Etch-Explorer: A Robust Bayesian Optimization Framework for Stringent Constrained Plasma Etching

Yujie Zhang
School of Integrated Circuits
Beijing University of Posts and Telecommunications
Beijing, China
zhangyujie0308@bupt.edu.cn

Xiao Yang
School of Integrated Circuits
Beijing University of Posts and Telecommunications
Beijing, China
yangxiao0618@bupt.edu.cn

Kang Zhao*
School of Integrated Circuits
Beijing University of Posts and Telecommunications
Beijing, China
zhaokang@bupt.edu.cn

Jianwang Zhai
School of Integrated Circuits
Beijing University of Posts and Telecommunications
Beijing, China
zhaijw@bupt.edu.cn

Abstract

Plasma etching is a critical process in semiconductor manufacturing, yet discovering optimal recipes is hindered by the constraint collapse challenge, in which the "golden" intervals that simultaneously satisfy multiple stringent interval constraints are extremely sparse in the high-dimensional design space. Traditional optimization methods often struggle with this sparsity and the complex physical coupling of plasma reactions. To address these limitations, we propose Etch-Explorer, a robust Bayesian Optimization (BO)-based framework that features three synergistic innovations. The framework first employs a Heterogeneous Active Sampling (HAS) strategy to capture space skeletons and physical boundaries, effectively mitigating initial search blindness in sparse regions. Subsequently, a Joint-Constraint-Aware Acquisition Function (JCAF) is leveraged to provide risk-aware navigation by explicitly modeling the joint satisfaction probabilities across multiple interval objectives. To support precise decision-making, a Deep Residual Process Emulator (ResSAN-DTS) is integrated to capture deep non-linear physical couplings with high fidelity. Experimental results demonstrate that Etch-Explorer significantly outperforms state-of-the-art methods in search efficiency and success rate, successfully locating optimal recipes within stringent constraints while substantially reducing wafer costs.

CCS Concepts

• **Theory of computation → Bayesian analysis**.

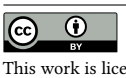
Keywords

Plasma Etching, Bayesian Optimization, Adaptive Sampling, Constrained Acquisition Function, Process Emulator Model

ACM Reference Format:

Yujie Zhang, Xiao Yang, Kang Zhao, and Jianwang Zhai. 2026. Etch-Explorer: A Robust Bayesian Optimization Framework for Stringent Constrained Plasma Etching. In *Great Lakes Symposium on VLSI 2026 (GLSVLSI '26), June 22–24, 2026, Canandaigua, NY, USA*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3787109.3815264

*Corresponding author.



GLSVLSI '26, Canandaigua, NY, USA

ACM ISBN 979-8-4007-2431-2/26/06
https://doi.org/10.1145/3787109.3815264

1 Introduction

The manufacturing of semiconductor chips is the cornerstone of modern information technology, where the relentless scaling of VLSI systems demands chips' nanometer-scale precision. The process involves hundreds of specialized steps, nearly half of which require complex chemical plasma processes such as etching and deposition. In the sub-5nm era, these manufacturing frontiers have become the primary determinants of Integrated Circuit (IC) performance, as even minor process variations can lead to significant degradation in Power, Performance, and Area (PPA) metrics. Traditional development relies heavily on the experience and trial-and-error of process engineers, leading to low efficiency and difficulty meeting modern precision requirements. The prohibitive costs of wafer experiments and the scarcity of high-fidelity experimental data further constrain the effectiveness of conventional approaches [1]. In plasma etching, process parameter tuning faces a unique and formidable challenge termed constraint collapse. Unlike general optimization tasks, etching recipes must balance the synergy between ion bombardment and chemical adsorption. Key performance indicators-such as the etch rate, the critical dimension loading, and the verticality of the sidewall-are often governed by competing physical mechanisms. As the number of process parameters increases, the "golden" window (the feasible region that satisfies all joint constraints) shrinks exponentially, becoming extremely sparse and fragmented within the high-dimensional design space. This structural complexity creates a "needle-in-a-haystack" scenario, where traditional heuristics often fail to locate any valid process window.

With the rapid advancement of technologies such as machine learning and deep learning in recent years, including the emergence of deep convolutional models [2], Genetic Algorithms (GA) [3], and random search algorithms [4], as well as probabilistic graphical models such as Bayesian networks for structure learning [5] and adaptive parameter control [6]. Alongside the increasing number of process parameters that require control, these advanced technologies have provided new momentum to address unique challenges in semiconductor manufacturing. In particular, Bayesian methods have shown promise in maintaining data distributions during sampling [7], which is relevant to our adaptive warm-up strategy that seeks to

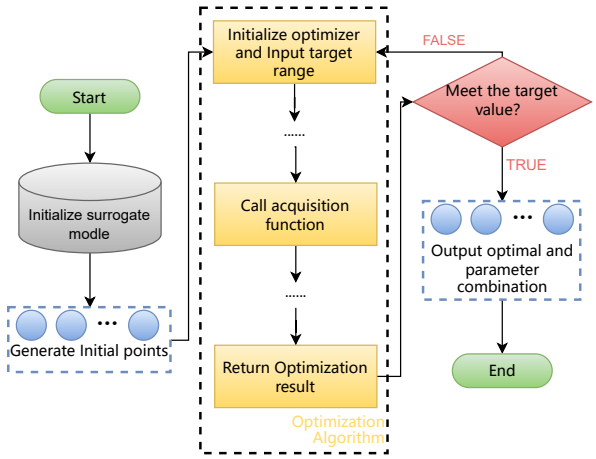


Figure 1: Illustrations of BO-based parameter tuning for plasma etching processes.

