

Author version

Published as:

Wang, B., Jeong, S., van Schie, S. P. C., Han, D., Min, J., & Hwang, J. T. (2026).
Knowledge-guided generative surrogate modeling for high-dimensional design optimization under
scarce data. *Journal of Computing and Information Science in Engineering*, 26(7).
<https://doi.org/10.1115/1.4070934>

BibTeX for this reference:

<https://lsdolab.github.io/lsdo-publications/publication/wang2026knowledge/>

```
@article{wang2026knowledge,  
  author = {Bingran Wang and Seongha Jeong and Sebastiaan P. C. van Schie and Dongyeon  
    Han and Jaeho Min and John T. Hwang},  
  title = {Knowledge-Guided Generative Surrogate Modeling for High-Dimensional Design  
    Optimization Under Scarce Data},  
  journal = {Journal of Computing and Information Science in Engineering},  
  year = {2026},  
  volume = {26},  
  number = {7},  
  doi = {10.1115/1.4070934},  
  url = {https://doi.org/10.1115/1.4070934}  
}
```

Knowledge-guided generative surrogate modeling for high-dimensional design optimization under scarce data

Bingran Wang

University of California San Diego, La Jolla, CA, USA
University of Sydney, Sydney, NSW, Australia

Seongha Jeong

Samsung Electronics Co., Ltd, Republic of Korea

Sebastiaan P. C. van Schie

University of California San Diego, La Jolla, CA, USA

Dongyeon Han, Jaeho Min

Samsung Electronics Co., Ltd, Republic of Korea

John T. Hwang

University of California San Diego, La Jolla, CA, USA

Surrogate models are widely used in mechanical design and manufacturing process optimization, where high-fidelity computational models may be unavailable or prohibitively expensive. Their effectiveness, however, is often limited by data scarcity, as purely data-driven surrogates struggle to achieve high predictive accuracy in such situations. Subject matter experts (SMEs) frequently possess valuable domain knowledge about functional relationships, yet few surrogate modeling techniques can systematically integrate this information with limited data. We address this challenge with RBF-Gen, a knowledge-guided surrogate modeling framework that combines scarce data with domain knowledge. This method constructs a radial basis function (RBF) space with more centers than training samples and leverages the null space via a generator network, inspired by the principle of maximum information preservation. The introduced latent variables provide a principled mechanism to encode structural relationships and distributional priors during training, thereby guiding the surrogate toward physically meaningful solutions. Numerical studies demonstrate that RBF-Gen significantly outperforms standard RBF surrogates on 1D and 2D structural optimization problems in data-scarce settings, and achieves superior predictive accuracy on a real-world semiconductor manufacturing dataset. These results highlight the potential of combining limited experimental data with domain expertise to enable accurate and practical surrogate modeling in mechanical and process design problems.

I. Introduction

In many mechanical design and manufacturing process optimization problems, high-fidelity physics-based simulation models are either unavailable or prohibitively expensive due to the complexity and proprietary nature of the underlying processes. As a result, engineers must often rely on experimental measurements or a limited number of simulations to make informed decisions on process designs. Such data are typically scarce, which makes the optimization problem particularly challenging—especially in high-dimensional settings.

To overcome these challenges, surrogate models are widely employed to approximate system responses from data. Their effectiveness in design optimization, however, depends critically on achieving high predictive accuracy despite data scarcity. Common surrogate modeling techniques include radial basis functions (RBF) [1], kriging [2, 3], inverse distance weighting (IDW) [4], and neural networks [5]. Each method provides unique advantages depending on the complexity, dimensionality, and noise characteristics of the problem. These techniques have demonstrated success in a wide range of applications, including semiconductor manufacturing [6, 7], aerospace system design [8–10], mechanical and materials design [11–16], and chemical process design [17, 18]. Across these domains, surrogate models enable engineers to explore complex design spaces, accelerate optimization, and reduce reliance on expensive experiments or

simulations.

