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

Yujie Zhang, Xiao Yang, Kang Zhao, Jianwang Zhai

Beijing University of Posts and Telecommunications

Outline

- Introduction
- Preliminaries
- Method
  - Overall Framework
  - Heterogeneous Active Sampling (HAS)
  - Joint Constraint-Aware Acquisition Function (JCAF)
  - ResSAN-DTS Emulator
- Experiments
- Conclusions
- Thanks & Q&A

Introduction

- **Background**
  - Semiconductor manufacturing enters sub-5nm era.
  - Plasma etching is a critical process.
- **Challenges: Constraint Collapse**
  - Constraint collapse: golden intervals are extremely sparse.
  - High-dimensional design space, complex physical couplings.
- **Limitations of traditional methods**
  - Trial-and-error: low efficiency, high wafer cost.
  - Generic BO methods struggles with sharp nonlinear boundaries.
- **Research motivation**
  - To automate and robustly navigate the collapsed constraint space.

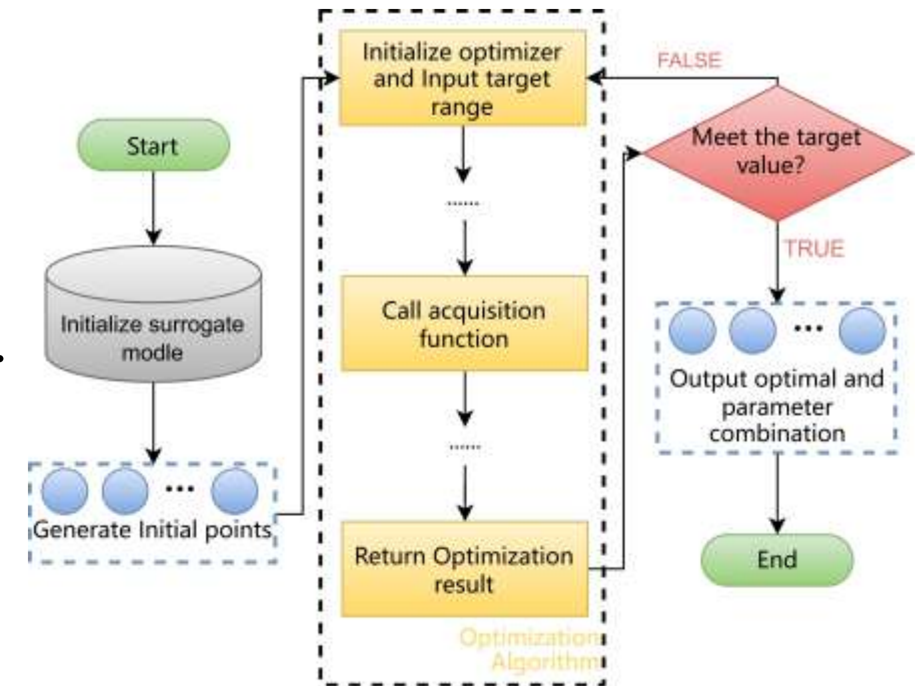

Figure: Illustrations of BO-based parameter tuning for plasma etching processes.

Preliminaries

- **Plasma etching workflow**
  - Inputs: Spacing, SRF, ESCTemp, Ar, etc.
  - Outputs: Rs, RsU, THK, THKU, etc.
- **Bayesian Optimization (BO)**
  - Surrogate model: Gaussian Process
  - Acquisition function: EI, PI, UCB
- **Problem formulation**
  - Find $\mathbf{x}^*$ such that $l_k \leq y_k(x) \leq r_k, \forall k$
  - Maximize joint satisfaction probability:

