Table 1: Base Case Fluxes to Wafer

Species	Flux ($\text{cm}^{-2}\text{s}^{-1}$)
C_3F_4	9.5×10^{16}
C_2F_3	6.8×10^{16}
CF	4.4×10^{16}
CF_2	9.4×10^{16}
CF_3	8.4×10^{15}
O	7.7×10^{16}
Ions	1.2×10^{16}

$\sin(2\pi f_{LF}) + V_{HF} \sin(2\pi f_{HFT})$ with $f_{LF} = 1$ MHz and $f_{HF} = 40$ MHz. For the base case, the low frequency voltage, V_{LF} and high frequency voltage V_{HF} , were adjusted to supply powers of $P_{LF} = 8.0$ kW and $P_{HF} = 2.5$ kW. A negative DC bias of -500 V was applied to the top electrode, to deliver 650 W. The gas phase reaction mechanism is based on previous work [15,49] and contains 87 neutral, excited and ion species and 1500 reactions.

The spatial distributions of the electron density $[e]$, electron temperature, T_e , negative fluorine ion density $[F^-]$, ionization by bulk electrons S_b , and ionization by secondary electrons S_s are shown in Fig. 3. Due to the large high frequency power which dominantly heats electrons, the electron density is high, with a maximum value of $9.5 \times 10^{10} \text{ cm}^{-3}$. The bulk electron temperature, shown for electron densities $> 1.0 \times 10^9 \text{ cm}^{-3}$, is 3.4 to 3.8 eV over the wafer. High energy electrons are initially produced by secondary electron emission by ion bombardment, followed by acceleration in the sheaths. The electron source generated by bulk

electrons S_b , mirrors that of $[e]$ and T_e . Most of the ionization is produced by bulk processes. The ionization produced by sheath accelerated beam electrons play a secondary but nevertheless important role in the overall ionization dynamics. Fluorine and oxygen containing plasmas can have significant densities of negative ions. The density of F^- , shown in Fig. 3c, has the largest density of negative ion in this process with a maximum density of $2.2 \times 10^{10} \text{ cm}^{-3}$. Due to the plasma having, on the average, a positive electric potential, the F^- ions are confined to the center of the plasma. The negative ions have only a small overall effect on plasma transport dynamics to the

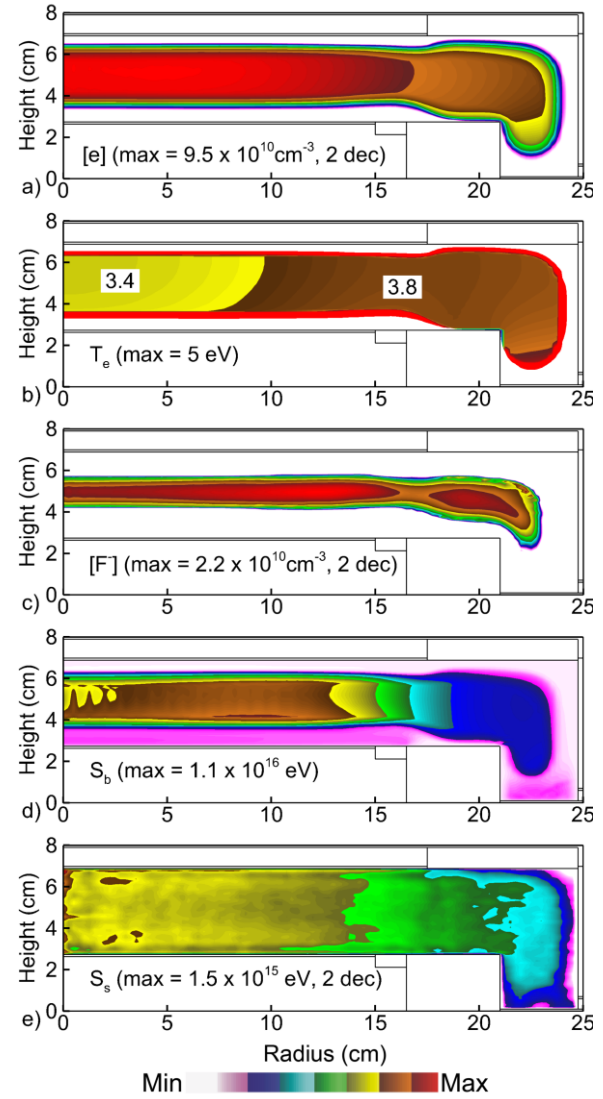


Figure 3: Plasma properties for the base case: (a) Electron density $[e]$, (b) electron temperature T_e , (c) negative fluorine ion density $[F^-]$, (d) ionization rate by bulk electrons and (e) ionization rate by secondary electrons.

wafer due to its negligible density close to the sheath region where its effect on electron heating dynamic would be the most important.

Plasma-surface interactions result from the fluxes of reactive species and ions, and their energy and angular distributions onto the surface. The fluxes of the most significant neutral and ion species incident onto the wafer at radius of 7.5 cm are listed in Table 1. These fluxes include the fluorocarbon radicals most responsible for polymer deposition as well as atomic oxygen which etches and removes the polymer. The ratio of the flux of ions and neutrals to the wafer is an important parameter in the overall process dynamics. The anisotropy of the process is enabled by the directionality of ions incident onto the surface whereas the fluxes of neutral fluorocarbon radicals determine passivation. This is especially true for HAR processes in which the overall performance requires a precise balance of deposition and removal of polymerizing radical fluxes, both of which are either indirectly or directly dependent on the neutral as well as ion fluxes.

In addition to the magnitude of the ion flux, the influence of energetic positive ion fluxes incident onto the surface and their transport through HAR features is sensitive to their energy and angular distribution (IEAD). The combined IEAD of all positive ion species onto the wafer is shown in Fig. 4a. Due to the large applied low frequency power, the sheath potential and commensurate DC self-bias, positive ions are accelerated vertically into the wafer while traversing the sheath. This leads to a narrow angular distribution of ions striking the wafer with energies up to 4800 eV, both properties being desirable in the context of HAR etch processes.

IV. Feature Scale Model

The process investigated in this work is a HAR plasma etch into SiO₂ using an amorphous carbon (AC) mask, schematically shown in Fig. 4b. A SiO₂ substrate is covered by a 850 nm thick

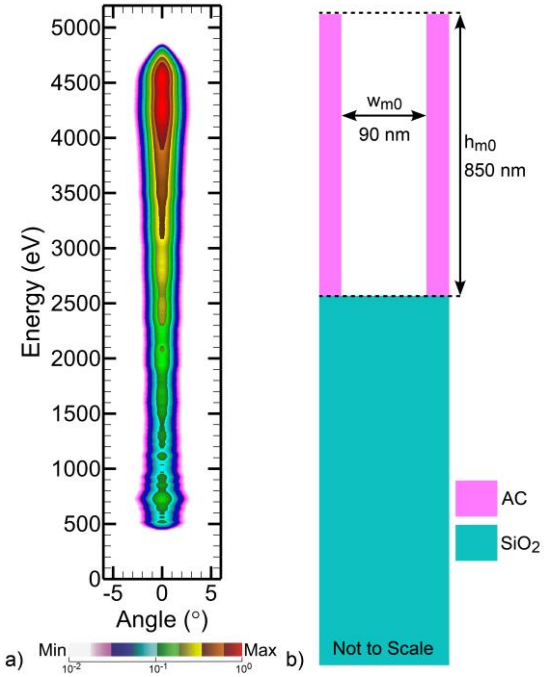


Figure 4: Inputs for feature profile simulation: a) Ion energy and angular distribution (IEAD) incident onto the wafer. b) Initial feature geometry (not to scale) consisting of a SiO₂ substrate with an amorphous carbon mask.

Table 2: High-level description of the SiO₂ and AC etch mechanism

<u>Species</u>			
M	Ion or hot neutral	AC	Amorphous carbon mask
S	Surface Site	(g)	Gas phase
P	Polymer	(s)	Solid
PC	Crosslinked polymer		
Reaction		Description	
SiO _{2(s)}	+ M _(g) → SiO _{2(g)} + M _(g)	Physical sputtering of SiO ₂	
S _(s)	+ SiO _{2(g)} → S _(s) + SiO _{2(s)}	SiO ₂ redeposition	
SiO _{2(s)}	+ C _x F _{y(g)} → SiO ₂ C _x F _{y(s)}	SiO ₂ -Fluorocarbon complex formation	
SiO ₂ C _x F _{y(s)}	+ M _(g) → SiO ₂ C _x F _{y(g)} + M _(g)	Complex sputtering	
S _(s)	+ C _x F _{y(g)} → S _(s) + P _(s)	Polymer deposition on Surfaces	
P _(s)	+ C _x F _{y(g)} → P _(s) + P _(s)	Polymer deposition on Polymer	
P _(s)	+ O _(g) → COF _{y(g)}	Polymer etch by O	
AC _(s)	+ C _x F _{y(g)} → AC _(s) + P _(s)	Polymer deposition on AC	
P _(s)	+ M _(g) → P _(g) + M _(g)	Polymer sputtering	
AC _(s)	+ M _(g) → AC _(g) + M _(g)	AC sputtering	
P _(s)	+ P _(s) → PC _(s) + PC _(s)	Crosslinking	
PC _(s)	+ PC _(s) + M _(g) → P _(s) + P _(s) + M _(g)	Breaking of Crosslinking	

AC film patterned to ideally yield a straight walled opening with an initial width of 90 nm. The etch was performed for 60 seconds.

The profile simulations were performed using the Monte Carlo Feature Scale Model (MCFPM) [7,15,48]. MCFPM resolves the bulk and surface properties on 3D cubic mesh. Here, the mesh cells – voxels - have equal side lengths of 1 nm. Each cell has a material identity. Gas phase species represented by Monte Carlo pseudo particles are launched with fluxes, and energy and angular distributions obtained from HPEM. In the MCFPM, cell identities are changed, or cells are removed or added for reactions including passivation, etching and deposition. The reactions of thermal (low energy) species with the surface for processes that do not have a threshold energy are described by a reaction probability p . Reactions of ions and hot neutrals (energetic ions

that have been neutralized during collisions with surfaces) have reaction yields having an energy and angular dependence. In the MCFPM the energy dependence of chemical and physical sputtering is implemented using the expression [15,49]

$$p(\varepsilon) = p_0 \frac{(\varepsilon - \varepsilon_{th})^q}{(\varepsilon_{th} - \varepsilon_0)^q} \quad (2)$$

for incident energy ε , reference energy ε_0 , reference probability p_0 and threshold energy ε_{th} . Unless indicated otherwise, the optimization of energy-dependent reaction probabilities refers to the value of p_0 . The SiO₂ reaction mechanism is based on previous work by Huang et al [49] and is described in detail in Ref. [15]. A high-level overview of the mechanism is listed in Table 2.

SiO₂ can be removed through physical sputtering by energetic ions and hot neutrals. The sputtered products can redeposit on other surfaces. Unsaturated fluorocarbons can chemisorb on the SiO₂ to form an oxide-fluorocarbon complex. This complex is in turn easier to sputter based on lower threshold energy and overall higher reaction probability as the site has a lower binding energy. Additional polymer can deposit on top of the complex as well as other surfaces, providing a physical barrier to further etching and providing sidewall passivation. The polymer can be removed by physical sputtering or chemical etching by oxygen radicals. The only O-containing radical having significant fluxes to the surface is ground state atomic oxygen.

The gas phase reactions with the mask and substrate are ultimately contained in a single global reaction mechanism where they interact and share certain reactions, such as polymer deposition. A high-level overview of the AC mask reaction mechanism is also in Table 2. See supplementary material at [URL will be inserted by AIP Publishing] for a listing of the full reaction mechanism. The mask mechanism includes polymer deposition by C_xF_y precursors on AC and on prior deposited polymer. Similar to the etch mechanism for SiO₂, polymer is removed by oxygen radicals and physical sputtering. The AC can be sputtered by ions and hot neutrals as well.

The polymer, as deposited, consists of individual C_xF_y radicals and so is strictly an assembly of monomers. This radical based film can subsequently crosslink to create an actual polymeric material that is more resistant to sputtering. Ion bombardment can then break bonds (chain scission) to produce a lower molecular weight polymer. A mechanism was developed to represent crosslinking between the deposited monomers and breaking of those crosslinks.

