Multi-Objective Bayesian Target Interval Optimization for Semiconductor Process Parameters

Paper ID: 25

Xiao Yang, Xingyu Qin, Yujie Zhang, Jianwang Zhai†, Kang Zhao
Beijing University of Posts and Telecommunications

Outline

Outline

Semiconductor Process Parameters Optimization — A High-Impact Yet Costly Task

- Why It Matters
  - Process parameters directly shape film quality, device performance, and yield
  - Small tuning errors → severe defects or variability

- What Makes It Difficult
  - Objective functions are black-box and multi-dimensional
  - Real experiments are slow, expensive, and non-repeatable
  - Classical tuning methods are inefficient and not generalizable

Why PVD Tuning Is Particularly Challenging

- A Critical Step in Modern Fabrication
  - PVD is key to forming conductive and barrier layers in interconnect stacks
  - Its output quality directly impacts performance
- Nonlinear and Coupled Parameter Effects
  - Outputs (e.g., thickness, uniformity) are sensitive to power, pressure, flow, etc.
  - Tuning often involves trade-offs across multiple conflicting goals

A PVD equipment used in advanced semiconductor fabrication

What Makes This Optimization Problem Unique

- Not Traditional Scalar Optimization
  - Aim to satisfy multiple intervals
  - Each output has a target range
- More than Multi-Objective Trade-offs
  - Our goal is not Pareto optimality, but to meet all constraints simultaneously
- A Constraint-Driven, Success-Oriented Task
  - We care about success rate
  - This setting arises naturally in industrial applications like PVD

Problem Formulation

- Choose input $\mathbf{x} \in \mathcal{X}$ so that all m outputs satisfy their target intervals:

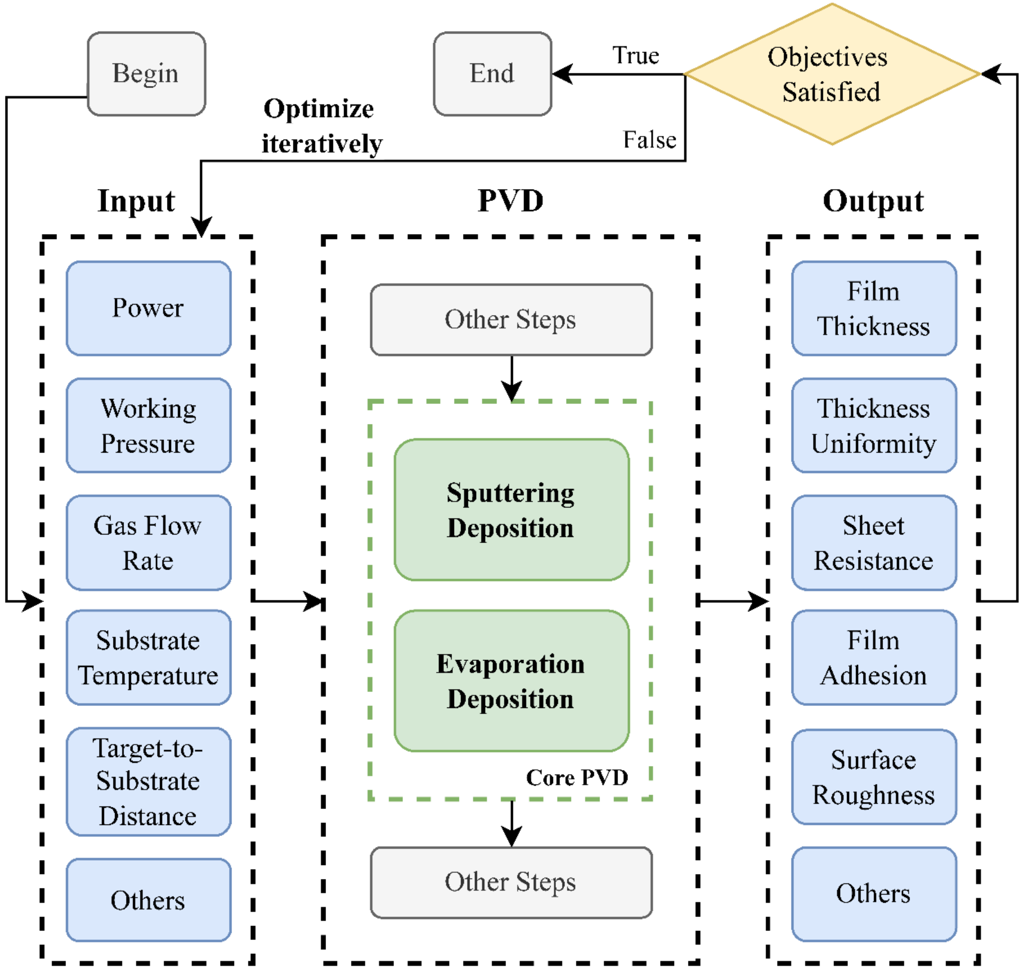
$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{X}} \mathbb{E}\left[\sum_{i=1}^{m} \mathbb{I}(y_i \notin [l_i, u_i])\right]$$