preserve parameter space coverage. Thus, integrating Artificial Intelligence (AI) and data-driven technologies offers transformative potential for the field [8]. Several AI techniques have been applied to semiconductor manufacturing, such as Bayesian Optimization (BO), Reinforcement Learning (RL), and GA. For example, Khakifirooz et al. [9] developed a framework based on Bayesian inference and Gibbs sampling to analyze large datasets from the semiconductor manufacturing industry, allowing fault detection and supporting smart production. Wang et al. [10] reviewed the reinforcement learning application framework in manufacturing scheduling and analyzed the integration model between reinforcement learning and metaheuristics. Park et al. [11] proposed a deep RL-based scheduling approach for semiconductor packaging facilities that adapts to fluctuations in machine availability and production demand. Heidary et al. [12] integrated GA with simulation to tackle flexible stochastic job-shop scheduling problems in semiconductor manufacturing.

Traditional Design Space Exploration (DSE) methods, particularly those relying on generic BO with Gaussian Process (GP) modeling, struggle to navigate this "collapsed" space. They often suffer from initial search blindness, where random samples fail to hit any feasible region, and cannot model the sharp nonlinear boundaries of plasma physical-chemical reactions. Recent work by Yang et al. introduced a target-interval BO framework using a Sparse Multi-task GP (SMTGP) and a Probability-guided Interval Search Mechanism (PRISM) [13]. Although effective in anticipated scenarios, SMTGP struggles to capture strong nonlinearities and noise, while PRISM lacks explicit handling of heteroskedasticity and joint constraint satisfaction. Consequently, a more robust, joint-constraint-aware framework is urgently needed to provide precise navigation through these narrow windows while minimizing expensive wafer costs. We identify two critical challenges that hinder efficient process design in plasma etching:

1) Emulator Process Gap: Standard models fail to capture the deep non-linear couplings and stochastic noise inherent in plasma reactions.

2) Navigation Failure: Conventional acquisition functions cannot effectively traverse "collapsed" design spaces where high-risk zones surround narrow feasible regions.

To address these critical bottlenecks, we propose Etch-Explorer, a systematic and automated BO framework specifically tailored for plasma etching optimization (Figure 1). The philosophy behind Etch-Explorer is to bridge the gap between advanced machine learning and the underlying physical constraints of the etching process. By integrating Heterogeneous Active Sampling (HAS), a Joint-Constraint-Aware Acquisition Function (JCAF), and a Deep Residual Emulator (ResSAN-DTS), our framework transforms the tuning procedure from a blind search into informed navigation. Etch-Explorer not only identifies the skeleton of the space and the physical boundaries during the early stages, but also maintains risk-aware decision-making to locate optimal recipes within the narrow "golden" window. Our main contributions are summarized as follows:

- *Heterogeneous Active Sampling (HAS):* We propose a multi-strategy probing method that synergizes global exploration with boundary-sensitive probing. By incorporating expertise-guided and corner sampling, HAS efficiently captures the space skeleton and physical failure boundaries, significantly reducing initial search blindness.
- *Joint-Constraint-Aware Navigation (JCAF):* We introduce a novel JCAF acquisition function that explicitly models the joint satisfaction probability of multiple interval constraints. This enables a risk-aware search logic that prioritizes high-performance regions with maximum feasibility, effectively navigating the narrow "golden" intervals.
- *ResSAN-DTS Emulator Model:* We implement a Deep Residual Network (ResSAN) with a Dynamic Training Strategy (DTS) as a high-fidelity virtual, etc. The SiLU-activated residual architecture captures deep non-linear physical couplings, providing precise uncertainty estimates for constrained optimization.
- *Experimental Validation:* We evaluate Etch-Explorer on real-world etching datasets. The results demonstrate that our framework achieves a state-of-the-art success rate and search efficiency, reducing the number of required physical experiments even under the most stringent process constraints.

The remainder of this paper is structured as follows. Section 2 provides the preliminaries of Bayesian optimization and defines the joint-constraint tuning problem in plasma etching. Section 3 details the proposed Etch-Explorer framework, including the Heterogeneous Active Sampling strategy (HAS), the Joint-Constraint-Aware Acquisition Function (JCAF), and the ResSAN-DTS emulator model. Section 4 evaluates the performance of the framework through comprehensive experiments and ablation studies, demonstrating its superiority in navigating stringent process windows. Finally, Section 5 concludes the paper and discusses future research directions.

2 Preliminaries

2.1 Optimization of Process Parameters in Plasma Etching

Contemporary advancements in semiconductor fabrication are defined by the imperative for increased precision and efficiency. Among various fabrication techniques, Plasma Etching is extensively utilized for transferring precise patterns onto a wafer, playing a crucial role in defining the feature dimensions of integrated circuits. The quality of the etching process, characterized by electrical and physical uniformity, is strongly influenced by a high-dimensional set of process parameters.