A polymer's physical and chemical properties can be dependent on crosslinks with neighboring species. The manner of linkage determines the reactivity of the polymer due to, for example, a lack of available radical sites for additional bonding or more resistance to physical processes due to an increase in total bond strength. These crosslinks can be broken by exposure to the plasma producing energetic particles or radiation. Since these energetic particles are typically delivered anisotropically to the surface, this spatially discriminating activation can result in shaping of the polymer deposition. A schematic of the crosslinking model is in Fig. 5. The creation of crosslinks during polymer deposition is represented in the left column (a-d) and the removal or breaking of crosslinks is represented in the right column (e-g).

Crosslinking occurs during the deposition of eligible materials (Fig. 5a, b). (By material, we refer to incident species, such as CF_3 or C_2F_3 , or their counterparts in the film.) Each material has a maximum number of neighbors to which it can crosslink based on the number of available bonds (3 in the example in Fig. 5).

For example, CF_2 would have a maximum of 2 crosslinks and CF_3 would have a maximum of a single crosslink. When a crosslink eligible deposition material is incident onto the surface, a randomly ordered search is conducted of nearest neighbors for other crosslink eligible materials that have available bonding sites. Based on a prescribed crosslinking probability and choice of a random number (0,1) crosslinks are made (or not made) with the eligible neighbors. The number of crosslinks (and the available bonding sites) are recorded for each cell (Fig. 5c). If crosslinking

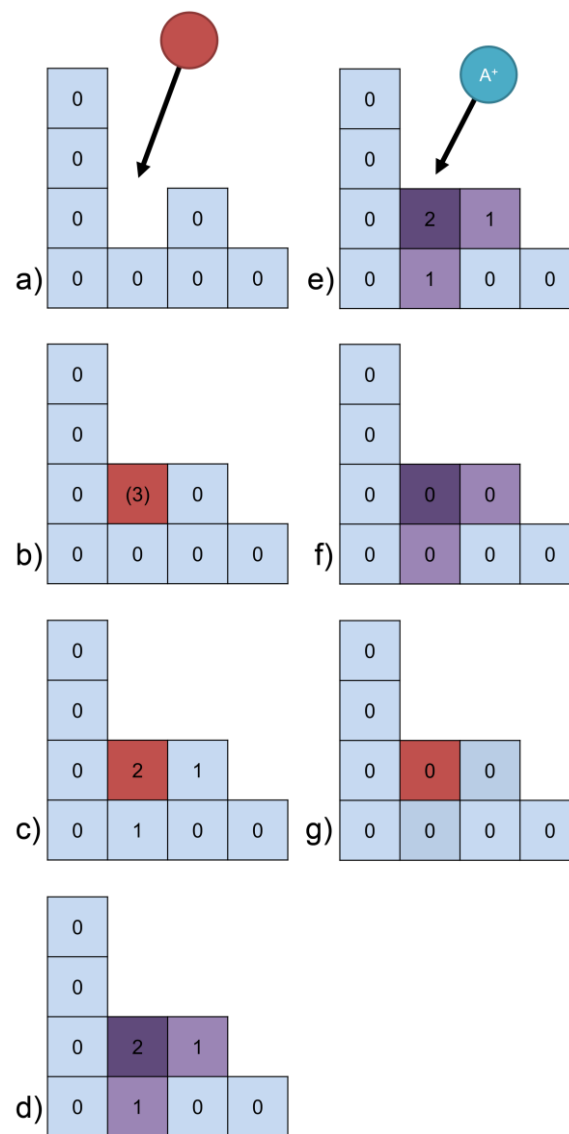


Figure 5: Schematic of the crosslink mechanism shown here in 2 dimensions. a) incident particle, b) deposition, c) crosslink availability, d) change in properties, e) energetic particle bombardment, f) bond breaking, g) change in properties.

occurs the cell identities are changed to represent the change in physical properties of cell (Fig. 5d) that may include, for example, sputtering probability.

Crosslinks can be broken by impinging particles such as ions, hot neutrals and photons (Fig. 5e). The bond breaking is represented by decreasing the crosslink number for both crosslink partners (Fig. 5f) and reversion of the material identify back to an unlinked state (Fig. 5g).

Since the control of necking and clogging by polymer deposition is of special interest we discuss an important control parameter. The amount of polymer growth is determined by the relative contributions of deposition and removal. A steady state polymer thickness occurs when these contributions balance. Polymer removal occurs by sputtering and O-radical based etching. In oxygen rich gas mixtures, polymer removal can be dominated by O-radical based etching, mostly ground state atomic oxygen. In the mechanism discussed here, necking and clogging (the amount of polymer deposition in the mask region) can ultimately be controlled by the reaction probability of the O based polymer etching. The final etch profiles for otherwise identical process conditions are shown in Fig. 6 while varying the probability of polymer etching by O-atoms. An etch probability of 0.005 results in net polymer growth which ultimately leads to a complete clog at the top of the feature. Higher removal probabilities, while still producing significant necking, do not fully clog the feature and allow for continued etching throughout the entire process, as shown in Fig. 6 for a probability of 0.02.

V. Target Metrics and Loss function

A scalar metric is needed to evaluate the quality of the simulation-experiment matching when tuning the reaction mechanism. In this work the loss function $L(\mathbf{p})$ of a given parameter set

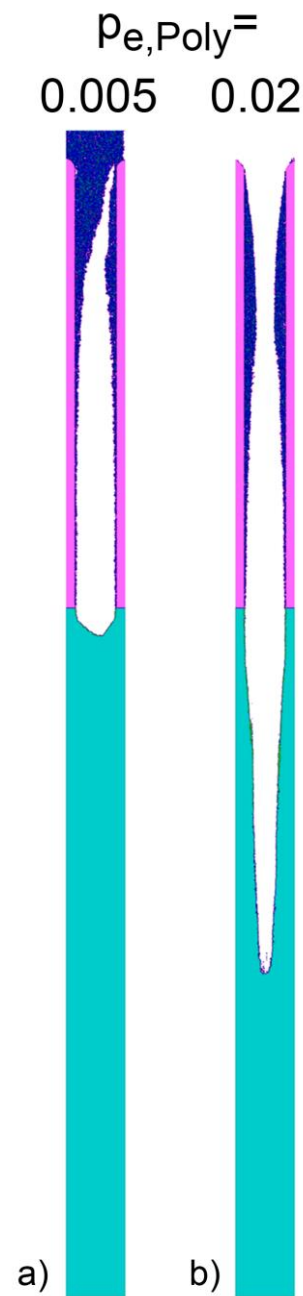


Figure 6: Etch features for different O based polymer etch probabilities of a) 0.005 and b) 0.02.

p is the root-mean-squared (RMS) error between the simulation and experiment of geometric parameters defining the shape of the etched feature, shown in Fig. 7. While these shape parameters are not independent of each other, they nevertheless each aim to represent a specific characteristic of the feature that is of technical relevance. The width of the mask opening w_m including deposition stands as a measure of the necking and clogging, a process which is dependent on the ratio of the fluxes of polymer depositing to removing species. The narrowed opening can impede neutral gas transport into the feature, trap etch products inside the feature and shadow the trajectories of ions and photons. Although the shape (curvature) of the polymer deposition is not a parameter in this investigation, the curvature of the polymer affects the angle of the trajectory with which ions (hot neutrals) reflect from its surface as they progress deeper into the surface. A higher-level approach might include shape (curvature) of the polymer in the optimization process.

The width at the top of the feature, w_t , aims to capture undercutting (etching under the mask) that most often occurs with excessive isotropic etching of the substrate or with there being insufficient passivation of the surface of the substrate compared to that of the mask material. The maximum width of the feature w_f , captures the occurrence of bowing or tapering, which can be caused by lack of passivation or broad ion angular distributions (bowing) or excessive passivation (tapering).

Mask selectivity refers to the relative rate of etching of the mask compared to the substrate. Masks for HAR etching must have high selectivity (low etch rate) compared to the substrate so that the mask has a reasonably small thickness. The thickness of the mask h_m must be controlled to ensure that the mask has a critical thickness at the end of the etch. The thickness of the mask

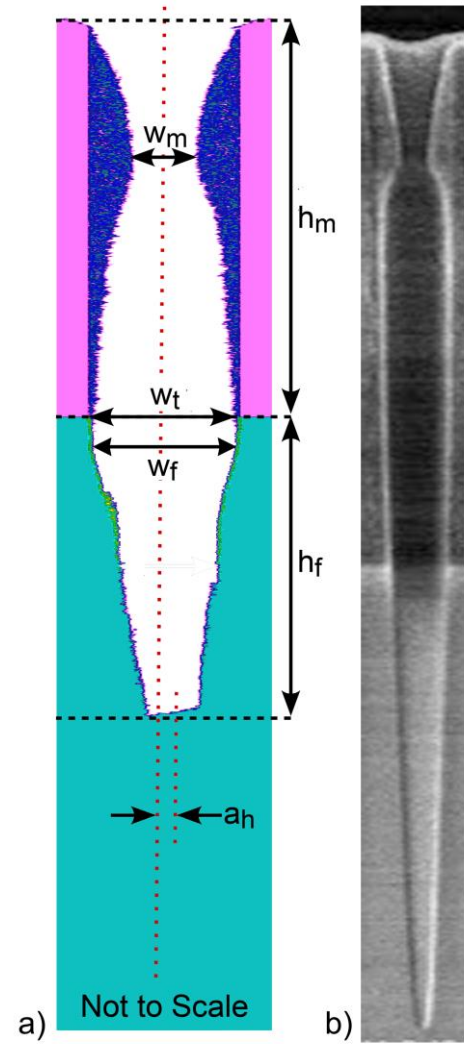


Figure 7: Feature specifications. a) Target metrics used to evaluate the experimental matching. b) SEM image from the base case experiment used for evaluation of the model.

determines the degree to which ions having broad angular spread are shadowed by and reflect off the mask prior to entering the feature. The depth of the final feature, h_f , fundamentally defines the characteristics of HAR features and for a given process time determines the etch rate. Undesired etch defects such as twisting [72] can be captured by measuring the asymmetry, a_h , across the vertical center axis (dotted vertical line in Fig. 7).

The loss function is then

$$L(\mathbf{p}) = \sqrt{a_{wm}(\hat{w}_m - w_m)^2 + a_{wt}(\hat{w}_t - w_t)^2 + a_{wh}(\hat{w}_h - w_h)^2 + a_{whm}(\hat{h}_m - h_m)^2 + a_{wf}(\hat{h}_f - h_f)^2 + a_{wah}(\hat{a}_h - a_h)^2} \quad (3)$$

where the hat accent (^) denotes the target value of a property. Each singular error component can be weighted in accordance to its importance or sensitivity by the weight factors a_i . In this work all components are equally weighted with the exception of the total asymmetry, a_{wah} , which has a higher weighting. The total asymmetry is computed based on the total number of asymmetric computational cells which has a substantially different scale than the other factors.

For any given evaluation of the loss function $L(\mathbf{p})$, the MCFPM is executed to simulate a full etch of 60 s which carries with it a computational cost. The method of evaluating and minimizing $L(\mathbf{p})$ through the optimization algorithm should have the goal of minimizing the number of evaluations. In this work, the target metrics are based on the scanning electron microscopy (SEM) image shown in Fig. 7. This image was produced after the process described in Sec. III was performed in a reactor on which the simulated geometry was based.

VI. Gradient Descent

Gradient descent (GD) is a method for minimizing a multivariate error function, which is convenient due to its conceptual simplicity, universality and fast convergence. The GD method has been applied to a wide variety of physical and engineering problems. [64,65,73–77] In short, GD is based on following the gradient of a loss function L to its minimum as shown in Fig. 8. For a given set of parameters $\mathbf{p}^i$ the selection of the next set of parameters is

$$\mathbf{p}^{i+1} = \mathbf{p}^i - \eta \nabla L(\mathbf{p}^i), \quad (4)$$

where η is the learning rate which controls the effective step size and the rate of descent. The loss function consists of a set of discrete points that results from successive runs of the MCFPM and so is not natively differentiable. A finite difference scheme was used to estimate the partial gradients in every dimension by evaluating $L(\mathbf{p})$ at a test location $L(\mathbf{p}')$,

$$L(\mathbf{p}'_j) = L(\mathbf{p} + \mathbf{x}_j p'_j), \quad (5)$$

with $\mathbf{x}_j$, the displacement unit vector and p'_j , the branch step size for every dimension of the parameter space n and computing the relative derivative $\Delta L / \Delta p_j$. Effectively, the discrete gradient descent formulation becomes

$$\mathbf{p}^{i+1} = \mathbf{p}^i - \eta \sum_{j=1}^n \mathbf{x}_j \frac{\Delta L}{\Delta p_j} = \mathbf{p}^i - \eta \sum_{j=1}^n \mathbf{x}_j \frac{L(\mathbf{p}^i) - L(\mathbf{p}'^i)}{p_j^i - p_j'^i} \quad (6)$$

A useful modification to the gradient descent method is the use of a momentum term that enables the method to overcome local extrema and avoid convergence around a non-global minimum. Analogous to the momentum in Newtonian physics, momentum in GD adds an inertia term to the updated parameters. The previous gradient is remembered and the update on the next iteration is determined by the effective gradient $\mathbf{g}^i$ which is a linear combination of the actual gradient and the previous update:

$$\mathbf{p}^{i+1} = \mathbf{p}^i - \eta \mathbf{g}^i \quad (1)$$

with

$$\mathbf{g}^i = \beta \mathbf{g}^{i-1} + \nabla L(\mathbf{p}^i), \quad (2)$$

where β is the momentum decay coefficient which determines the overall strength of the momentum term. Akin to a heavy ball, rolling down a gravitational potential (referred to as heavy ball method) the gradient descent is less affected by small scale perturbation or “roughness” of the loss function surface.