  - The indicator $\mathbb{I}(\cdot)$ penalizes violations
  - Problem = "Hit all intervals with minimal trials"
  - This is a constraint-focused optimization problem, solved using probabilistic modeling

Outline

PVD Process

- What is PVD?
  - A vacuum-based process for thin-film deposition in back-end semiconductor fabrication
- How it works: Core Process Flow
  - Input: tunable parameters
  - Core: physical deposition via sputtering or evaporation
  - Output: performance metrics

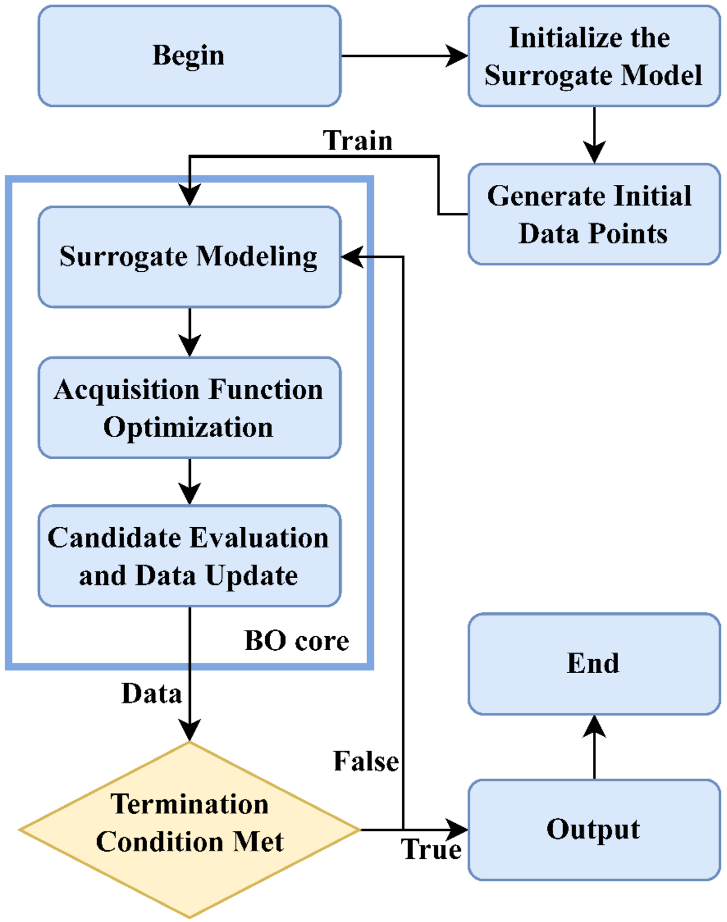


PVD Workflow and Key Process Parameters

Bayesian Optimization Basics

- Surrogate-Based Optimization
  - Use Gaussian Process (GP) to approximate expensive function
  - Acquisition function balances exploration and exploitation

- Standard BO Loop
  - Sample → Fit surrogate → Maximize acquisition → Evaluate → Update



Typical Workflow of BO

Inside the Surrogate: GP & SMTGP

- Gaussian Processes
  - Predict mean + uncertainty for any unseen input
  - Enables uncertainty-aware decisions