However, in practice, the performance of surrogate models often deteriorates when only a small amount of data is available. On one side, surrogate modeling methods such as radial basis functions (RBF) and kriging are known to perform well for low-dimensional problems with small datasets, but they lack mechanisms to directly incorporate engineering knowledge into the modeling process. In high-dimensional, data-scarce regimes, surrogates built solely from limited measurements often suffer from poor predictive accuracy, leading to suboptimal designs and minimal improvement in the quantities of interest. On the other side, neural network methods such as physics-informed neural networks (PINNs) [19] and generative adversarial networks (GANs) [20] can embed governing equations and distributional priors into training, thereby compensating for data scarcity. While these approaches reduce dependence on large datasets compared to purely data-driven neural networks, they remain limited in the types of prior knowledge they can incorporate and still face significant challenges in high-dimensional, data-scarce settings.

In many real-world settings, subject matter experts (SMEs) possess valuable domain knowledge about the functional relationships between design variables and process responses. This knowledge may include monotonicity with respect to certain variables, known output bounds, convexity or concavity, similarity to known analytical functions, or qualitative sensitivity trends. When properly incorporated into surrogate modeling, such information has the potential to substantially improve predictive accuracy and, in turn, the quality of optimization results. Moreover, these priors can guide the construction of a family of candidate functions that interpolate the available data, enabling more reliable predictions of design improvements. Despite this potential, no existing surrogate modeling framework systematically integrates such knowledge into the training process.

This paper addresses these challenges by introducing RBF-Gen, a novel surrogate modeling framework that integrates scarce data with domain knowledge. The method constructs an overcomplete radial basis function (RBF) representation and leverages a generative modeling approach to explore the null space of admissible interpolants, enabling the systematic incorporation of structural priors (e.g., monotonicity, convexity, and bounds) and distributional knowledge provided by domain experts. In contrast to existing physics-guided generative models, RBF-Gen does not rely on explicit governing equations or physical residuals. Instead, it embeds qualitative expert knowledge directly into the surrogate training process through penalty- and information-based objectives, ensuring exact interpolation of known data while allowing controlled variability consistent with physical intuition. By combining limited data with domain-informed priors, RBF-Gen serves as a natural extension of the standard RBF method, expanding its applicability to data-scarce problems where domain knowledge can be partially encoded as structural or distributional priors.

We evaluate the effectiveness of RBF-Gen on three case studies: (i) a one-dimensional cantilever beam optimization problem using simulation data, (ii) a two-dimensional shell thickness optimization problem using simulation data, and (iii) a semiconductor etching process surrogate modeling problem with real-world manufacturing data. Across all cases, RBF-Gen consistently achieves higher predictive accuracy and enables larger design improvements compared to standard RBF surrogates in data-scarce settings. These results demonstrate the value of combining limited experimental data with engineering knowledge, and highlight the potential of RBF-Gen as a practical and reliable surrogate modeling framework for complex mechanical design and manufacturing process problems.

The remainder of the paper is organized as follows. Section 2 reviews background on surrogate modeling, kernel methods, and generative adversarial networks. Section 3 presents the proposed RBF-Gen methodology. Section 4 reports numerical results on three case studies. Section 5 concludes with a summary and directions for future work.

II. Background

A. Surrogate modeling from scarce data

We consider a single-output surrogate modeling problem with N experimental measurements, denoted as

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N,$$

where $\mathbf{x}_i \in \mathbb{R}^d$ are the d -dimensional design or process variables, and $y_i \in \mathbb{R}$ is the corresponding quantity of interest (QoI). The objective is to construct a surrogate $\hat{f} : \mathbb{R}^d \rightarrow \mathbb{R}$ that accurately approximates the unknown mapping $f(\mathbf{x})$ based only on the limited dataset $\mathcal{D}$.

In many mechanical and manufacturing applications, the input dimension d can be large while the number of available data points N is very small, owing to the high cost and long turnaround time of experiments. This imbalance between d and N makes surrogate construction severely ill-posed: the data admit many possible interpolants, yet only a few are physically meaningful or predictive outside the training region.

B. Kernel-based surrogate modeling methods: RBF and kriging

Kernel methods are widely used for surrogate modeling because they provide flexible, nonlinear approximations from scattered data. The core idea is to represent the response $f(\mathbf{x})$ in terms of pairwise similarities between input points, encoded by a kernel function $\phi(r)$ or $k(\mathbf{x}, \mathbf{x}')$. Two of the most popular kernel-based surrogates in engineering design optimization are radial basis function (RBF) interpolation and kriging (also known as Gaussian process regression).