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

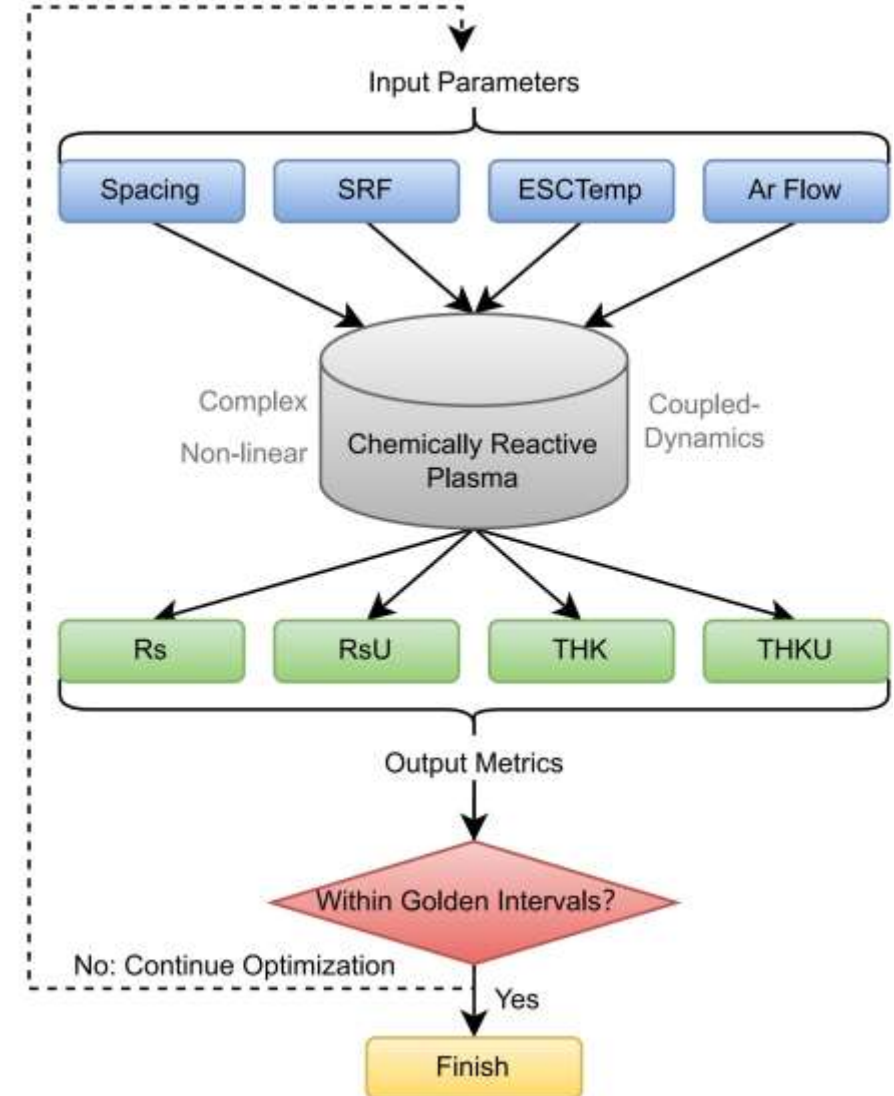

Figure: Illustrations of etching workflow.

Method - Overall Framework

- Etch-Explorer:
  - a joint-constraint-aware navigation engine via three synergistic architectural upgrades.
- HAS - Heterogeneous Active Sampling
  - Builds space skeleton and physical boundaries
- JCAF - Joint-Constraint-Aware Acquisition
  - Risk-aware navigator
- ResSAN-DTS - Deep residual process emulator
  - High-fidelity modeling of nonlinear couplings