Based on this methodology, an initial optimization was performed. The tuning parameters as well as the range of their permitted values are listed in Table 3. In the context of the surface reaction mechanism these 5 parameters affect a class of reactions with many individual reactions. For example, p_{s,SiO_2} , the sputter probability of SiO_2 and $p_{s,\text{SiO}_2\text{CFXY}}$, the sputter probability of SiO_2 -polymer complex affect all ion and fast neutral

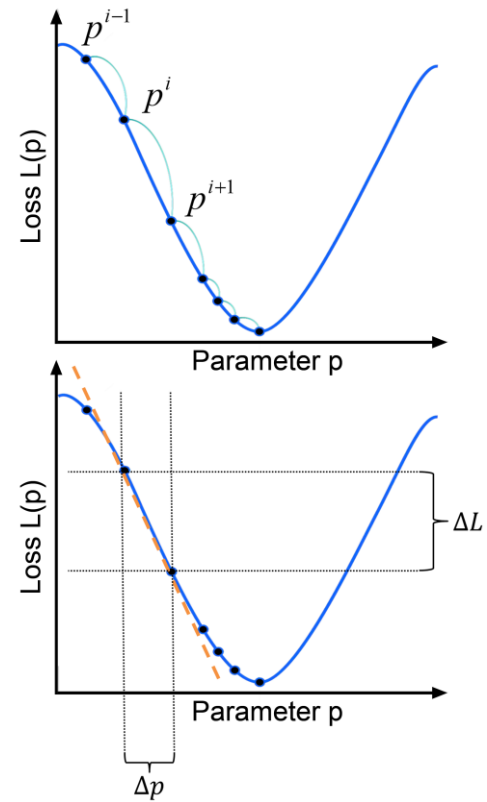


Figure 8: Outline of the gradient descent method for a 1D parameter space.

Table 3: Tuning parameters for optimization

Symbol	Description	Minimum Value	Maximum Value
p_{s, SiO_2}	Sputter probability of SiO_2	0.0	0.3
p_{s, SiO_2CFXY}	Sputter probability of SiO_2 -polymer complex	0.1	0.5
p_{p, SiO_2}	SiO_2 -polymer-complex formation probability	0.1	0.5
$p_{e, poly}$	O based polymer etch	0.0	0.5
$p_{d, poly-AC}$	Polymer deposition probability on mask	0.0	0.5

surface reactions, 44, respectively. The target metrics obtained from experimental data are listed in Table 4. The choice of the range of permitted values was based on prior experience and knowledge of the dominating physics. A judicious choice of the range of permitted values speeds the optimization process while also preventing the optimization process from converging on a non-physical set of parameters. For example, if a non-physical value of polymer etching probability is allowed, the optimization algorithm might compensate for that value with non-physical value of another parameter.

The initial optimization was performed for a total of 100 epochs . (An epoch is single round of evaluating the loss function with a given set of parameters.) The evolution of the total loss function is shown in Fig. 9a as a function of the number of elapsed epochs. The total loss quickly decreased during the first 20 epochs which indicates a partially successful optimization. This convergence is mirrored by the evolution of the target metrics, shown in Fig. 9b. The three metrics, etch depth h_f (blue squares), minimum mask opening w_m (red pyramid) and maximum feature width w_f (green delta) quickly approach their target values, represented by dashed horizontal lines. The remaining metrics were fairly insensitive to changes in the mechanism. In order to reduce clutter, they were omitted from the plot.

Some of the trends in computed metrics are correlated to the underlying adjustments to the parameters defining the

etching mechanism, a selection of which are shown in Fig. 9c. The increase in final predicted etch depth, for example, can be explained by changes to the

Table 4: Target metrics for optimization

Symbol	Description	Target value
w_m	Width of mask opening	45 nm
w_t	Width at the top of the feature	90 nm
w_f	Maximum width of the feature	90 nm
h_f	Etch depth	825 nm
h_m	Remaining mask thickness	850 nm
a_h	Asymmetry	0

SiO₂-polymer complex formation probability, p_{p, SiO_2} , and its sputter probability, $p_{s, \text{SiO}_2 \text{CFXY}}$, both of which increase in tandem with the increase in final etch depth. However, beyond epoch 20, the clear trend towards error minimization breaks down and instabilities both in the model error and in the metrics emerge. In terms of the loss function composition, this breakdown seems to be dominated by large swings in h_f , suggesting its especially high sensitivity to abrupt changes in the mechanism. This breakdown results from the underlying definition of the loss function.

The step size with which the parameter space is traversed is inherently tied to the magnitude of the gradient (Eq. (6)) and should diminish as one approaches the minimum (zero gradient). $L(\mathbf{p})$ is derived from predicted profiles produced by the MCFPM which has a finite grid resolution – the dimension of the voxel. As the step size diminishes, producing a small change in the physics coefficients, the physical change in the predicted profile may not exceed the size of the voxel and there is no change in the loss function. At this point, the loss function is no longer differentiable, as shown in Fig. 10. This condition

can lead to an absence of a gradient when the change in the profile stays within a single voxel, producing a breakdown of the functional relation between $\mathbf{p}$ and $L(\mathbf{p}) = L = \text{constant}$. Another outcome is an overestimation of the gradient when a small change in $\mathbf{p}$ causes the previous static profile to change to a different voxel providing a finitely large response.

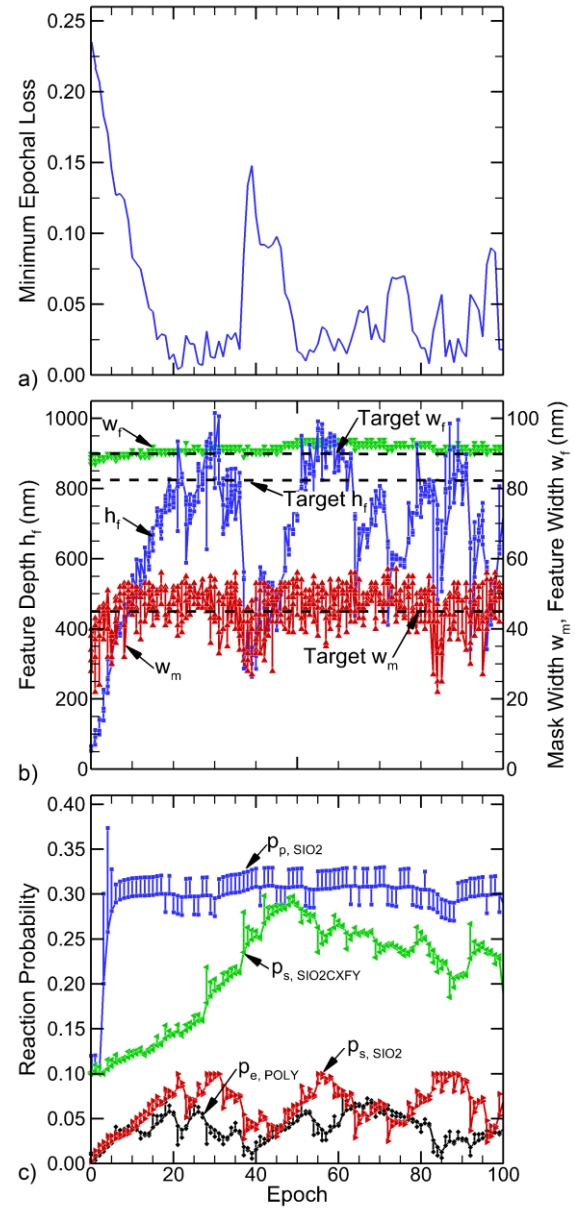
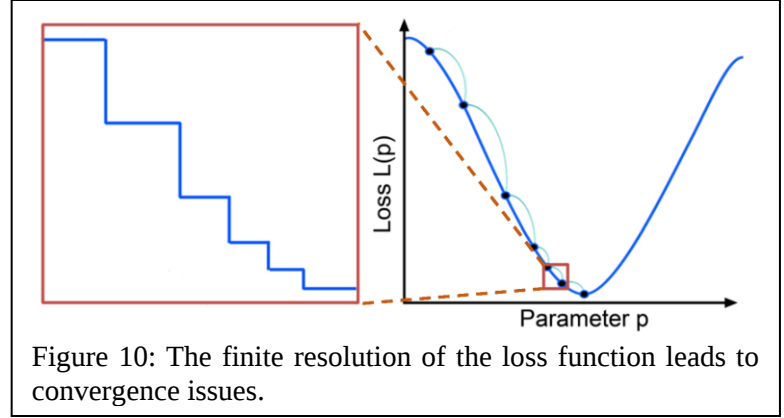


Figure 9: Results from the gradient descent optimization for etch parameters as a function of epoch. a) Minimal epochal loss, b) values of select metrics as a function of epochs and c) select model parameters as a function of epochs.

The GD method provides rapid initial convergence but is potentially erratic when approaching the optimum, in part due to the discrete changes in the loss as the resolution approaches that of the voxel-based mesh. The behavior of the GD and control of the speed of



the initial descent speed is a function of the learning rate η . The evolution of $L(\mathbf{p})$ for different learning rates $\eta = 1 \times 10^{-4}$, 5×10^{-4} , 1×10^{-3} and 5×10^{-3} is shown in Fig. 11. The rate of descent is correlated to η . A smaller learning rate leads to smaller step sizes used to traverse across the parameter space and, consequently, a slower overall descent. Due to the limitations of the method that result from the erratic differentiation near convergence (an outcome of the finite voxel size), all optimizations eventually result in a convergence failure, independent of the learning rate.

In spite of the limitations imposed by the finite difference approach, the GD method has the ability to overcome local minima (to some degree) while having rapid and controllable initial descent. This rapid initial convergence is particularly important in minimizing the number of loss function evaluations and so minimizing computational cost. As a result, one is motivated to retain the GD method to approach the optimized values, followed by a more robust though possibly more slowly converging optimization method. In this work the Nelder-Mead method was selected to perform this latter task.

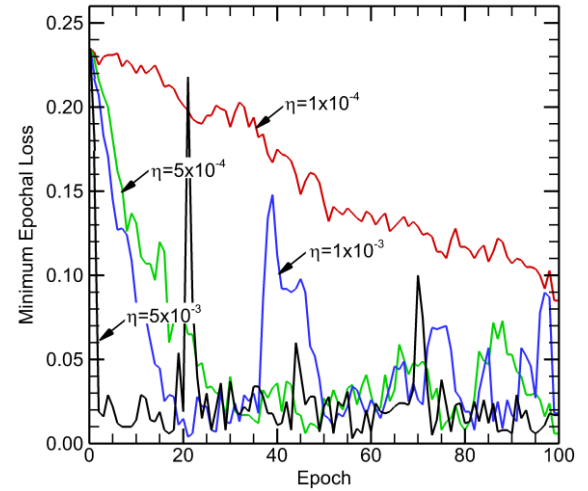


Figure 11: Evolution of the total loss for different learning rates $\eta = 1 \times 10^{-4}$, 5×10^{-4} , 1×10^{-3} and 5×10^{-3} .