Radial basis function (RBF) The RBF method approximates $f(\mathbf{x})$ as a linear combination of radially symmetric basis functions centered at the training points:

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^N w_i \phi(\|\mathbf{x} - \mathbf{x}_i\|), \quad (1)$$

where w_i are interpolation weights and $\phi(r)$ is a radial kernel, such as the Gaussian

$$\phi(r) = \exp(-\epsilon^2 r^2) \quad (2)$$

or the thin-plate spline

$$\phi(r) = r^2 \log r. \quad (3)$$

Enforcing $\hat{f}(\mathbf{x}_i) = y_i$ for all training points leads to the square linear system

$$\Phi w = y, \quad \Phi_{ij} = \phi(\|\mathbf{x}_i - \mathbf{x}_j\|), \quad (4)$$

which has a unique solution if Φ is non-singular.

RBFs are effective in many settings but face two drawbacks in scarce-data regimes: (i) they yield a *single* interpolant, preventing exploration of alternative functions equally consistent with the data, and (ii) the basis is tied to the data locations, which can limit generalization when the experimental points are sparse or unevenly distributed.

Kriging (Gaussian process regression) Kriging, or Gaussian process regression (GPR), models $f(\mathbf{x})$ as a Gaussian process:

$$f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (5)$$

with mean function $\mu(\mathbf{x})$ and covariance kernel $k(\mathbf{x}, \mathbf{x}')$. The predictor at a new point $\mathbf{x}_*$ is

$$\hat{f}(\mathbf{x}_*) = \mu(\mathbf{x}_*) + k(\mathbf{x}_*, X) K^{-1} (y - \mu(X)), \quad (6)$$

where $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ encodes pairwise covariances. Kriging naturally provides uncertainty estimates and is widely used in design optimization for exploration–exploitation tradeoffs. However, its performance depends heavily on kernel hyperparameters and likelihood optimization, and in the small- N , high- d regime it often suffers from overfitting or numerical instability.

Together, RBF and kriging highlight both the promise and the limitations of kernel-based surrogates: while they interpolate sparse data well, they lack mechanisms to incorporate domain knowledge, and their predictive accuracy can degrade sharply when $N \ll d$.

C. Information maximization in generative modeling

The principle of *information maximization* (InfoMax) originated in neural network research as a way to learn informative internal representations by maximizing the mutual information between inputs and outputs [21, 22]. This information-theoretic perspective led to the development of models such as independent component analysis and the information bottleneck method [23], which emphasize learning representations that preserve task-relevant information while filtering out noise.

Building on these ideas, Chen et al. [24] proposed the Information Maximizing Generative Adversarial Network (InfoGAN), which encourages interpretable latent spaces by maximizing the mutual information between a subset of latent codes and the generated outputs:

$$I(c; \hat{f}_z) = H(c) - H(c | \hat{f}_z). \quad (7)$$

This formulation allows controllable variations in generated samples and has inspired subsequent work on mutual-information-based autoencoders and contrastive representation learning.

These methods have since found applications in areas such as computer vision [25], natural language processing [26], mechanical design [27] and aerodynamics [28]. Despite these advances, InfoGAN and related neural-network-based models are not directly suitable for surrogate modeling in engineering design problems, particularly in data-scarce settings. First, stable training typically requires thousands of samples, whereas only a few dozen experiments may be available in real-world mechanical design and manufacturing process problems. Second, they lack mechanisms to encode domain knowledge such as monotonicity, positivity, or convexity, which frequently arise in mechanical and manufacturing applications.

III. Methodology

Algorithm 1 Training procedure of the RBF-Gen framework

Require: Training dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, number of RBF centers $K > N$, kernel function $\phi(r)$, and expert-provided prior knowledge.

Step 1: Construct RBF basis and interpolation system.

- 1: Place K RBF centers $\{c_j\}_{j=1}^K$ uniformly or quasi-randomly in the design domain.
- 2: Form the interpolation matrix $\Phi_{ij} = \phi(\|x_i - c_j\|)$.

Step 2: Compute null space of the interpolation system.

- 3: Solve for a particular solution w_0 such that $\Phi w_0 = y$.
- 4: Compute null-space basis $N \in \mathbb{R}^{K \times (K-N)}$ such that $\Phi N = 0$.
- 5: Represent admissible interpolants as $w = w_0 + N\alpha$, where α are free coefficients.

Step 3: Define generator for null-space exploration.