The etching workflow operates as a complex input-output system driven by chemically reactive plasma, as shown in Figure 2. As defined in the process configuration, the system state is controlled

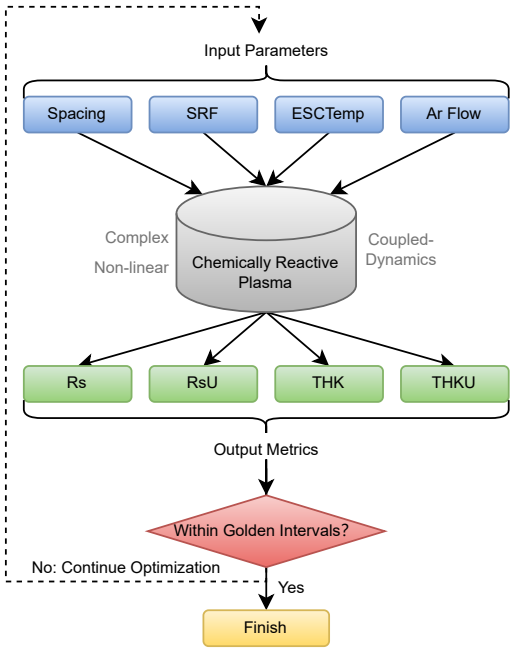


Figure 2: Illustrations of etching workflow.

by specific input parameters: electrode spacing (Spacing), source RF power (SRF), electrostatic chuck temperature (ESCTemp), and argon flow rate (Ar), et al. These inputs determine the dynamics of the plasma, which in turn dictates the output metrics typically evaluated to gauge performance: resistance (Rs), resistance uniformity (RsU), film thickness (THK), and thickness uniformity (THKU), et al. The goal of process optimization is to find a set of input parameters that drives all output metrics to fall simultaneously within their prespecified "golden" intervals, a requirement for high product yield. This is challenging due to the complex, non-linear, and coupled relationship between inputs and outputs being highly. Furthermore, similar to other high-precision fabrication steps, experimental validation is extremely costly and time-consuming, making exhaustive search strategies infeasible and highlighting the need for an efficient, constraint-aware optimization framework.

2.2 Bayesian Optimization (BO) Algorithm

In plasma etching, process parameter optimization is fundamentally a black-box optimization problem with stringent constraints. Due to the extremely high cost of actual etching process experiments, traditional trial-and-error methods struggle to achieve effective optimization within a limited number of trials. BO is an efficient optimization algorithm particularly suited for high-cost black-box problems [14], as shown in Figure 3. BO constructs a probabilistic surrogate model and uses dynamic sampling to transform optimization into sequential decision-making. The core of BO lies in establishing a probabilistic mapping between the parameter space and process objectives using GP, guiding the direction of the parameter search with an expected improvement criterion, and approximating the optimal solution within a finite number of iterations [15].

Three core components of Bayesian optimization:

• *Probabilistic Surrogate Model:* GP are commonly used for probabilistic modeling of surrogate functions [16]. Given the observed dataset $D_{1:t} = \{(x_i, y_i)\}_{i=1}^{t}$, GP characterizes the correlation between parameters via a kernel function $k(x, x')$, establishing the posterior distribution of the objective function:

$$p(f(x)|D_{1:t}) \sim \mathcal{GP}\left(\mu_t(x), \sigma_t^2(x)\right), \tag{1}$$

where $\mu_t(x)$ is the mean function, reflecting the expected performance of parameter x; $\sigma_t^2(x)$ is the variance function, representing the uncertainty of the prediction.

• *Acquisition Function:* A utility function guides the search by balancing the exploration of uncertain regions and the exploitation of promising areas, identifying the next point for evaluation. This includes the acquisition functions: Probability of Improvement (PI), Upper Confidence Bound (UCB), and Expected Improvement (EI).

• *Optimization Execution Strategy:* An iterative loop that updates the surrogate model and maximizes the acquisition function to select new points sequentially until convergence, as illustrated in the workflow diagram in Figure 3.

2.3 Process Emulator Model

The process emulator plays a critical role throughout the flow, establishing a high-fidelity mapping from the process parameters to the target performance metrics. By acting as a "digital twin" of the actual reactor, it defines the optimization landscape and provides objective targets for the search algorithm. This paper employs a Deep Residual to construct the emulator. Unlike a standard Multilayer Perceptron (MLP), which typically includes simple input, hidden, and output layers [17], our architecture captures deeper physical couplings. Mathematically, given an input parameter vector $x \in \mathbb{R}^d$, the output prediction is computed via forward propagation:

$$\hat{y} = W^{(L)} \cdot \sigma\left(W^{(L-1)} \cdots \sigma(W^{(1)}x + b^{(1)})\right) + b^{(L)}, \tag{2}$$

where $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias vector of the l-th layer, respectively, and $\sigma(\cdot)$ is the activation function. The MLP is trained by minimizing the prediction error $L = \|y - \hat{y}\|_2^2$ using the backpropagation algorithm to update network parameters. Tunable hyperparameters in the MLP include the number and nodes of hidden layers, activation functions (Sigmoid, ReLU, Tanh, etc.), learning rate, regularization parameters, batch size, and optimizer. Obtaining the optimal hyperparameter combination through experimental adjustment can enhance the model's fit to the task, thereby improving the overall algorithm performance.