VII. Nelder-Mead optimization

The Nelder-Mead method is a gradient free optimization algorithm that has been used in a wide variety of optimization problems [78–84]. Conceptually, the Nelder-Mead (NM) method

represents the tuning parameters as a point in an n -dimensional solution space, where n is the number of tuning parameters to be optimized. A simplex, an object with $n+1$ vertices, each representing a different set of parameters, is employed to traverse the solution space through geometric transformations along its centroid $\mathbf{x}_0$. An example is shown in Fig. 12 with $n = 2$ – a 2D solution space where the simplex is a triangle:

- **Sorting:** The points are sorted according to their loss function from best (lowest loss) to worst (highest loss)
 $L(\mathbf{u}) < L(\mathbf{v}) < L(\mathbf{w})$
- **Reflection:** Reflect the worst point, $\mathbf{w}$, across the centroid; $\mathbf{r} = \mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{w})$, with $\alpha > 0$.
- **Extension:** Extend the reflection beyond point $\mathbf{r}$; $\mathbf{e} = \mathbf{x}_0 + \gamma(\mathbf{r} - \mathbf{x}_0)$, with $\gamma > 1$.
- **Contraction:** Generate two points, $\mathbf{c}_0$ and $\mathbf{c}_1$, that represent contractions of $\mathbf{r}$ towards $\mathbf{w}$; $\mathbf{c}_0 = \mathbf{x}_0 + \rho(\mathbf{r} - \mathbf{x}_0)$ and $\mathbf{c}_1 = \mathbf{x}_0 + \rho(\mathbf{w} - \mathbf{x}_0)$ with $0 < \rho < 0.5$
- **Shrinkage:** Shrink the simplex towards the best point; $\mathbf{w}' = \mathbf{u} + \sigma(\mathbf{w} - \mathbf{u})$ and $\mathbf{v}' = \mathbf{u} + \sigma(\mathbf{v} - \mathbf{u})$, with $0 < \sigma < 1$

Using these geometric transformations, the NM algorithm consists of a specific set of decisions, which leads to an iterative loop that requires at least one successive evaluation of the loss function and a maximum of 4 evaluations per epoch. A flow chart of the Nelder-Mead algorithm is shown in Fig 13. Every epoch begins by sorting the simplex points according to their loss function value, with $L(\mathbf{u}) < L(\mathbf{v}) < L(\mathbf{w})$. Reflection is always the first geometric operation to be tested. If the reflected point, $\mathbf{r}$, is better than the previous second point but worse than the best ($L(\mathbf{u}) < L(\mathbf{r}) < L(\mathbf{v})$), the reflected point replaces the previously worst point ($\mathbf{w} = \mathbf{r}$) and the epoch ends. If $\mathbf{r}$ is better than the previously most preferred point ($L(\mathbf{r}) < L(\mathbf{u})$) the extension point $\mathbf{e}$ is

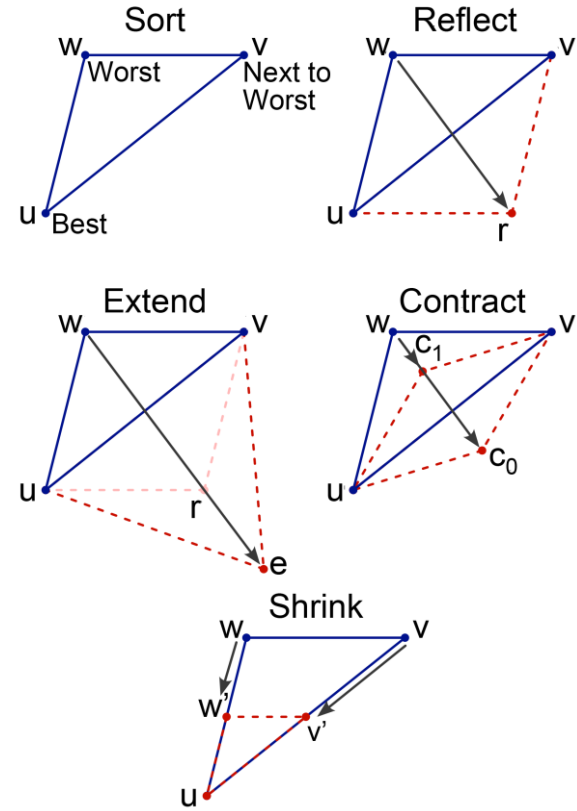


Figure 12: Examples of the geometric transformations of the parameter simplex in a 2D parameter space; reflection, extension; contraction; shrinkage.

tested and added to the simplex if it outperforms $\mathbf{r}$ ($L(\mathbf{e}) < L(\mathbf{r})$). If the reflected point $\mathbf{r}$ is worse than all other simplex points, the contraction is performed. If neither $\mathbf{c}_0$ nor $\mathbf{c}_1$ have a lower loss than $\mathbf{r}$, the simplex is shrunk and the next epoch is initiated. This loop continues until the maximum number of iterations is reached or the specified convergence has been reached. Since the individual parameters represent physical quantities, additional constraints can be placed on the values of the parameters. For example, sputter yields less than zero or reaction probabilities greater than one can be accounted for by limiting respective transformations to valid ranges.

Since NM is a direct search method (only relies on direct evaluations of the loss function itself), NM does not suffer from the same issues that arise with GD which relies on discrete differentiation. NM is regarded as heuristic optimization in this context. Although convergence is technically not formally guaranteed, from a practical perspective NM does nearly always converge. This convergence results from the finite grid size. The loss function effectively becomes insensitive to perturbations below a certain threshold. Once the optimization process switches from GD to NM, no instabilities or non-stationary convergences were encountered.

The use of NM as the only convergence method is somewhat vulnerable to the presence of local minima. That is, NM may converge around a non-global minimum. This propensity for local convergence increases for higher order problems as the size of the parameter space increases.

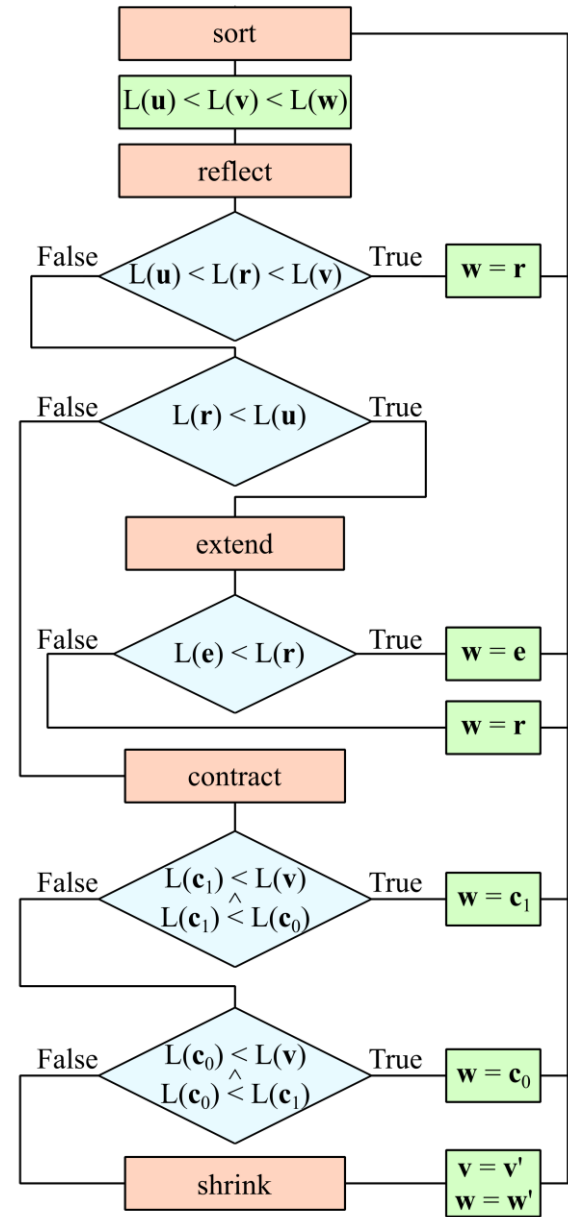


Figure 13: Flow chart describing the Nelder-Mead optimization algorithm

This tendency for local convergence is greatly reduced by initializing the first simplex in the vicinity of the absolute minimum as provided by the initial GD method. That is, the first simplex is not initialized at a random location. As a result, the NM portion of the optimization does not need to perform complete coverage of the parameter space nor conduct an exhaustive search. For our conditions, the sequential GD-NM approach minimized the computational load by reducing the number of loss function evaluations compared to random scatter initialization or swarm optimization methods [85].

VIII. Hybrid GD-NM Optimization

Sequentially executing GD and NM optimization algorithms produces a hybrid scheme whose goal is to address the weakness of each singular approach while capturing their positive qualities. The two optimizers are executed sequentially. The initial gradient descent method is the same as described in Sec. XI and is executed for a prescribed number of epochs. After the final GD epoch, the NM method is initialized. The NM method requires at least $n+1$ points the parameter space to construct the simplex object. The set of initial simplex points contains the best and last point (in terms of loss) obtained from the GD method as well as the best $n-2$ points taken from a randomly distributed scatter across the entire parameter space. In this case 20 random samples were used. This technique provides for the translation of the GD results as well as an adequate spread across the parameter space to prevent an early false convergence of the NM method.

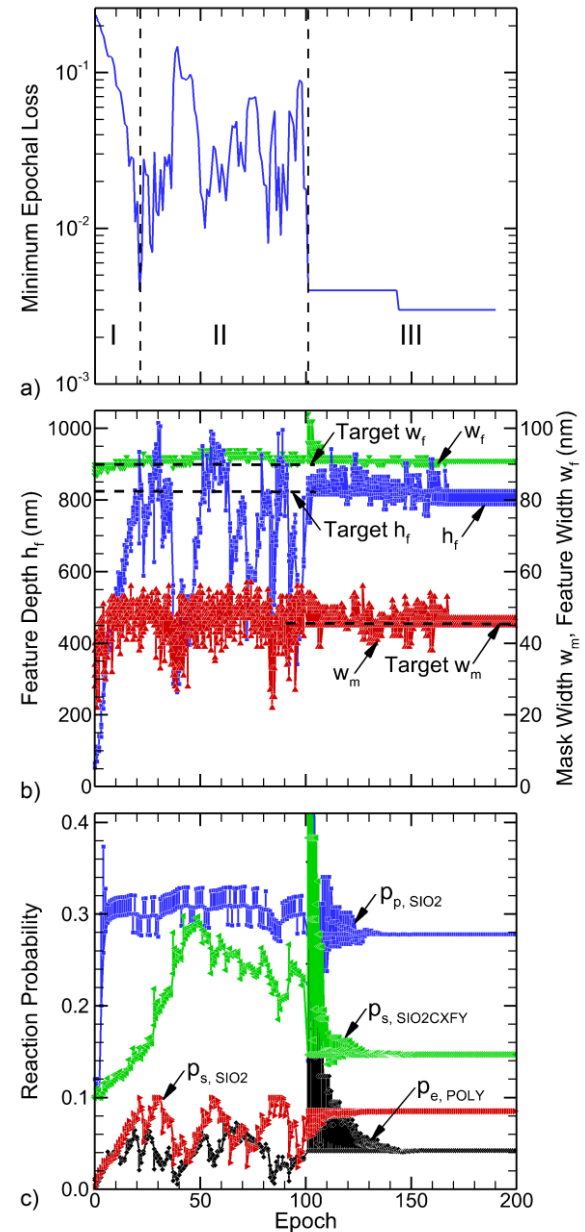


Figure 14: Results from the combined gradient descent and Nelder-Mead optimization. a) Minimum epochal loss, b) values of select metrics as a function of epochs and c) select model parameters as a function of epochs.

The convergence and output parameters produced by the hybrid optimization scheme are shown in Fig. 14 as a function of epoch for the same conditions as in Fig. 9: a) minimum epochal error, b) selected metrics and c) evolution of the tuning parameters. The evolution of the total model error has 3 stages. The first stage consists of

the initial rapid descent and nearly monotonic reduction of the loss function, enabled by the gradient descent with an appropriate learning rate. Starting at approximately epoch 20, Stage II suffers from convergence instabilities which is reflected in fluctuations in the target metrics (Fig. 14b) and model parameters (Fig. 14c). At epoch 100, Stage 3, the gradient descent is terminated and, based on its results, the NM algorithm is initiated. The fluctuations are rapidly dampened and a stable convergence commences. The error drops to approximately 3×10^{-3} which given the grid resolution is on the order of a single cell or less. This minimization in error is mirrored in the target metrics (Fig. 14b) which converge to the desired values. The final parameter set is listed in Table 5.