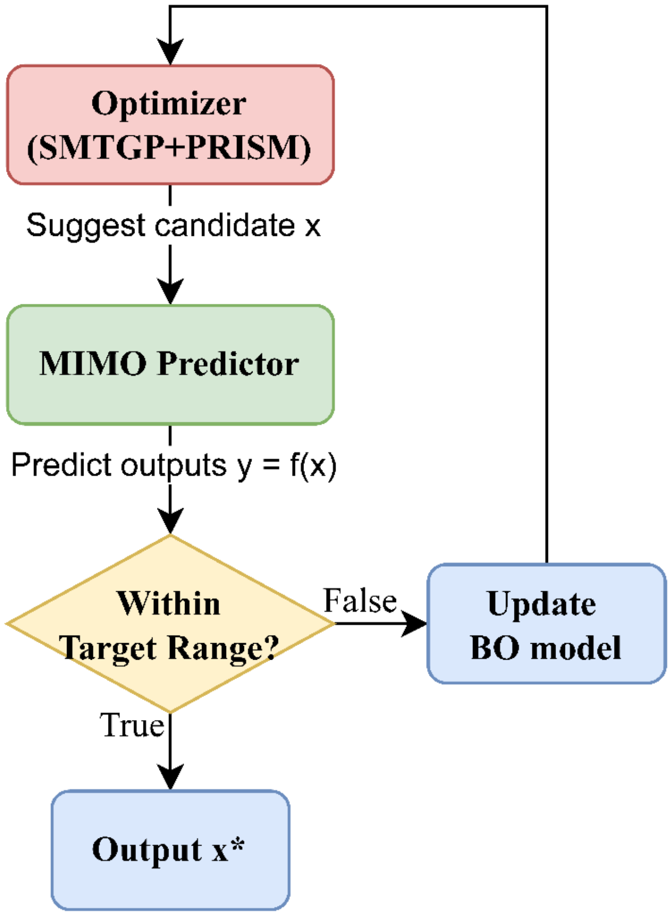
- SMTGP: **Sparse Multi-Task Gaussian Process**
  - Captures cross-objective correlations
  - Uses a small set of inducing points $P \ll N$ to reduce complexity from $O(N^3M^3) \to O(P^3M^3)$

P: *number of inducing points* N: *number of data points* M: *number of output tasks*

Outline

1. Introduction
2. Preliminaries
3. Proposed Method
4. Experiments
5. Conclusion

Framework Overview



- Two Core Modules
  - **Optimizer**: BO integrating SMTGP & PRISM
  - **Predictor**: Lightweight neural model for system behavior approximation
- Closed Optimization Loop
  - Suggest → Predict → Evaluate → Refine → Repeat

Key Innovations

- MIMO Predictor with Soft Physical Constraints
  – Learns from data while encouraging known physical trends