2.4 Problem Formulation

The parameter optimization problem in plasma etching is defined as follows. Let $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$ denote the d dimensional vector of the input process parameters (e.g., Spacing, Power). Let $\mathbf{y} = [y_1, y_2, \ldots, y_K]^T$ represent the K output metrics of interest (e.g., Rs, THK). The core objective is to find an input configuration $\mathbf{x}^*$ such that all outputs simultaneously satisfy their respective strict interval constraints:

$$l_k \le y_k(\mathbf{x}) \le r_k, \quad \forall k \in \{1, \ldots, K\}, \tag{3}$$

where $[l_k, r_k]$ is the desired target interval for the k-th output. Given the noisy and expensive nature, the goal is achieved by maximizing the probability of satisfying all constraints:

$$\max_{\mathbf{x}\in\mathcal{X}} \mathbb{P}\left(\bigcap_{k=1}^{K} \{l_k \le y_k(\mathbf{x}) \le r_k\}\right). \tag{4}$$

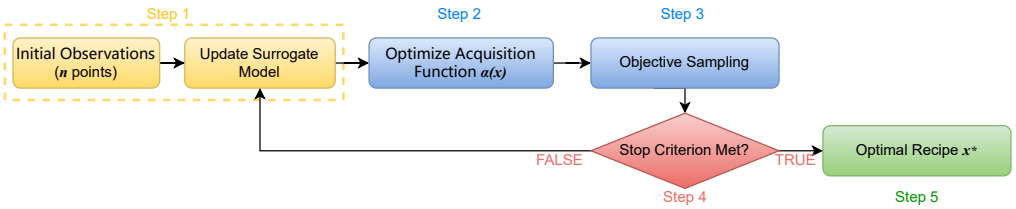


Figure 3: Flowchart of the Bayesian optimization algorithm.

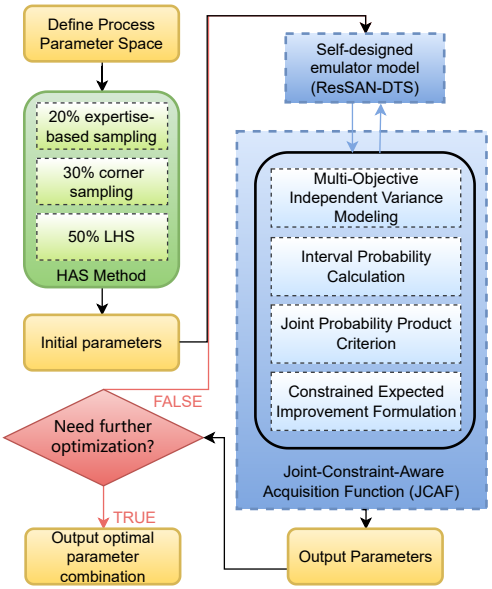


Figure 4: The Etch-Explorer framework for process parameter optimization in plasma etching.

3 Methodology

3.1 Framework Overview

The standard BO paradigm typically relies on GP modeling and generic acquisition functions (e.g., EI, UCB). However, these conventional tools often struggle with the constraint collapse inherent in plasma etching, where the highly nonlinear, noisy process physics and narrow joint-interval constraints lead to frequent search failures in the emulator landscape. To overcome these limitations, we propose Etch-Explorer, a robust framework specifically engineered for plasma process design.

As illustrated in Figure 4, Etch-Explorer transforms the standard BO loop into a joint-constraint-aware navigation engine through three synergetic architectural upgrades:

• *Boundary-Aware Initialization:* Unlike blind random sampling, we implement Heterogeneous Active Sampling (HAS). By integrating expertise-guided anchors and corner-probing, this strategy constructs an initial "space skeleton" that captures the physical boundaries of the process window from the outset.

• *Risk-Aware Decision Engine:* We introduce the Joint-Constraint-Aware Acquisition Function (JCAF). It serves as the framework's navigator, explicitly calculating the joint probability of satisfying all interval constraints to prioritize candidates that offer both performance improvement and process safety.

• *High-Fidelity Process Emulation:* We establish a Deep Residual Network (ResSAN) enhanced by a Dynamic Training Strategy (DTS) to serve as the optimization target. This combination acts as a high-fidelity "virtual etcher," capturing the deep non-linear couplings and heteroskedastic noise that traditional GP-based analytical functions fail to represent.

Algorithm 1 Heterogeneous Active Sampling

Input: n: number of initial samples; *bounds*: parameter bounds
Output: sampled parameter combinations
$S = \emptyset$
1) Expertise-based sampling (20%):
Exploit plasma physics relationships for intelligent sampling
$S = S \cup \text{ExpertiseBasedSampling}(0.2n)$
2) Corner sampling (30%):
Explore boundaries to capture extreme condition effects
$S = S \cup \text{CornerSampling}(0.3n)$
3) LHS (50%):
$n_l = n - |S|$
if $n_l > 0$ **then**
  Ensure uniform coverage with space-filling design
  $S = S \cup \text{LatinHypercubeSampling}(n_l)$
end if
return $S[0:n]$

The overall workflow (Figure 4) operates as a closed-loop system, enabling automatic recipe discovery while significantly reducing expensive wafer costs. The following subsections detail the rationale for the design and technical formulation of each core component.