The definition of model physics parameters based on this optimizer algorithm is based on the reduced scalar representation of the feature. The method will not capture effects that are not included in the target metrics. For example, a subset of the features produced by the MCFPM after the full etch process are shown in Fig. 15 as a function of epochal progression. (These features are predictions at the end of the full etch period for a particular set of model parameters corresponding to an epoch. The sequence of features is not the temporal evolution of the feature itself). The SEM from which the target metrics were extracted is also shown (grayscale). Here, too, the optimization stages can be identified. During epochs 0 – 20, belonging to stage I (initial descent), the monotonic behavior is reflected in the increasing etch depth and widening of the necking. The increased etch rate directly correlates with increasing oxide-polymer complex formation, its removal and increases in the SiO₂ sputter rate. The decreased polymer deposition at the top of the mask is a consequence of the increasing polymer etch by oxygen.

Stage II (instabilities) produces erratic behavior in the predicted profiles with an absence

Table 5: Final tuning parameters after coupled optimization

Tuning Parameter	Value	
	Single Feature	4-Feature Array
p _{s,SiO2}	0.0852	0.0909
p _{s,SiO2CFXY}	0.1471	0.1384
p _{p,SiO2}	0.278	0.2729
p _{e,poly}	0.0423	0.0628
p _{d,poly-AC}	0.094	0.0842

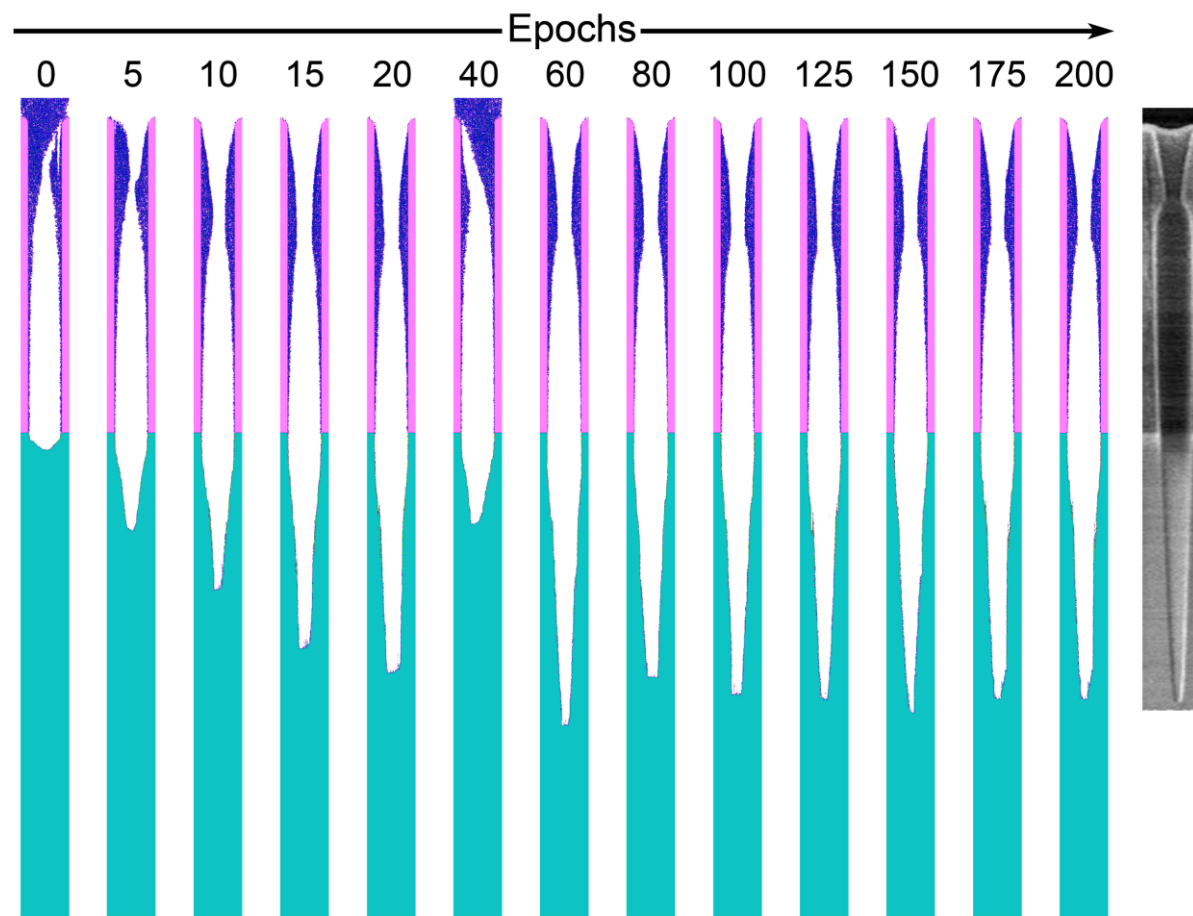


Figure 15: Final features after etch completion for different parameter sets as a function of epochal evolution and the actual experimental target feature.

of clear trends. Between epochs 20 and 100 the etch depth and polymer deposition vary seemingly randomly without clear trends. In Stage III and with the change to the Nelder-Mead algorithm for epochs 100-200, the mechanism converges with a lack of non-monotonic evolution of the metrics. In the absence of a gradient (as in the Nelder-Mead method) the convergence failure is eliminated with there being no changes to the simulation setup, structure of the reaction mechanism and process sensitivities. This outcome supports the proposal that the inadequate gradient determination is the primary failure mode.

Overall the final simulated etch feature at epoch 200 is in good agreement with the experimental counterpart in terms of the metrics used in the optimization process. There are, however, some key differences between the predicted profiles and the experimental SEM. For example, there are differences in the vertical position of the minimum in necking and the taper of the mask. The mismatch in these properties is due to the fact that they were not part of the optimization

metrics. Assuming that the optimization process produced *good model parameters*, these differences indicate that a) the mechanism lacks processes that would otherwise determine the necking location or taper, b) the non-optimized physics parameters are not accurate and/or c) the solution is not unique and a second solution might better capture these phenomena.

The lack of guaranteed uniqueness is a drawback of most optimization schemes, including this one. The physics informed nature of the this approach does, however, enhance physical understanding through the choice of reaction mechanism and the derived parameters. Omitting known minor reactions and setting reasonable ranges of parameters may reduce the applicability of the reaction mechanism while increasing the likelihood of a physically valid and unique outcome. There is inevitably a tradeoff between general applicability and uniqueness of the solution for a more narrowly defined problem. These results stress the importance of including metrics that address the most critical properties of the features in the optimization process, as well as the physics in the reaction mechanism that correlate to those properties. For example, taper of the feature is known to be sensitive to the chemical sputtering probability as a function of angle of incidence of energetic particles. This physics parameter was included in our mechanism but was not part of the optimization process.

In this context, metrics and the optimization processes are as a whole coupled, requiring at least a partial simultaneous evaluation. That is, independently optimizing each metric will not produce the desired result. This coupling can be demonstrated by the relationship between necking and clogging, and feature evolution. Necking or, in the extreme, clogging reduces or entirely impedes the transport of reactive species and ions into the feature, slowing or terminating the etch process. In the event of clogging, the etch process is stopped which prevents evaluation and optimization of the etch process as a whole. Since in this work metrics are only derived from the final feature after a fixed etch time, there is no mechanism to differentiate between two fully clogged features based on the mask metrics alone. The width of mask opening is $w_m = 0$ in both cases even if the rates of the deposition were different and the clogging occurred at different times during the etch. However, differences in rates of, for example, deposition can affect the time that the feature finally clogs. The longer the time that the feature is not clogged, the deeper the total etch depth. Through this mechanism, the etch depth acts as a secondary metric for the clogging mechanism, without which prior attempts to optimize the isolated mask mechanism failed.

Overall, the hybrid optimization scheme performed satisfactory in that it produced a

mechanism that reproduces the desired feature with high accuracy. However, aspects of the process have potential for improvement, particularly with respect to the required computer time. For example, A single optimization requires approximately 20 CPU hours Intel Xeon Gold 6254 processor. This does not include the initial HPEM run required as input. The epoch at which switching between GD and NM methods should be carefully chosen. Epochs spent in Stage II produce little additional information and are not a good use of computing resources. Such hybrid optimization schemes should include a mechanism or criterion to identify convergence failure of the GD and automatically switch to the secondary algorithm. A careful choice of the needed convergence would also minimize the number of epochs spent in Stage III.

IX. Transferability of the converged mechanism

A test of whether the derived model parameters are physically relevant is to evaluate their transferability to *related processes*. Given the small sample set, overfitting or circumstantial unphysical results are a major concern. We use the term *related processes* deliberately. The parameters that have been derived from the optimization process are physics parameters that should, in principle, apply to a wide range of process conditions. However, there is a practical limit. For example, introduction of new gases or new method of excitation (inductively coupled vs capacitively coupled) would likely exceed that limit or require additional parameters in the optimization process.

i. Variation of O₂ in flow

To test the applicability of the derived

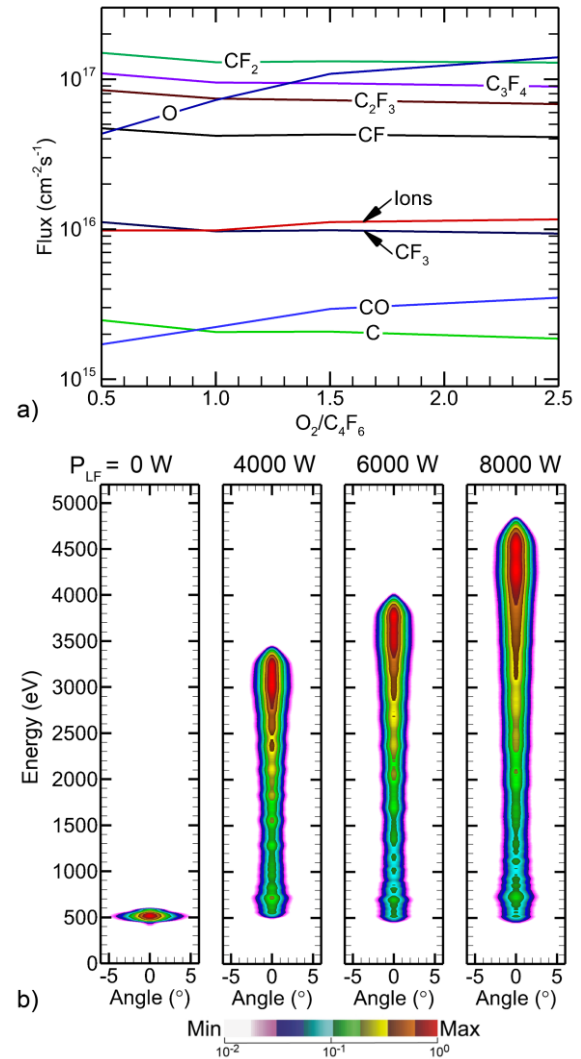


Figure 16: Parametric data for application of derived parameters to other etch conditions. a) Fluxes incident on the wafer surface as a function of the O₂/C₄F₆ feedstock gas ratio. b) IEADs onto the wafer resulting for the low frequency power P_{LF} = 0, 4, 6 and 8 kW.

model parameters, a process with a different input power ($P_{LF} = 6.0$ kW and $P_{HF} = 2.5$ kW) was simulated with differing oxygen to fluorocarbon feedstock gas ratios ($O_2/C_4F_6 = 0.5, 1, 1.5$ and 2.5). All other conditions were kept the same as the base case. The most significant changes compared to the base case were in the particle fluxes incident on the wafer surface, which are shown as function of the feed gas ratio in Fig. 16a. The variation in the O_2/C_4F_6 ratio was achieved, experimentally and in the simulation, by adjusting the rate of oxygen inflow only. As a consequence, the fluxes of the major contributing fluorocarbons remain nearly constant. The O-atom flux increases from 4.1×10^{16} to $1.5 \times 10^{17} \text{ cm}^{-2}\text{s}^{-1}$, a factor of 3.6, with an increase of the oxygen inflow by factor of 5. These trends indicate that mole fractions of feedstock gases do not translate one-to-one to the surface fluxes.

The features generated by MCFPM using the derived reaction parameters and their experimental counterparts are shown in Fig. 17. Although the reaction mechanism used in MCFPM was not specifically tuned for these process conditions, relevant trends were qualitatively reproduced. For example, trends in polymer deposition on the mask are reproduced. At low O_2 fraction and low O-atom flux, the O based etching of deposited polymer is low which leads to full clogging of the feature in both the experiment and simulation. With increasing O_2 the rate at which polymer is removed increases and the necking is reduced, being essentially absent for the $O_2/C_4F_6 = 2.5$ case.