- 6: Initialize a generator network $G(z; \theta)$ that maps latent variables $z \sim \mathcal{N}(0, I)$ to null-space coefficients $\alpha = G(z; \theta)$.
- 7: Each latent sample z produces a valid interpolant $f_z(x) = \Phi(x)^\top (w_0 + NG(z))$.

Step 4: Incorporate prior knowledge.

- 8: Define loss terms for structural priors and distributional priors.
- 9: Construct total loss following Equation 13.

Step 5: Train the generator.

- 10: Train the generator neural network to optimize θ .
-

To overcome the limitations of existing surrogate modeling methods, we propose the RBF-Gen framework, which extends the classical radial basis function (RBF) formulation to incorporate prior domain knowledge. By combining data with domain knowledge from subject-matter experts, the framework enables the construction of more accurate and reliable surrogate models for design optimization, particularly in data-scarce and high-dimensional settings. Constructing a surrogate model with the RBF-Gen framework consists of three steps:

- 1) Define RBF basis with K basis functions
- 2) Compute the nullspace of the RBF basis after interpolating $N < K$ data points
- 3) Train nullspace coefficients with domain knowledge

We cover these steps in the next sections.

A. Relaxed RBF centers: building a richer function space

Classical RBF interpolation ties each basis function center to a training point, yielding a unique interpolant. While this ensures exact data fit, it limits flexibility and prevents exploration of alternative functions equally consistent with the data. To overcome this limitation, we relax the formulation by decoupling the choice of RBF centers from the data locations.

We place $K > N$ centers $\{c_j\}_{j=1}^K$ uniformly (or quasi-randomly) across the design domain and represent the surrogate

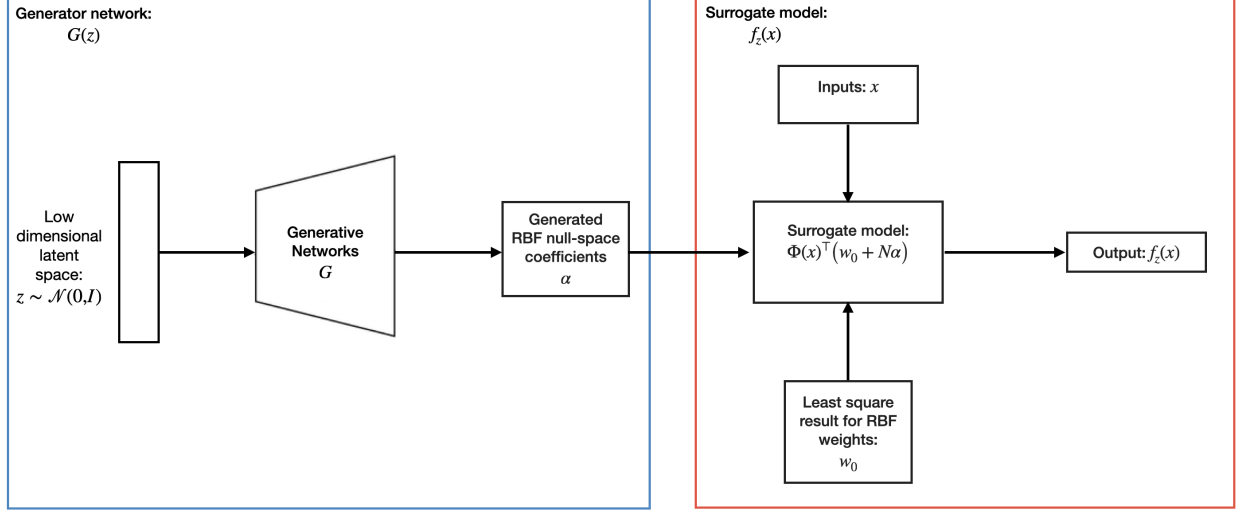


Fig. 1 Schematic illustration of the RBF-Gen framework.

as

$$\hat{f}(\mathbf{x}) = \sum_{j=1}^K w_j \phi(\|\mathbf{x} - c_j\|), \quad (8)$$

where $\phi(r)$ is a radial kernel and w_j are the RBF weights. Enforcing interpolation at the N data points leads to the underdetermined system

$$\Phi w = y, \quad (9)$$

where

$$\Phi_{ij} = \phi(\|x_i - c_j\|), \quad \Phi \in \mathbb{R}^{N \times K}, \quad (10)$$

with $w \in \mathbb{R}^K$ the RBF weight vector and $y = [y_1, \dots, y_N]^T$ the vector of data points located at $x = [x_1, \dots, x_N]$.