3.2 HAS: A Heterogeneous Probing Strategy

The performance of BO is fundamentally governed by the quality of the initial sample set, particularly in plasma etching, where the golden intervals are extremely sparse. Traditional Design of Experiments (DOE) strategies, such as Latin Hypercube Sampling (LHS), often waste expensive wafer resources on expansive invalid regions, while pure expertise-based sampling may introduce human bias and miss non-intuitive optima.

To overcome these limitations, we propose Heterogeneous Active Sampling (HAS), a process-aware probing strategy designed to delineate the geometric contours of the process window with minimal experimental trials. Unlike conventional warm-up methods, HAS synergistically integrates three complementary techniques to construct a comprehensive space skeleton:

• *Latin Hypercube Sampling:* We utilize LHS to ensure uniform coverage across the high-dimensional parameter space, preventing sample clustering and providing a baseline understanding of global workflow trends [18].

• *Expertise-based Sampling:* To leverage domain knowledge, we inject expertise-guided samples as physical pivots. By biasing samples toward regions with known high potential (e.g., correlations between power and temperature), Heterogeneous Active Sampling

(HAS) avoids blind search and accelerates convergence toward physically plausible regimes.

• *Corner Sampling:* To address the constraint collapse, we explicitly employ corner sampling to probe the limits of the parameter space. This component is critical for identifying physical failure boundaries-the points where plasma discharge becomes unstable or etching selectivity drops sharply.

As illustrated in Algorithm 1, HAS employs an optimized weighting mechanism (50% LHS, 20% Expertise-based, and 30% Corner Sampling). This heterogeneous composition ensures that the BO engine possesses "Boundary Awareness" from the very first iteration. By simultaneously capturing the global space skeleton, high-potential anchors, and critical process limits, HAS transforms the initial probing from a random trial into a strategic mapping of the etching process window, effectively mitigating the inherent sparsity of the search space.

3.3 JCAF: A Joint Constraint Navigator

The search for an optimal etching recipe is essentially a navigation problem within a "collapsed" space. Traditional acquisition functions, such as EI and UCB, primarily focus on objective minimization, often ignoring the catastrophic risk of scrapping a wafer when any single metric drifts outside its narrow target interval. To address this, we propose the Joint-Constraint-Aware Acquisition Function (JCAF). Acting as the framework's navigator, JCAF implements a scientific decision-making procedure by decoupling performance improvement from the probability of constraint satisfaction, then fusing them into a robust joint optimization criterion.

• *Heteroskedastic Uncertainty Modeling:* To preserve the unique physical characteristics of different process metrics (e.g., the high sensitivity of etch rate versus the relative stability of profile angle), we avoid weighted summation, which often masks heteroskedastic noise. Instead, we maintain independent modeling for each objective k. For a given parameter set x, the prediction distribution for the k-th metric is defined as:

$$p_k(y_k|x) = \mathcal{N}\left(\mu_k(x), \sigma_k^2(x)\right), \tag{5}$$

where $\mu_k(x)$ and $\sigma_k^2(x)$ are the predicted mean and variance for the objective k, and $\mathcal{N}$ denotes the Gaussian distribution. This allows the framework to explicitly account for the distinct uncertainty structures of various plasma-chemical responses.

• *Joint Probability of the Golden Window:* In the constraint collapse scenario, a recipe is only valid if all metrics simultaneously reside within their respective target intervals $[l_k, r_k]$. We transform these hard physical constraints into a differentiable probabilistic landscape. For each metric, the probability of satisfying the interval constraint is computed as:

$$\Pr(y_k) = \Phi\left(\frac{r_k - \mu_k(x)}{\sigma_k(x)}\right) - \Phi\left(\frac{l_k - \mu_k(x)}{\sigma_k(x)}\right), \quad \text{for } l_k \le y_k \le r_k \tag{6}$$

where $\Phi(\cdot)$ is the standard normal Cumulative Distribution Function (CDF), $\mu_k(x)$ and $\sigma_k(x)$ are the mean and standard deviation predicted for the objective k. To ensure robustness and avoid the overconfidence of independent thresholding, we define the Joint Feasibility Probability as the product of individual probabilities:

$$\Pr_{\text{joint}}(x) = \prod_{k=1}^{K} \Pr(l_k \le y_k \le r_k). \tag{7}$$

This multiplicative formulation acts as a strict safety filter, penalizing any region where even a single metric shows a high risk of failure.