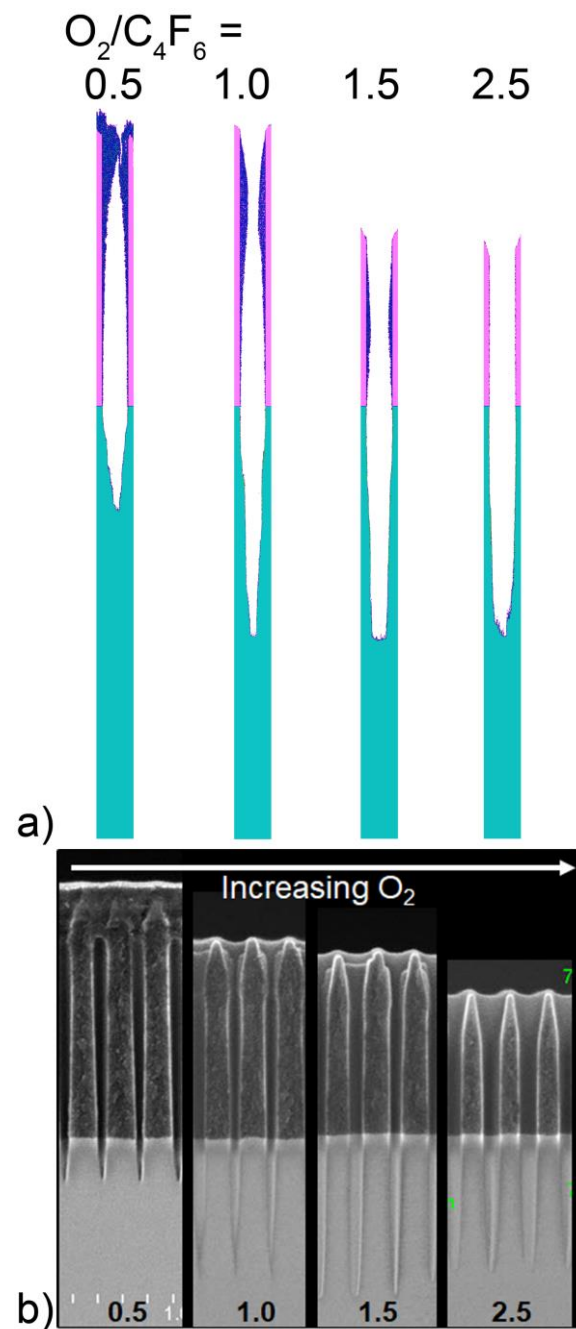


Figure 17: Variation of oxygen in O_2/C_4F_6 gas mixtures: a) Predicted features. b) SEM images of the features resulting from the corresponding experiments.

The lack of polymer film on the mask leads to increased mask erosion with increasing O_2 inflow. The amount of mask erosion in the simulated features is not in quantitative agreement with the experimental results. This lack of quantitative agreement is likely due to not having the necessary physical processes in the mechanism despite not being specifically required for the initial optimization process. For example, direct oxidation of the AC mask by O_2 was not included in the mechanism, but may become important at larger O_2 flow rates. The etch depth is related to the mask necking through reduction in particle transport into the feature and through onset of an etch stop with full clogging. The full clogging of the feature for $O_2/C_4F_6 = 0.5$ is reproduced by the simulation, however etch depth for the fully clogged feature is not reproduced as this depth depends on when the feature was clogged.

The experiments show maximum etch depth for $O_2/C_4F_6 = 1.5$, which is not the case for the simulation. The fact that the etch rate does not increase from $O_2/C_4F_6 = 1.5$ to 2.5 cannot be explained by differences in physical transport alone. In the reaction mechanism, the C_xF_y radicals can deposit on the mask and in the feature, and subsequently be removed by O atom etching. At this point, the C_xF_y becomes unavailable at the etch front. The relative increase of O atom flux may produce a certain level of *fluorocarbon-starvation* that limits the etch rate.

Overall, however, the predicted trends generally agree with experiments. This outcome is not necessarily the expectation given the fact that the mechanism was tuned for only a single different process. The emergence of these untrained trends is a consequence of using a *physics informed* optimization approach where the derived parameters are physics parameters and not simply fitting parameters. The ability to reproduce the experimental trends, to some degree, speaks to this approach's ability to avoid overfitting issues. While not exhaustively demonstrated, the effective parameter space reduction introduced by the *physics informed* approach seems to have enabled the extrapolation of the model beyond its training regime. These trends also rely on the fluxes produced by the HPEM being *ground truth* and accurately representing experimental fluxes, which adds additional variability to the process.

ii. Variation of low frequency power

As second test of the model parameters derived from the optimization process was prediction of feature profiles while varying the low frequency power, $P_{LF} = 0, 4, 6$ and 8 kW. The low frequency power is often used to control the ion energy at the surface of the wafer. IEADs onto the wafer are a first-order process variable in HAR etching. The IEADs produced by HPEM for

this range of P_{LF} are shown in Fig. 16b) with there being a direct correlation between P_{LF} and the mean and maximum energies of ions at the wafer. While the IEAD for $P_{LF} = 0$ kW has the lowest energies, the distribution does contain ions exceeding 500 eV. The high frequency RF power alone produces oscillation in the plasma potential and high energy ions. The maximum angular spread has little variation as a function of P_{LF} , with the exception of $P_{LF} = 0$ kW which has significant broadening of the angular distribution.

Using the derived model parameters, MCFPM simulations were performed for the different values of $P_{LF} = 0$. Those results and the experimental SEM references are shown in Fig. 18. The profiles, experiment and simulation, for $P_{LF} = 0$ kW produce total clogging of the mask opening, indicating that ion energy through sputtering plays an important role in removing excess polymer. The remaining cases ($P_{LF} = 4, 6$ and 8 kW) have unclogged features and full etching with unexpected little variation of the etch depth as a function of low frequency.

In the experimental data a doubling of the P_{LF} (4 kW to 8 kW) produces few differences in the final features. These trends indicate that above a certain threshold energy the etch progression and the mask removal process are not ion starved, but rather limited by neutral gas transport. To some degree this trend is reproduced by the simulations where etch depth does increase with increasing low frequency power however the rate of increase is substantially sublinear with , 635 nm, 715 nm, 720 nm of final etch depth respectively. These outcomes indicate that the effect of ion energy (for example in sputter yield or related processes) might be overestimated in the mechanism.

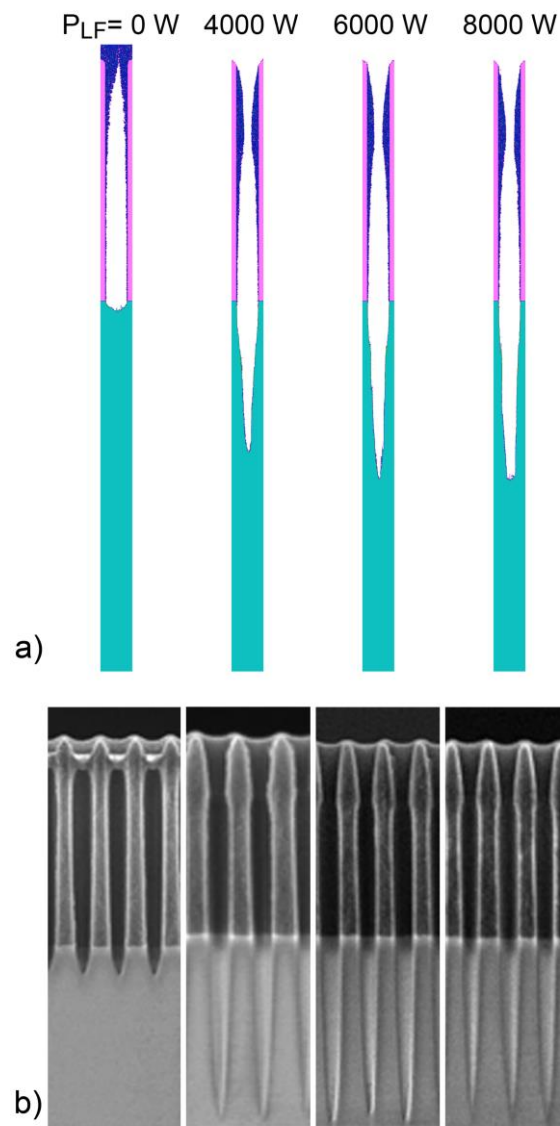


Figure 18: Variation of low frequency power. a) Predicted features. b) SEM images of the features resulting from the corresponding experiments.

Again, this assessment assumes that the results produced by the HPEM are *ground truth*.

X. Optimization based on multiple features

Monte Carlo based simulations are inherently subject to statistical variations. Since the loss function used for optimization is based on a single feature's properties, overfitting of the target metrics may occur. To partially mitigate and to roughly estimate the degree of overfitting, the optimization was performed on a linear array of four features. The individual features in the 4-feature array are identical to those used in optimizations to this point. The width of the features is 90 nm, with a center-to-center pitch spacing of 130 nm, and distance between the edges of features 40 nm. Using the 4-feature array, the total loss function to be minimized is the average of the loss functions of the individual features. The pseudo particles directed towards the feature array from the top of the simulation domain in the MCFPM are launched based on a shared random number pool. The pseudo particles are launched randomly onto the surface encompassing all 4 features, a process which introduces statistical variation between the features. The number of pseudo particles used in the simulation was increased so that during a given time interval, the particle flux (particles/area) was the same as for a single feature.

Apart from the change in feature geometry, the optimization process is unchanged. Analogous to the results discussed in Sec. VIII for a single feature, the epochal evolution of the loss function for the 4-feature array is shown in Fig. 19a. As with the single feature optimization, 3

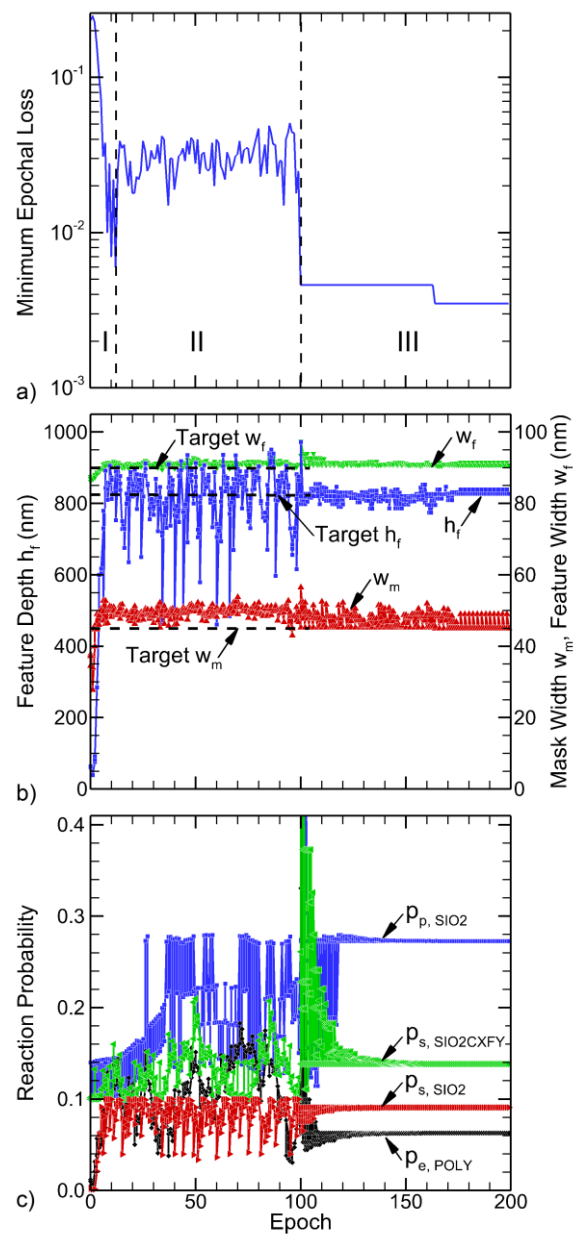


Figure 19: Results of the feature array optimization. a) Minimal epochal loss, b) values of select metrics as a function of epochs and c) select model parameters as a function of epochs.

distinct stages in the optimization can be identified. The initial descent (Stage I) commences more rapidly than with the single feature in spite of what should be the same learning rates. This outcome may be the result of the induced statistical variation due to the launching of particles which is shared among the features. That variation, to a degree, remedies some of the initial overfitting to statistical noise and enables a smoother gradient descent. While the convergence failure (Stage II) still occurs, the random oscillation of the loss is significantly dampened compared to that for a single feature (Fig. 14a). This outcome is likely due to the propensity for incorrect gradient calculations for any single feature being partially negated by acceptable gradient calculations in other features. The errant gradient calculations are unlikely to occur in all features at the same time. The final convergence (Stage III) is similar to that of the single feature optimization in that the loss rapidly converges to a stable value. Although the final loss is slightly larger than that for a single feature, it is still determined by feature variations on the order of a single numerical cell.