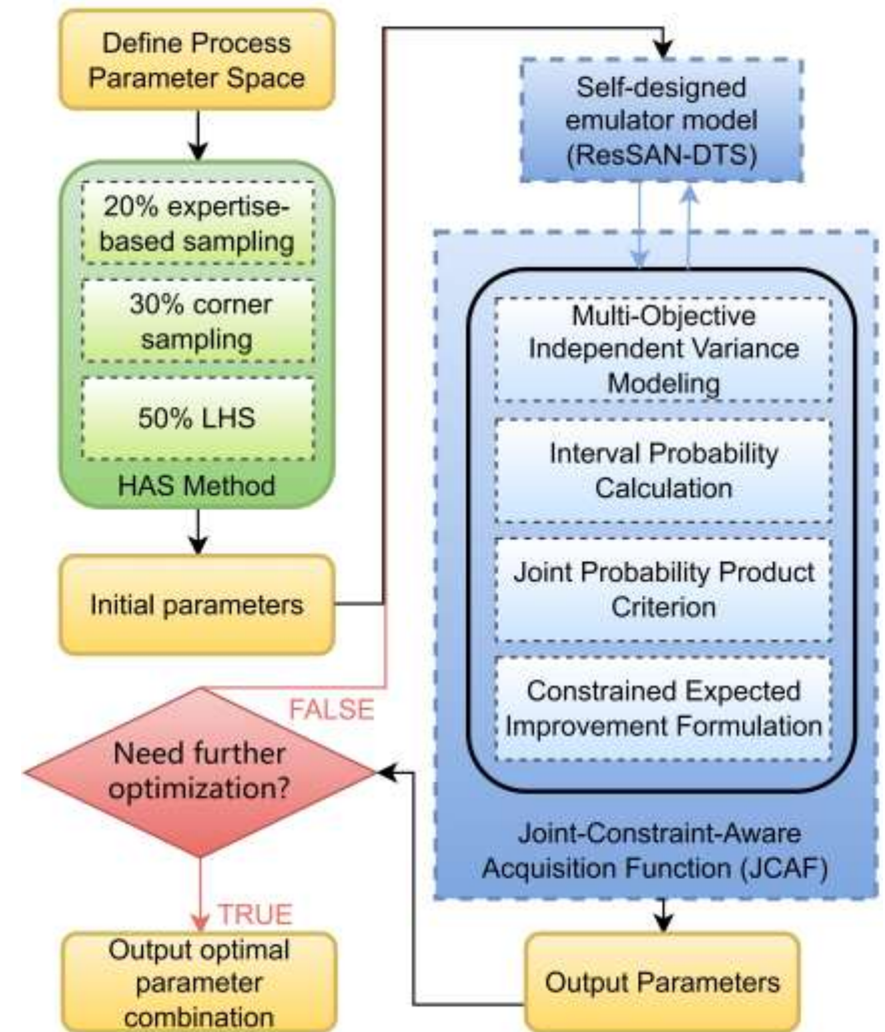

Figure: The Etch-Explorer framework for process parameter optimization in plasma etching

Method – Heterogeneous Active Sampling (HAS)

- **Problem**
  - Random/LHS wastes wafers on invalid regions;
  - Expertise alone misses non-intuitive optima.
- **Proposed Solution: Three Complementary Strategies**
  - LHS (50%) + Expertise-based (20%) + Corner sampling (30%)
- **Contribution**
  - This "space skeleton" gives the BO engine boundary awareness from the first iteration:
  - LHS ensures global coverage
  - Expertise anchors high-potential regions
  - Corner sampling identifies failure zones
  - The procedure is summarized in Algorithm 1

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 $S[0 : n]$

Method – Joint-Constraint-Aware Acquisition Function (JCAF)

- **Limitation of Traditional Functions**
  - EI, PI and UCB ignore the catastrophic risk of violating any single constraint.
- **JCAF Formula**
  - $\mathrm{JCAF}(x) = \mathrm{EI}(x) \cdot \mathrm{Pr}_{\mathrm{joint}}(x).$
- **Contribution**
  - Decouple performance improvement from constraint satisfaction and fuses them into a joint criterion.
  - Employ a multiplicative safety filter for joint constraint satisfaction.
  - Prioritize high-potential, high-confidence regions, avoiding infeasible local optima.

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

$$\mathrm{Pr}(y_k) = \Phi\left(\frac{r_k - \mu_k(x)}{\sigma_k(x)}\right) - \Phi\left(\frac{l_k - \mu_k(x)}{\sigma_k(x)}\right),$$

$$\mathrm{Pr}_{\mathrm{joint}}(x) = \prod_{k=1}^{K} \mathrm{Pr}(l_k \le y_k \le r_k).$$

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

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

Method – ResSAN-DTS Emulator

- **Limitation of Standard Models**
  - Standard models (MLP/GP) miss deep non-linear couplings and heteroskedastic noise.
- **ResSAN Architecture**
  - Residuals + SiLU + Dropout + BN
- **Dynamic Training Strategy (DTS)**
  - Adaptive learning rate
  - Early stopping → robust to stochastic noise
- **Contribution**
  - Provide a reliable predictive distribution N(μ, σ²) needed for the JCAF to make informed, risk-aware decisions.