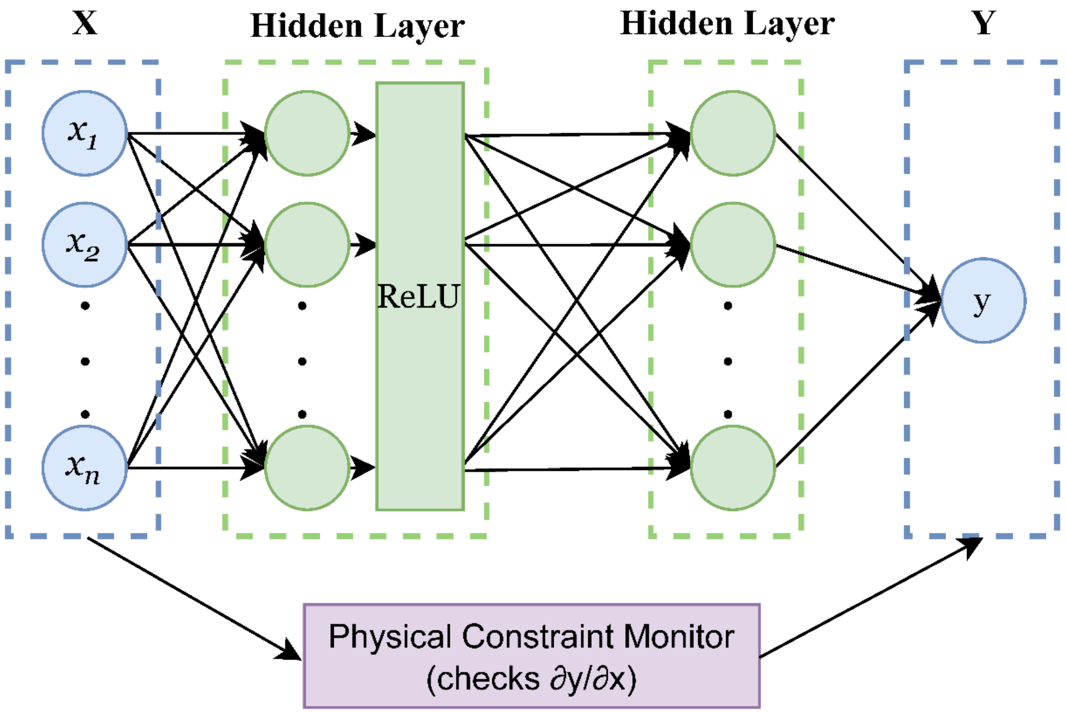
- Bayesian Optimization Framework with Interval Constraints

  A. SMTGP Surrogate Model
  – Multi-output GP with sparse inducing points
  – Captures cross-objective dependencies efficiently

  B. Acquisition Function: PRISM
  – Interval-probability-guided search
  – Selects points with highest joint interval satisfaction probability

MIMO Predictor

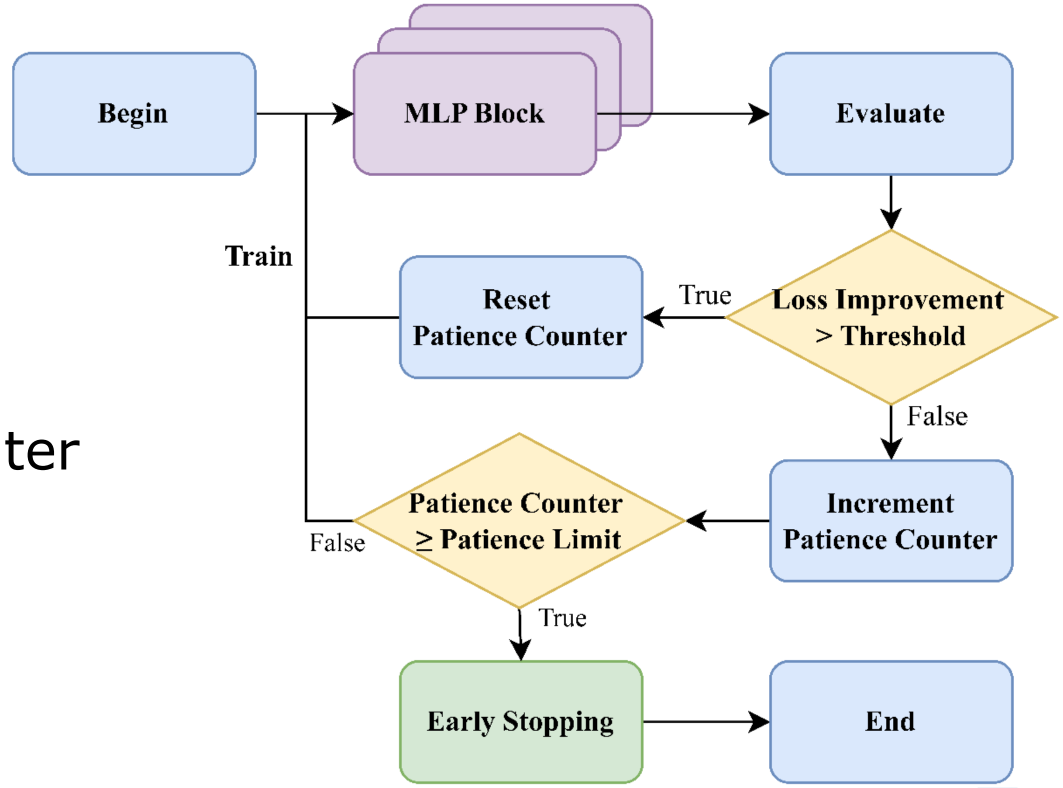
- Model Architecture
  - Shallow MLP with ReLU activation
  - Low cost, captures nonlinear patterns
- Constraint Regularization
  - Gradient-based penalty term that checks physical consistency (e.g., $\partial y/\partial x \geq 0$) during training
- Prior Integration
  - Soft constraints help improve accuracy with limited data