These trends are confirmed by the convergence of the metrics shown in Fig. 19b, which are the average of the 4 features. All metrics converge to their respective targets. Compared to the single feature, the initial approach of the metrics to their targets is faster and the intermediate non-convergence oscillations are less erratic and of smaller magnitude. The corresponding tuning parameters are shown in Fig. 19c. While their epochal evolution differs from that of the single feature, their final values are similar. The final values of the turning parameters for the 4-feature array are listed in Table 5.

The feature arrays produced as a function of epoch are shown in Fig. 20. As with the single feature optimization, the initial mechanism produces clogged features, caused by insufficient polymer etching and which leads to an early etch stop. With progressing epochs (left to right in Fig. 20), the net polymer deposition is reduced and the etch rate and etch depth increase. At epoch 100, immediately before stage III, the scalar metrics (Fig. 19b) and 2D feature representation both indicate a good match in etch depth but excessive feature bowing which is largely removed with subsequent ongoing optimization. As in experimental etching of arrays of features, there is some statistical variation from feature to feature. For example, the SEMS in Figs. 17 and 18 show feature to feature variation due to the statistical nature of the reactant fluxes entering each feature. This feature to feature variation occurs in the simulations for the same reason.

While these findings indicate that the optimization based on feature arrays is generally preferable to that using only a single feature, these improvements must be weighed against the

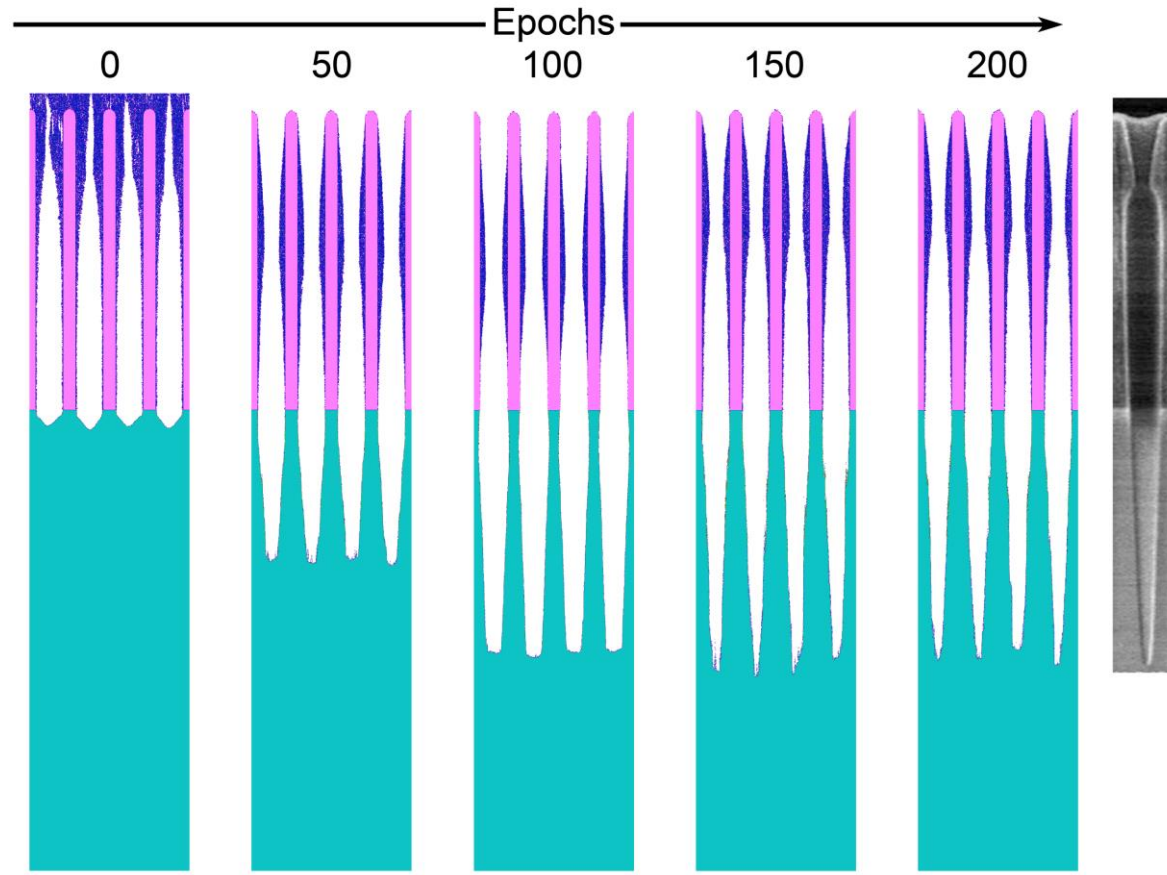


Figure 20: Final feature grids after etch completion for different parameter sets as a function of epochal evolution and the actual experimental target feature.

increase in computational cost that is incurred by the larger geometry and subsequent increase in total number of cells and in the number of pseudoparticles that must be launched. Overall, the required computation time increases linearly with the number of features. Approximately 80 CPU hours were required for the optimization process using 4-features compared to 20 CPU hours for the single feature.

XI. Concluding Remarks

A two-stage hybrid optimization scheme was discussed aimed at optimizing physical parameters in a reaction mechanism that is used in profile simulations of high aspect ratio plasma etching into a SiO_2 substrate using a $\text{C}_4\text{F}_6/\text{Ar}/\text{O}_2$ gas mixture. The process was performed in a dual frequency capacitively coupled plasma reactor with independently controlled high frequency and low frequency powers applied to the bottom electrode. The optimization scheme consisted of consecutive execution of finite difference gradient descent (GD) and Nelder-Mead (NM) algorithms. Doing so enabled taking advantage of some of the desirable qualities of each method while

at least partially negating some of their respective drawbacks. The fast, initial descent of the GD method was maintained while its instability was dampened by the subsequent NM process. The reaction mechanism was optimized to best represent the geometric properties of a final etched feature extracted from scanning electron microscopy images. Using appropriate target metrics, the optimization produced a mechanism that was able to reproduce the experimental reference with sub-1% error based on the metrics describing the feature. Performing the optimization process using multiple features improved the stability of the method.

The transferability of the mechanism was tested by applying the mechanism to etch processes outside the training parameter space using different gas compositions and low frequency power. The mechanism captured the relevant trends. In a variation of the O_2/C_4F_6 ratio, the mechanism reproduced the onset of feature clogging and, to a lesser degree, trends in necking and mask erosion. Similarly, a variation of the low frequency power produced a comparable onset of clogging at low power which was also captured by the mechanism. Trends in feature etch rate (or lack thereof) are reproduced to a lesser degree.

Overall, the optimization process described here to refine physics parameters in a plasma etch mechanism shows promise in that with a limited set of data transferable physics parameters were produced. Using a single SEM image, the process produced a mechanism that was able to reproduce a good match to the training target and was able to reproduce relevant trends outside of its immediate training regime. This degree of success is a consequence of the *physics informed* optimization approach. A base mechanism that already contained explicit formulations of the relevant physical processes was refined by the optimization process. As a consequence the outcome is more universally applicable than the singular case to which it was matched.

Improvements to this proof of concept study includes a more effective criterion for the algorithm switching between GD and NM, and a robust formulation to detect convergence and initiate an early termination of the optimization process to reduce unnecessary compute overhead. Further parallelization may reduce the time to completion while possibly increasing the total required computational cost. This could be implemented by parallelizing execution of individual MCFPM instances or parallel execution of an ensemble of single features. Another potential improvement is the expansion of the metrics to include not only a single process condition but rather match to one or more parametric sweeps, improving the overall statistical quality and physical validity. Additional geometrical feature metrics could be added to better define the features, or

replace the finite feature metrics with image-to-image comparisons. The latter could leverage image-to-image translation techniques that have developed for image processing. Alternatively, the inclusion of multiple features per process condition could prove useful to include statistical variance and prevent overfitting, which is of great concern whenever only a small set of data is used.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of Interest

The authors have no conflicts of interest.

Acknowledgements

This work was supported by Tokyo Electron Ltd. and Samsung Electronics Co. These results were also based upon work supported by the U.S. Department of Energy, Office of Science, Office of Fusion Energy Sciences under award number DE-SC0020232.

References

- [1] V. M. Donnelly and A. Kornblit *J. Vac. Sci. Technol. A Vacuum, Surfaces, Film.* **31**, 050825 (2013)
- [2] I. Adamovich, S. D. Baalrud, A. Bogaerts, P. J. Bruggeman, et al *J. Phys. D. Appl. Phys.* **50**, 323001 (2017)
- [3] G. Packard, A. Rosenfeld, G. S. Oehrlein and S. Hamaguchi *Plasma Sources Sci. Technol.* **27**, 023001 (2018)
- [4] Y. Suzuki, S. Iwashita, T. Sato, H. Yonemichi, H. Moki and T. Moriya, "Machine Learning Approaches for Process Optimization," *2018 International Symposium on Semiconductor Manufacturing (ISSM)*, Tokyo, Japan, 2018, pp. 1-4, doi: 10.1109/ISSM.2018.8651142.
- [5] R. Anirudh, R. Archibald, M. S. Asif, M. M. Becker, et al *IEEE Trans. Plasma Sci.* **51**, 1750 (2023)
- [6] Y. G. Yook, H. S. You, J. H. Park, W. S. Chang, et al *J. Phys. D. Appl. Phys.* **55**, 255202 (2022)
- [7] D. Zhang, S. Rauf and T. Sparks *IEEE Trans. Plasma Sci.* **30**, 114–5 (2002)
- [8] D. Zhang, S. Rauf, T. G. Sparks and P. L. G. Ventzek *J. Vac. Sci. Technol. B* **21**, 828–36 (2003)
- [9] T. Iwase, Y. Kamaji, S. Y. Kang, K. Koga, et al *Jpn. J. Appl. Phys.* **58** SE0802 (2019)
- [10] N. Kuboi *Journal of Micro/Nanopatterning, Materials, and Metrology*, **22**, 041502 (2023)
- [11] J. A. Sethian and D. Adalsteinsson *IEEE Trans. Semicond. Manuf.* **10**, 167 (1997)
- [12] M. Chopra, S. Helpert, R. Verma, Z. Zhang, X. Zhu, and R. Bonnacaze, "A model-based, Bayesian approach to the CF₄/Ar etch of SiO₂," *Proc. SPIE 10588, Design-Process-Technology Co-optimization for Manufacturability XII*, 105880G (20 March 2018); <https://doi.org/10.1117/12.2297482>
- [13] T. Shimada, T. Yagisawa and T. Makabe *Japanese J. Appl. Physics, Part 2 Lett.* **45**, L132 (2006)
- [14] H. Tsuda, Y. Takao, K. Eriguchi and K. Ono *Jpn. J. Appl. Phys.* **51**, 08HC01 (2012)
- [15] S. Huang, C. Huard, S. Shim, S. K. Nam, et al *J. Vac. Sci. Technol. A* **37**, 031304 (2019)
- [16] C. M. Huard, Y. Zhang, S. Sriraman, A. Paterson, et al *J. Vac. Sci. Technol. A* **35**, 031306 (2017)
- [17] H. Tsuda, K. Eriguchi, K. Ono and H. Ohta *Appl. Phys. Express* **2**, 116501 (2009)
- [18] N. A. Mauchamp and S. Hamaguchi *J. Vac. Sci. Technol. A* **40**, 53004 (2022)
- [19] C. M. Huard, S. Sriraman, A. Paterson and M. J. Kushner *J. Vac. Sci. Technol. A* **36**, 06B101 (2018)
- [20] C. Qu, Y. Sakiyama, P. Agarwal and M. J. Kushner *J. Vac. Sci. Technol. A* **39**, 52403 (2021)
- [21] R. Dussart, R. Ettouri, J. Nos, G. Antoun, et al *J. Appl. Phys.* **133**, 113306 (2023)
- [22] G. Antoun, T. Tillocher, P. Lefauchaux, J. Faguet, et al *Sci. Reports* **11**, 357 (2021)
- [23] R. Dussart, T. Tillocher, P. Lefauchaux and M. Boufnichel *J. Phys. D. Appl. Phys.* **47**, 123001 (2014)
- [24] A. Mesbah and D. B. Graves *J. Phys. D. Appl. Phys.* **52**, 30LT02 (2019)
- [25] A. D. Bonzanini, K. Shao, D. B. Graves, S. Hamaguchi, et al *Plasma Sources Sci. Technol.* **32**, 024003 (2023)
- [26] J. Moyne and J. Iskandar *Process* **5**, 39 (2017)
- [27] D. Gidon, D. B. Graves and A. Mesbah *Plasma Sources Sci. Technol.* **26**, 085005 (2017)
- [28] T. F. Edgar, S. W. Butler, W. J. Campbell, C. Pfeiffer, et al *Automatica* **36**, 1567 (2000)
- [29] M. Hankinson, T. Vincent, K. B. Irani and P. P. Khargonekar *IEEE Trans. Semicond.*