B. RBF nullspace: avenue for integration of additional knowledge

Since $K > N$, infinitely many solutions to equation (9) exist. Any solution can be expressed as

$$w = w_0 + N\alpha, \quad (11)$$

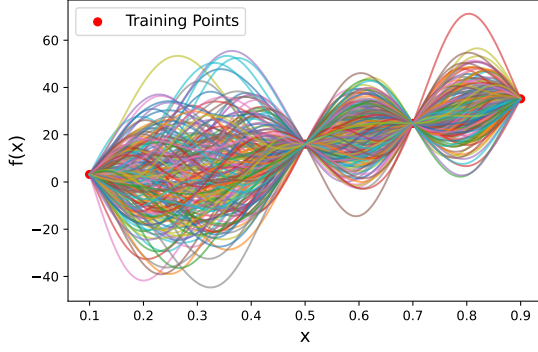
where w_0 is a minimum-norm particular solution to equation (9), $N \in \mathbb{R}^{K \times (K-N)}$ is a basis for the null space of Φ , and $\alpha \in \mathbb{R}^{K-N}$ is a free parameter vector. Each choice of α defines a distinct surrogate model $\hat{f}$ that interpolates the given data points.

This relaxed construction offers three key benefits:

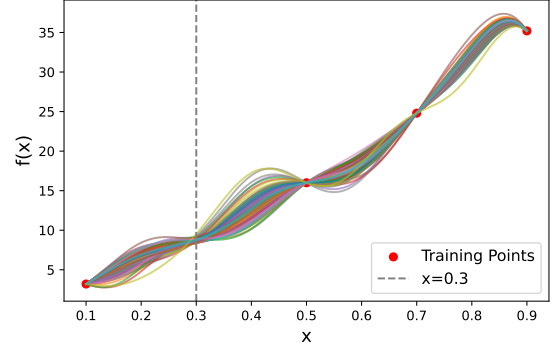
- **Flexibility:** The null space enables systematic generation of an entire family of admissible interpolants, rather than a single fixed function.
- **Improved generalization:** By distributing centers across the domain rather than restricting them to the data points, the surrogate covers unexplored regions more effectively.
- **Knowledge integration:** The free parameters α serve as controllable degrees of freedom, allowing prior domain knowledge to guide the selection of admissible interpolants.

C. RBF-Gen: Generator-based RBF surrogate modeling

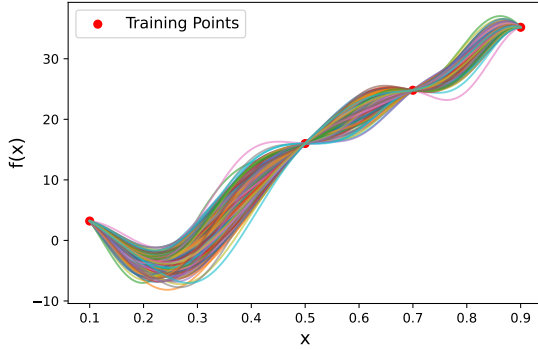
Building upon the RBF-based function generation framework, we extend it to incorporate prior knowledge using a generative modeling approach. Our goal is to generate functions that not only interpolate the known data points, but also exhibit desired statistical and structural characteristics that reflect domain-specific knowledge.



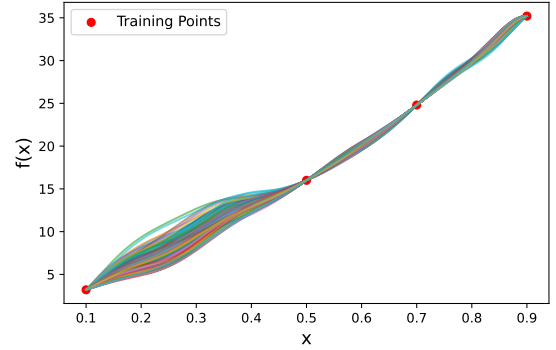
(a) RBF with relaxed centers



(b) RBF-Gen with distribution prior:
 $f(0.3) \sim \mathcal{N}(\text{true}, 1)$



(c) RBF-Gen with distribution prior:
Hessian $\sim \mathcal{N}(\text{estimated}, 1)$



(d) RBF-Gen with structural prior: monotonically increasing

Fig. 2 RBF-generated functions with and without generator extensions.