MLP Block (Detailed View)

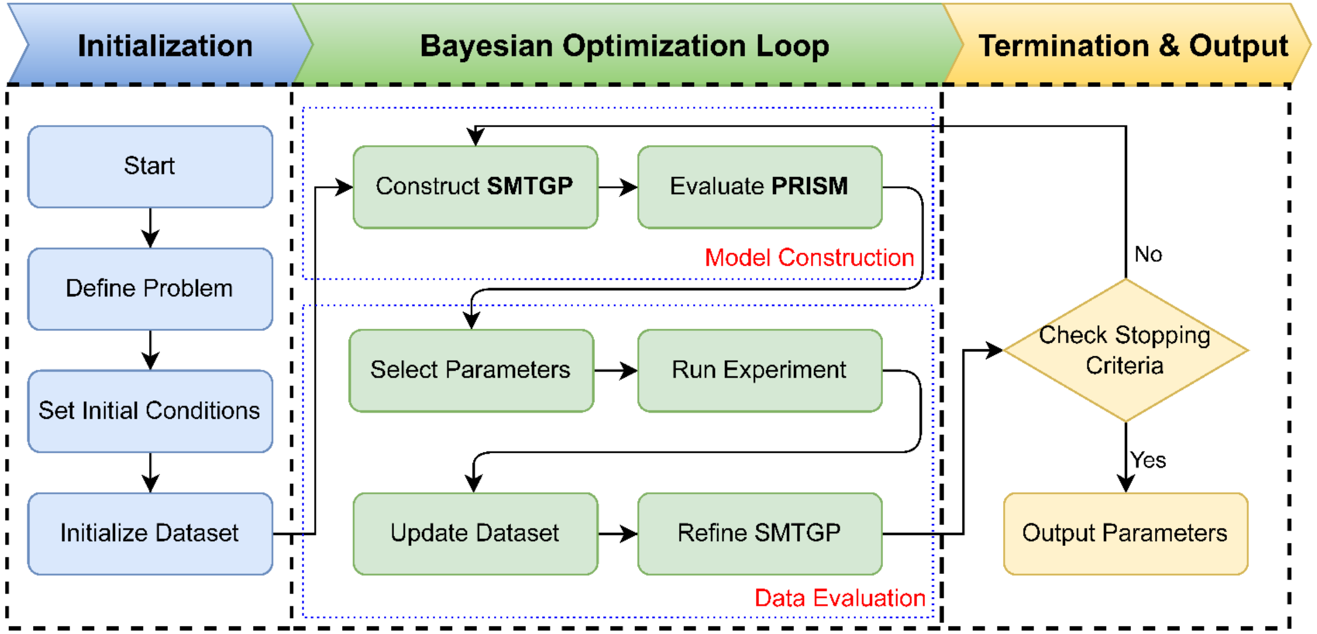
MIMO Predictor - Training Flow

- Training Workflow
  - Train MLP on initial data
  - Evaluate loss after each epoch
- Early Stopping Mechanism
  - If no improvement, increase patience counter
  - Stop training when patience exceeds limit



Training Workflow of MIMO Predictor

Intelligent Target Interval Optimization Algorithm



Enhanced BO Framework for
Target Interval PVD Parameter Tuning

- Motivation
  - Handle multi-objective constraints under scarce data
  - Enable interval-based optimization
- Framework Highlights
  - Tailored BO system for constrained industrial optimization
  - Combines model accuracy and constraint-aware decision making