- Manuf.* **10**, 121 (1997)
- [30] D. Stokes and G. S. May *IEEE Trans. Semicond. Manuf.* **13**, 469 (2000)
- [31] J. H. Xia, Rusli and A. S. Kumta *IEEE Trans. Plasma Sci.* **38**, 142 (2010)
- [32] W. Q. Xiong, Y. Qiao, L. P. Bai, M. Ghahramani, et al *IEEE Trans. Semicond. Manuf.* **34**, 207 (2021)
- [33] J. H. Xia, Rusli and A. Kumta *IEEE Trans. Plasma Sci.* **38**, 1091 (2010)
- [34] C. E. Davis and G. S. May *IEEE Trans. Electron. Packag. Manuf.* **31**, 104 (2008)
- [35] P. Seleson, M. Mustafa, D. Curreli, C. D. Hauck, et al *Comput. Phys. Commun.* **279**, 108436 (2022)
- [36] T. Gergs, B. Borislavov and J. Trieschmann *J. Vac. Sci. Technol. B* **40**, 12802 (2022)
- [37] G. Dong, X. Wei, J. Bao, G. Brochard, et al *Nucl. Fusion* **61**, 126061 (2021)
- [38] S. Morosohk and E. Schuster *Contrib. to Plasma Phys.* **63**, e202200153 (2023)
- [39] C. Ma, B. Zhu, X. Q. Xu and W. Wang *Phys. Plasmas* **27**, 42502 (2020)
- [40] A. Merlo, D. Böckenhoff, J. Schilling, U. Höfel, et al *Nucl. Fusion* **61**, 096039 (2021)
- [41] J. C. Martínez-Loyola, A. A. Siller-Ceniceros, M. E. Sánchez-Castro, et al *Nucl. Fusion* **62**, 126067 (2022)
- [42] P. Rodriguez-Fernandez, A. E. White, A. J. Creely, M. J. Greenwald, et al *Fusion Sci. Technol.* **74**, 65 (2018)
- [43] S. Dasbach and S. Wiesen *Nucl. Mater. Energy* **34**, 101396 (2023)
- [44] Y. Kim, S. Lee, T. Jung, B. Lee, N. Kwak, S. Park, "Challenges in high-aspect ratio contact (HARC) etching for DRAM capacitor formation," Proc. SPIE 9428, Advanced Etch Technology for Nanopatterning IV, 942806 (17 March 2015); <https://doi.org/10.1117/12.2087765>
- [45] Y. Yang and Y. Xu *J. Vac. Sci. Technol. B* **41**, 52602 (2023)
- [46] M. Omura, J. Hashimoto, T. Adachi, Y. Kondo, et al *Jpn. J. Appl. Phys.* **58**, SEEB02 (2019)
- [47] K. J. Owen, B. VanDerElzen, R. L. Peterson and K. Najafi, "High aspect ratio deep silicon etching," 2012 IEEE 25th International Conference on Micro Electro Mechanical Systems (MEMS), Paris, France, 2012, pp. 251-254, doi: 10.1109/MEMSYS.2012.6170138.
- [48] B. Wu, A. Kumar and S. Pamorthy *J. Appl. Phys.* **108**, 51101 (2010)
- [49] S. Huang, S. Shim, S. K. Nam and M. J. Kushner *J. Vac. Sci. Technol. A* **38**, 023001 (2020)
- [50] Z. Hu, H. Shao, J. Li, P. Lai, et al *J. Vac. Sci. Technol. A* **41**, 63113 (2023)
- [51] J. Bobinac, T. Reiter, J. Piso, X. Klemenschits, et al *Micromachines* **14**, 665 (2023)
- [52] R. P. Brinkmann *J. Phys. D: Appl. Phys.* **44**, 042002 (2011)
- [53] W. Park, J. Han, S. Park and S. Y. Moon *Vacuum* **216**, 112466 (2023)
- [54] X. Xiao, X. Ke, B. Su and H.-Y. Zhang *ECS Trans.* **104**, 201 (2021)
- [55] T. Matsushita, T. Matsumoto, H. Mukai, S. Kyoh, K. Hashimoto, "Dry etch challenges for CD shrinkage in memory process," Proc. SPIE 9428, Advanced Etch Technology for Nanopatterning IV, 942807 (17 March 2015); <https://doi.org/10.1117/12.2085628>
- [56] A. A. Kaptanoglu, K. D. Morgan, C. J. Hansen and S. L. Brunton *Phys. Rev. E* **104**, 015206 (2021)
- [57] E. P. Alves and F. Fiuza *Phys. Rev. Research* **4**, 033192 (2022)
- [58] K. Willcox and J. Peraire *AIAA Journal* **40**, 2323 (2002)
- [59] Y. Xing, M. A. Gosálvez, K. Sato, M. Tian, et al *J. Micromechanics Microengineering* **22**, 085020 (2012)
- [60] M. A. Gosálvez, N. Ferrando, Y. Xing, P. Pal, et al *J. Micromechanics Microengineering* **21**, 065017 (2011)

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1116/6.0003554

- [61] Y. Li, Y. Xing, M. A. Gosálvez, P. Pal and Y. Zhou, "Particle swarm optimization of model parameters: Simulation of deep reactive ion etching by the continuous cellular automaton," *2013 Transducers & Eurosensors XXVII: The 17th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS & EUROSENSORS XXVII)*, Barcelona, Spain, 2013, pp. 1087-1090, doi: 10.1109/Transducers.2013.6626960.
- [62] C. -C. Liu, Y. Li, Y. -S. Yang, C. -Y. Chen and M. -H. Chuang, "Automatic Device Model Parameter Extractions via Hybrid Intelligent Methodology," *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Kobe, Japan, 2020, pp. 355-358, doi: 10.23919/SISPAD49475.2020.9241613.
- [63] T. Xiao and D. Ni *Process.* **9**, 151 (2021)
- [64] M. S. Daoud, M. Shehab, H. M. Al-Mimi, L. Abualigah, et al *Arch. Comput. Methods Eng.* **30**, 2431 (2022)
- [65] M. Jalalitabar, I. G. Tsoulos, Y. Tian, Y. Zhang, et al *Math.* **11**, 682 (2023)
- [66] L. Zhou, X. Zhou, C. A. Yi, L. Zhou, et al *Electron.* **12**, 994 (2023)
- [67] S. Takenaga, Y. Ozaki and M. Onishi *Optim. Lett.* **17**, 283 (2023)
- [68] Z. Huang, S. Yan, and M. Rosenbusch, "Optimizing ion transport in a multi-reflection time-of-flight mass spectrograph by a modified Nelder--Mead simplex algorithm" Authorea. August 21, 2023. DOI: 10.22541/au.169265002.28126249/v1
- [69] M. J. Kushner *J. Phys. D. Appl. Phys.* **42**, 194013 (2009)
- [70] T. Piskin, Y. Qian, P. Pribyl, W. Gekelman, et al *J. Appl. Phys.* **133**, 173302 (2023)
- [71] C. Qu, S. J. Lanham, S. C. Shannon, S. K. Nam, et al *J. Appl. Phys.* **127**, 133302 (2020)
- [72] M. Shen, T. Lill, J. Hoang, H. Chi, et al *Jpn. J. Appl. Phys.* **62**, SI0801 (2023)
- [73] H. J. M. Shi, M. Qiming Xuan, F. Oztoprak and J. Nosedal *Optim. Methods Softw.* **38**, 289 (2023)
- [74] E. J. Paul, M. Landreman and T. Antonsen *J. Plasma Phys.* **87** 905870214 (2021)
- [75] D. Shin and S. J. Hong *J. Vac. Sci. Technol. B* **41**, 64002 (2023)
- [76] J. A. Désidéri *Comptes Rendus Math.* **350**, 313 (2012)
- [77] S. Ruder (2016), <https://arxiv.org/abs/1609.04747> .
- [78] R. R. Barton and J. S. Ivey *Management Science* **42**, 954 (1996)
- [79] K. R. Mahmoud *Electromagnetics* **31**, 578 (2011)
- [80] I. Fajfar, Á. Bürmen and J. Puhan *Optim. Lett.* **13**, 1011 (2019)
- [81] C. Audet and C. Tribes *Comput. Optim. Appl.* **71**, 331 (2018)
- [82] M. A. Luersen and R. Le Riche *Comput. Struct.* **82**, 2251 (2004)
- [83] Y. Ozaki, M. Yano and M. Onishi *IPSJ Trans. Comput. Vis. Appl.* **9**, 1 (2017)
- [84] Y. Lee, A. Resiga, S. Yi and C. Wern *J. Manuf. Mater. Process.* **4**, 66 (2020)
- [85] E. Zahara and Y. T. Kao *Expert Syst. Appl.* **36**, 3880 (2009)

List of Tables

Table 1: Base Case Fluxes to Wafer

Table 2: High-level description of the SiO₂ and AC etch mechanism

Table 3: Tuning parameters for optimization

Table 4: Target metrics for optimization

Table 5: Final tuning parameters after coupled optimization

Figures Captions

1. Global schematic of the optimization process.
2. CCP reactor geometry and circuit.
3. Plasma properties for the base case: (a) Electron density [e], (b) electron temperature T_e , (c) negative fluorine ion density $[F^-]$, (d) ionization rate by bulk electrons and (e) ionization rate by secondary electrons.
4. Inputs for feature profile simulation: a) Ion energy and angular distribution (IEAD) incident onto the wafer. b) Initial feature geometry (not to scale) consisting of a SiO_2 substrate with an amorphous carbon mask.
5. Schematic of the crosslink mechanism shown here in 2 dimensions. a) incident particle, b) deposition, c) crosslink availability, d) change in properties, e) energetic particle bombardment, f) bond breaking, g) change in properties.
6. Etch features for different O based polymer etch probabilities of a) 0.005 and b) 0.02.
7. Feature specifications. a) Target metrics used to evaluate the experimental matching. b) SEM image from the base case experiment used for evaluation of the model.
8. Outline of the gradient descent method for a 1D parameter space.
9. Results from the gradient descent optimization for etch parameters as a function of epoch. a) Minimal epochal loss, b) values of select metrics as a function of epochs and c) select model parameters as a function of epochs.
10. The finite resolution of the loss function leads to convergence issues.
11. Evolution of the total loss for different learning rates $\eta = 1 \times 10^{-4}$, 5×10^{-4} , 1×10^{-3} and 5×10^{-3} .
12. Examples of the geometric transformations of the parameter simplex in a 2D parameter space; reflection, extension; contraction; shrinkage.
13. Flow chart describing the Nelder-Mead optimization algorithm
14. Results from the combined gradient descent and Nelder-Mead optimization. a) Minimal epochal loss, b) values of select metrics as a function of epochs and c) select model parameters as a function of epochs.
15. Final features after etch completion for different parameter sets as a function of epochal evolution and the actual experimental target feature.
16. Parametric data for application of derived parameters to other etch conditions. a) Fluxes incident on the wafer surface as a function of the O_2/C_4F_6 feedstock gas ratio. b) IEADs onto the wafer resulting for the low frequency power $P_{LF} = 0, 4, 6$ and 8 kW.
17. Variation of oxygen in O_2/C_4F_6 gas mixtures: a) Predicted features. b) SEM images of the features resulting from the corresponding experiments.
18. Variation of low frequency power. a) Predicted features. b) SEM images of the features resulting from the corresponding experiments.
19. Results of the feature array optimization. a) Minimal epochal loss, b) values of select metrics as a function of epochs and c) select model parameters as a function of epochs.
20. Final feature grids after etch completion for different parameter sets as a function of epochal evolution and the actual experimental target feature.