• *Risk-Aware Navigation Logic:* We formulate the JCAF by combining the traditional Expected Improvement with the joint feasibility probability. JCAF dynamically balances objective improvement against feasibility, ensuring efficient search based on constraints. Traditional EI is defined as:

$$z = \frac{f_{\min} - \mu(x)}{\sigma(x)}, \tag{8}$$

$$\text{EI}(x) = \begin{cases} (f_{\min} - \mu(x))\Phi(z) + \sigma(x)\phi(z), \\ \qquad \text{for } \sigma > 0, \\ 0, \quad \text{for } \sigma = 0, \end{cases} \tag{9}$$

where $f_{\min}$ is the current best observed objective value, $\mu(x)$ and $\sigma(x)$ are the predicted mean and standard deviation of the objective function, and $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal cumulative distribution function and probability density function, respectively. Finally, the complete JCAF acquisition function is formulated as:

$$\text{JCAF}(x) = \text{EI}(x) \cdot \Pr_{\text{joint}}(x). \tag{10}$$

Here, EI(x) targets the improvement of a primary performance indicator. The core philosophy of JCAF is Risk-Aware Path Planning. As a navigator, it mimics the decision-making of an experienced process engineer: it automatically balances the pursuit of aggressive performance gains with the need to remain within the "golden" intervals. By prioritizing regions that offer both high potential and high confidence of compliance, JCAF effectively prevents the optimization process from becoming trapped in locally optimal but physically infeasible zones, ensuring a high success rate in searching the narrowest process windows.

3.4 ResSAN-DTS: A High-Fidelity Model

In this work, since direct physical wafer trials are prohibitively expensive, we develop ResSAN-DTS as a high-fidelity process emulator to serve as the objective black-box function for optimization. To bridge the gap between theory and industrial practice, ResSAN-DTS is trained on a high-fidelity dataset that captures the intricate plasma dynamics, serving as a reliable "digital twin" for algorithm validation where physical wafer trials are cost-prohibitive. It captures the intricate parameter-to-performance mappings by leveraging residual learning and SiLU activation, providing a continuous and differentiable landscape that represents the actual etching physics. The DTS (Dynamic Training Strategy) further ensures that this emulator remains robust against the stochastic noise inherent in real-world plasma reactions.

To accurately emulate the intricate parameter-to-performance mapping of the etching process, Deep Residual SiLU-Activated Network (ResSAN) incorporates several advanced structural features (Figure 5):

• *Deep Residual Learning:* Inspired by ResNet [19], we employ residual connections to handle the deep hierarchical dependencies between multi-physical variables (e.g., the coupling of gas chemistry, RF power, and pressure). This architecture ensures stable gradient

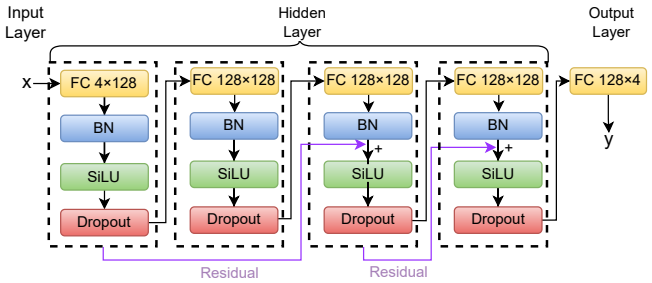


Figure 5: The architecture of the process emulator model.

flow and allows the model to capture strong non-linear effects without the degradation typical of standard MLPs.

• *SiLU Activation for Physical Continuity:* Unlike the piecewise linear ReLU, the Sigmoid-weighted Linear Unit (SiLU) [20] provides a smooth and non-monotonic activation profile. This mathematical smoothness is particularly suited for simulating the continuous shifts in plasma density and radical concentrations, better reflecting the underlying physical-chemical transitions.

• *Regularization and Stability:* We integrate Batch Normalization (BN) [21] to stabilize training and Dropout [22] as a dual-role component. While it acts as a regularizer during training to prevent overfitting, it is further utilized for Monte Carlo Dropout during inference. Multiple stochastic passes generate a predictive distribution $\mathcal{N}(\mu, \sigma^2)$, providing the variance needed for JCAF to balance exploration and exploitation.

To mimic the heuristic learning process of a process engineer-moving from "coarse tuning" to "fine precision"-we introduce the Dynamic Training Strategy (DTS). DTS governs the evolution of the virtual etcher through two key mechanisms:

• *Adaptive Learning Rate Scheduling:* DTS dynamically adjusts the learning rate based on the current optimization stage, ensuring rapid convergence in early global exploration and refined stability during local exploitation within the "golden" intervals.

• *Early-Stopping-Based Robustness:* By monitoring the loss trends on noisy experimental data, DTS prevents the model from memorizing stochastic outliers, ensuring that the emulator remains a reliable guide for the acquisition function throughout the iterative BO.

By synergizing the expressive power of ResSAN with the adaptive robustness of DTS, the emulator model provides the high-fidelity landscape necessary for navigating the "collapsed" design space, ensuring that the final optimized recipes are both high-performing and physically sound.

4 Evaluation

4.1 Experimental Setup

The objective of the evaluation is to identify input parameters that simultaneously satisfy all output constraints within a predefined number of experimental iterations. The proposed framework is compared with several baseline optimizers to benchmark its performance: Random search (RS) [4], Simulated Annealing (SA) [23], BO-PI/UCB/EI, and Y-BO [13].

The datasets are derived from actual semiconductor manufacturing scenarios. Three different target intervals (i.e., Large, Medium, and Small) are used to evaluate the performance of the methods. To better distinguish the performance of algorithms and models in the experiment, different maximum iteration (Max Iter.) counts were also set. In this evaluation, we use two key metrics: Success Rate (SR) and Average Experiments (AvgE).