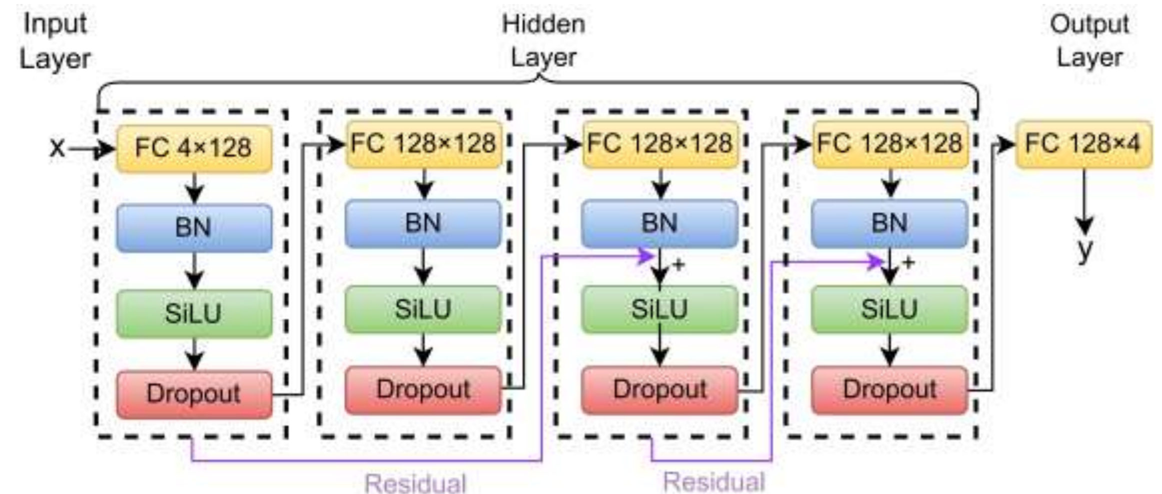

Figure: The architecture of the process emulator model.

Experiments - Setup

- **Baselines**
  - RS, SA, BO-PI/UCB/EI, Y-BO[1]
- **Dataset**
  - Real-world semiconductor manufacturing scenarios
- **Constraint Difficulties**
  - Large / Medium / Small (progressively stricter)
- **Metrics**
  - Success Rate (SR)
  - Average Experiments (AvgE)
- **Environment**
  - Intel Xeon, NVIDIA RTX 3090, PyTorch, GPyTorch

[1] 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.

Experiments

- **Setting:**
  - Large/Medium/Small interval, max 50 iterations
- **Analysis:**
  - Etch-Explorer is the only method that reliably succeeds under stringent constraints, with far fewer experiments.

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**

- **Goal:**
  - Isolate the HAS contribution. Same JCAF + ResSAN-DTS.
- **Setting:**
  - Small interval, max 15 iterations
- **Analysis:**
  - HAS achieves the highest success rate by synergistically combining global coverage (LHS), domain priors (Expertise-based), and boundary detection (Corner).

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**

Experiments

- **Goal:**
  - Isolate the JCAF contribution. Same HAS + ResSAN-DTS.
- **Setting:**
  - Medium/Small interval, max 15 iterations
- **Analysis:**
  - The JCAF is the primary driver of search efficiency under constraint collapse.

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**

- **Goal:**
  - Dissect the contributions of ResSAN-DTS components.
- **Setting:**
  - Small interval, max 50/15 iterations
- **Analysis:**
  - Residuals capture deep couplings.
  - SiLU ensures smoothness.
  - DTS is essential for preventing overfitting to noise.
  - The full model provides a high-fidelity "virtual etcher."

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**

Conclusions

- **What We Achieved**
  - Proposed Etch-Explorer for constrained plasma etching optimization.
  - Introduced three core innovations: HAS, JCAF, and ResSAN-DTS.
  - Significantly outperformed existing methods in success rate and sample efficiency.
- **Practical Impact**
  - Reduce both wafer experiments and R&D cost.
  - Enable automated discovery of optimal recipes under stringent constraints.
- **Future Work**
  - Cross-process knowledge transfer via meta-learning.

Thank You!

Authors: Yujie Zhang, Xiao Yang, Kang Zhao, Jianwang Zhai
Contact: zhangyujie0308@bupt.edu.cn
Beijing University of Posts and Telecommunications