SMTGP Surrogate

- Multi-Task Modeling
  - Shared kernel across objectives → Captures correlations
- Sparse Approximation
  - 20 inducing points → Retain accuracy while reducing cost
- Prediction Equations

$$\boldsymbol{\mu}(\mathbf{x}^*) = \mathbf{K}_{\mathbf{x}^*\mathbf{Z}}\mathbf{K}_{\mathbf{ZZ}}^{-1}\mathbf{m}_{\mathbf{Z}}$$

$$\Sigma(\mathbf{x}^*) = \mathbf{K}_{\mathbf{x}^*\mathbf{x}^*} - \mathbf{K}_{\mathbf{x}^*\mathbf{Z}}\mathbf{K}_{\mathbf{ZZ}}^{-1}\mathbf{K}_{\mathbf{Zx}^*}$$

$\mu(x*)$: *predicted mean at* $x*$ $\Sigma(x*)$: *predictive covariance matrix at* $x*$

K: *kernel (covariance) matrix*

Probability-guided Interval Search Mechanism (PRISM)

- Optimization Target
  - Maximize $P(L \leq y \leq U)$ over predicted distribution

- Mechanism: Interval Probability via Multivariate Gaussian
  1. For each candidate input x, SMTGP predicts a multivariate normal distribution of outputs:

$$y \sim N(\mu(x), \Sigma(x))$$

  2. PRISM computes the joint probability that all predicted outputs fall within their respective intervals:

$$P(\mathbf{L} \leq \mathbf{y} \leq \mathbf{U}) = \int_{\mathbf{L}}^{\mathbf{U}} \mathcal{N}(\mathbf{y} | \boldsymbol{\mu}, \Sigma) d\mathbf{y}$$

μ: *predicted mean vector* Σ: *covariance matrix* L, U: *target intervals*

Outline

Experimental Settings

- Dataset
  - Origin: Industry challenge from NAURA
  - High-dim variants synthesized with domain knowledge
- Benchmark Tasks
  - 4-input 4-output, 6-input 6-output, 10-input 10-output × Wide/Narrow ranges
- Evaluation Metrics
  - Aver., average number of evaluations
  - SR, success rate (all outputs fall into respective target intervals)

Main Results

- Performance Overview
  - SMTGP models output correlations to enhance sample efficiency in multi-objective settings
  - Interval-oriented PRISM directly optimizes for target satisfaction, not scalar objective
  - The only method that consistently succeeds in high-dimensional, constrained settings

TABLE I Main Experimental Results

Scenario	RS		BO		Ours	
	Aver.	SR	Aver.	SR	Aver.	SR
4I4O-W	217	21%	68	72%	**25**	**92%**
4I4O-N	476	14%	103	71%	**37**	**89%**
6I6O-W	N/A	N/A	183	52%	**105**	**81%**
6I6O-N	N/A	N/A	248	36%	**138**	**77%**
10I10O-W	N/A	N/A	N/A	N/A	**264**	**65%**
10I10O-N	N/A	N/A	N/A	N/A	**308**	**64%**

N/A: Success rate too low to produce statistically meaningful evaluation

Ablation Study

TABLE II Ablation Study: Impact of PRISM and SMTGP

Scenario	Full (Ours)		No PRISM		No SMTGP	
	Aver.	SR	Aver.	SR	Aver.	SR
4I4O-W	**25**	**92%**	70	37%	52	77%
4I4O-N	**37**	**89%**	107	36%	77	73%
6I6O-W	**105**	**81%**	285	35%	219	67%
6I6O-N	**138**	**77%**	391	31%	305	63%

- Key Insight: SMTGP and PRISM are both indispensable
  - SMTGP improves sample efficiency by modeling output correlations
  - PRISM guides the optimizer by focusing on target intervals
  - Joint effect: accurate modeling + goal-aware decision making

Outline

Conclusion Summary

- Addressed Problem
  - Optimize process parameters under multi-objective interval constraints
- Method
  - BO framework integrating SMTGP & PRISM + MIMO Predictor with Soft Physical Constraints
- Results
  - Achieves higher success rate with fewer evaluations, even under high-dimensional tuning challenges

Thank you for listening!