Table 1: Main Results Compared with Baselines

Model	Large		Medium		Small	
	SR (%)	AvgE	SR (%)	AvgE	SR (%)	AvgE
RS [4]	0	N/A	0	N/A	0	N/A
SA [23]	2.70	17.93	0.40	15.50	0	N/A
BO-PI	14.70	15.54	6.40	30.48	0	N/A
BO-UCB	10.90	31.10	1.40	24.50	0	N/A
BO-EI	38.30	31.45	6.60	35.83	0	N/A
Y-BO [13]	40.00	37.90	15.00	33.30	0	N/A
Etch-Explorer	**100.00**	**10.07**	**99.80**	**12.17**	**98.30**	**15.97**

Table 2: Comparison Study of Sampling Strategies

Sampling Strategy	SR (%)	AvgE
Etch-Explorer (Random)	91.50	9.66
Etch-Explorer (LHS)	92.00	9.47
Etch-Explorer (Sobol)	90.00	10.57
Etch-Explorer (Expertise-based)	95.00	10.82
Etch-Explorer (Corner)	92.00	9.05
Etch-Explorer (HAS)	**97.00**	**9.20**

• *Success Rate (SR)* is the percentage of independent optimization runs that find a feasible solution within the maximum iteration budget. This serves as the primary metric for evaluating overall effectiveness under constraints.

• *Average Experiments (AvgE)* is the mean number of experiments (iterations) required to find a feasible solution in successful runs. This metric measures sample efficiency and is related to cost savings.

Experiments run on an *Intel Xeon NVIDIA RTX 3090* system using *Python* with *PyTorch*, *GPyTorch*, and *BayesianOptimization*.

4.2 Main Results

To evaluate the overall performance of the proposed Etch-Explorer in navigating "collapsed" design spaces, we conduct a comprehensive benchmarking study against several state-of-the-art optimization methodologies. With a maximum of 50 iterations, the comparative results across different target intervals are summarized in Table 1.

The comparative data indicate that the integrated Etch-Explorer framework consistently yields superior search efficiency and reliability compared to existing methodologies. Specifically, our Etch-Explorer achieves a high success rate in both the Large and Medium intervals, requiring significantly fewer experiments (10.07 and 12.17, respectively) than other methods. In contrast, the best-performing baseline Y-BO attains success rates of only 40% and 15% for Large and Medium intervals, respectively. Classical BO variants exhibit even lower success rates, while Random Search and Simulated Annealing are largely ineffective. Notably, in the most challenging Small interval scenario, our framework remains the only method capable of achieving a non-zero success rate (98.30%), whereas all other baselines fail. These results demonstrate that Etch-Explorer effectively identifies valid recipes satisfying all joint-interval constraints within significantly fewer iterations, offering a more robust solution for complex process window discovery.

4.3 Efficacy of Initial Sampling Strategies

Based on the Etch-Explorer architecture, we further investigate the impact of different initial sampling strategies. To isolate the effect of sampling, all variants in this study utilize the same JCAF and ResSAN-DTS configurations. The results for the Small interval scenario with a maximum of 15 iterations are summarized in Table 2.

Table 3: Comparison Study of Acquisition Function

Acquisition Function	Medium		Small	
	SR (%)	AvgE	SR (%)	AvgE
Etch-Explorer (PI)	1.60	14.77	0	N/A
Etch-Explorer (UCB)	0.20	14.97	0	N/A
Etch-Explorer (EI)	72.30	10.47	57.80	11.54
Etch-Explorer (JCAF)	**99.10**	**11.90**	**97.00**	**9.20**

Table 4: Ablation Study on Emulator Model

Model Structure	Max Iter.: 50		Max Iter.: 15	
	SR (%)	AvgE	SR (%)	AvgE
Basic Model	**37.60**	**28.15**	**3.00**	**12.63**
+Deep+Res	39.60	38.46	2.80	10.86
+Deep+Res+DTS	100.00	15.75	53.80	11.88
+SiLU+DTS	100.00	15.18	61.00	11.26
+Deep+Res+BN+DTS	98.50	21.08	17.10	12.94
+Deep+Res+Dro+DTS	98.80	23.04	29.90	10.94
+Deep+Res+Dro+SiLU+DTS	100.00	14.30	66.40	11.76
Basic Model w. ResSAN	63.20	28.38	4.80	12.65
Basic Model w. ResSAN-DTS	**100.00**	**10.55**	**91.50**	**9.66**

Our proposed HAS achieves the highest success rate (97.00%) and the lowest average evaluation score (9.20). Compared to standard LHS, which uniformly distributes samples across the vast invalid space, HAS places physical pivots near process boundaries and high-potential regions strategically. This boundary awareness allows the emulator model to construct a reliable space skeleton early on, and proves that our HAS effectively mitigates search blindness in the initial stages of process tuning.

4.4 Efficacy of Acquisition Function

To isolate the contribution of the acquisition logic, we conduct a controlled variable study comparing JCAF with three classical Bayesian optimization strategies: PI, UCB, and EI. In this section, all methods utilize the same HAS and ResSAN-DTS to ensure a rigorous comparison, with a maximum of 15 iterations.

The results in Table 3 indicate that while traditional functions struggle with initial search blindness in sparse regions, the logic of JCAF provides the risk-aware navigation necessary to handle competing physical constraints. This confirms that the JCAF strategy is a primary driver of search efficiency when the feasible region is exceedingly narrow.

4.5 Ablation Study on Emulator Model

To dissect the contribution of each component within our proposed emulator, we conduct an ablation analysis under the Small interval scenario. This study evaluates the performance gains attributed to the ResSAN-DTS structure, with results summarized in Table 4.

The full ResSAN-DTS configuration achieves optimal performance, increasing the SR from 37.60% to 100% within 50 iterations. Thanks to the smooth SiLU activation and residual structure, the model accurately captures the sharp non-linear transitions that traditional Gaussian Process (GP) models tend to smooth over. Moreover, the DTS prevents overfitting to process noise, ensuring that uncertainty estimates remain reliable, which is a prerequisite for the JCAF navigator to make informed, risk-aware decisions.

5 Conclusion

This work proposes Etch-Explorer, a systematic Bayesian optimization framework designed to navigate the constraint collapse in plasma etching. By integrating boundary-aware sampling (HAS), risk-aware navigator (JCAF), and a high-fidelity deep modeling (ResSAN-DTS), our approach transforms recipe tuning from a blind search into a joint-constraint-aware discovery exploration. Experimental results confirm that Etch-Explorer significantly outperforms traditional methods, achieving a higher success rate with fewer wafer trials under stringent interval constraints. This work provides a robust and automated solution for semiconductor process development, reducing both research and development costs and complexity. Future research will focus on cross-process knowledge transfer via meta-learning.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62404021), the State Key Lab of Processors, Institute of Computing Technology, CAS (No. CLQ202504), the Fundamental Research Funds for the BUPT (No. 2025AI4S20), and the Research Initiation Project for Introduced Talents of BUPT (No. 510124058, 510224062).

References

[1] Keren J Kanarik et al. 2023. Human–machine collaboration for improving semiconductor process development. *Nature* 616, 7958 (2023), 707–711.
[2] Christian Szegedy et al. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.
[3] Darrell Whitley. 1994. A genetic algorithm tutorial. *Statistics and computing* 4, 2 (1994), 65–85.
[4] Zelda B Zabinsky et al. 2009. Random search algorithms. *Department of Industrial and Systems Engineering, University of Washington, USA* (2009), 1–16.
[5] Jun-Gang Xu, Yue Zhao, Jian Chen, and Chao Han. 2015. A structure learning algorithm for Bayesian network using prior knowledge. *Journal of Computer Science and Technology* 30, 4 (2015), 713–724.
[6] Concha Bielza et al. 2013. Parameter control of genetic algorithms by learning and simulation of Bayesian networks—a case study for the optimal ordering of tables. *Journal of Computer Science and Technology* 28, 4 (2013), 720–731.
[7] Jiao-Yun Yang et al. 2021. A heuristic sampling method for maintaining the probability distribution. *Journal of Computer Science and Technology* 36, 4 (2021), 896–909.
[8] Cristina De Luca et al. 2022. AI in semiconductor industry. In *Artificial Intelligence for Digitising Industry–Applications*. River Publishers, 105–112.
[9] Marzieh Khakifirooz et al. 2018. Bayesian inference for mining semiconductor manufacturing big data for yield enhancement and smart production to empower industry 4.0. *Applied Soft Computing* 68 (2018), 990–999.
[10] Ling Wang et al. 2021. A review of reinforcement learning based intelligent optimization for manufacturing scheduling. *Complex System Modeling and Simulation* 1, 4 (2021), 257–270.
[11] In-Beom Park et al. 2021. Scalable scheduling of semiconductor packaging facilities using deep reinforcement learning. *IEEE Transactions on Cybernetics* 53, 6 (2021), 3518–3531.
[12] Ensieh Ghaedy-Heidary et al. 2024. A simulation optimization framework to solve stochastic flexible job-shop scheduling problems—Case: Semiconductor manufacturing. *Computers & Operations Research* 163 (2024), 106508.
[13] Xiao Yang et al. 2025. Multi-Objective Bayesian Target Interval Optimization for Semiconductor Process Parameters. In *2025 International Symposium of Electronics Design Automation (ISEDA)*. 609–615.
[14] Sila Guler et al. 2021. Bayesian optimization for tuning lithography processes. *IFAC-PapersOnLine* 54, 7 (2021), 827–832.
[15] Peter I Frazier. 2018. A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811* (2018).
[16] Jasper Snoek et al. 2012. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* 25 (2012).
[17] Ilya O Tolstikhin et al. 2021. Mlp-mixer: An all-mlp architecture for vision. *Advances in neural information processing systems* 34 (2021), 24261–24272.
[18] Jon C Helton et al. 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety* 81, 1 (2003), 23–69.
[19] Kaiming He et al. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
[20] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. 2018. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural networks* 107 (2018), 3–11.
[21] Sergey Ioffe et al. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*. pmlr, 448–456.
[22] Nitish Srivastava et al. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15, 1 (2014), 1929–1958.
[23] Scott Kirkpatrick et al. 1983. Optimization by simulated annealing. *science* 220, 4598 (1983), 671–680.