1. Generator formulation

We adopt a generator-based framework. The generator G maps latent variables $z \sim \mathcal{N}(0, I)$ into coefficient vectors α that lie in the null space of the interpolation system:

$$\hat{f}_z(x) = \Phi(x)^\top (w_0 + NG(z)), \quad (12)$$

where w_0 is a particular solution of the RBF interpolation system and N is a basis for its null space. Each latent sample z thus yields a valid interpolant that exactly satisfies the training data, while the variability of $G(z)$ enables sampling an ensemble of admissible surrogates.

2. Incorporating prior knowledge

Unlike classical surrogate models, which rely solely on the training data, RBF-Gen is able to flexibly incorporate different sources of prior knowledge, from pointwise physical constraints to global distributional properties. Two complementary mechanisms are used to introduce different types of prior knowledge: (i) *penalty terms*, which enforce structural properties such as monotonicity or positivity of $\hat{f}_z$, and (ii) *divergence terms*, which align the distribution of generated function statistics with prescribed target distributions.

The resulting penalty function for the generator is

$$\mathcal{L}_{\text{gen}} = \sum_i \lambda_i \text{pen}_i(\hat{f}_z) + \sum_j \gamma_j \text{KL}(p_{\text{gen}}(s_j) \parallel p_{\text{target}}(s_j)), \quad (13)$$

where pen_i are soft structural penalty terms (e.g., violations of pointwise quantities such as monotonicity or positivity),

KL denotes the Kullback–Leibler divergence, s_j are functional statistics of interest (e.g., extrema, averages, curvature), and λ_i, γ_j are tunable hyperparameters that allow one to place different weights on the various prior knowledge terms.

The resulting surrogate modeling approach, which we term **RBF-Gen**, produces ensembles of interpolants that both satisfy sparse data constraints and reflect expert-specified physical knowledge, without requiring adversarial discriminator training. A schematic illustration of the RBF-Gen framework is shown in Figure 1, and the corresponding pseudocode outlining its main steps is provided in Algorithm 1.

D. Different forms of information in RBF-Gen

We give an overview of how various common types of domain knowledge can be formulated for use in the RBF-Gen framework:

- **Penalty terms (structural priors)** These act pointwise on generated functions $\hat{f}_z(x)$ and are implemented as soft penalties on violations:

- *Monotonicity (increasing or decreasing):*

$$\text{pen}_{\text{mono}} = \frac{1}{|\mathcal{G}|} \sum_{x_k \in \mathcal{G}} \text{ReLU}(\pm (\hat{f}_z(x_{k+1}) - \hat{f}_z(x_k))) \quad (14)$$

where $+$ enforces non-decreasing and $-$ enforces non-increasing trends.

- *Positivity / bounds:*

$$\text{pen}_{\text{pos}} = \text{ReLU}(m - \min_{x \in \mathcal{G}} \hat{f}_z(x)) \quad (15)$$

ensures $\hat{f}_z(x) \geq m$ across probe points $\mathcal{G}$.

- *Lipschitz / slope bound:*

$$\text{pen}_{\text{Lip}} = \sum_{(x,y) \in \mathcal{P}} \text{ReLU}\left(\frac{|\hat{f}_z(x) - \hat{f}_z(y)|}{\|x - y\|} - L\right) \quad (16)$$

- *Smoothness / curvature:*

$$\text{pen}_{\text{curv}} = \sum_{x_k \in \mathcal{G}} \left(\hat{f}_z(x_{k+1}) - 2\hat{f}_z(x_k) + \hat{f}_z(x_{k-1}) \right)^2 \quad (17)$$

- *Convexity / concavity:*

$$\text{pen}_{\text{conv}} = \sum_{x_k \in \mathcal{G}} \text{ReLU}(-\Delta^2 \hat{f}_z(x_k)) \quad (18)$$

enforces convexity ($\Delta^2 f \geq 0$); flip sign for concavity.

- *Boundary / symmetry conditions:*

$$\text{pen}_{\text{bnd}} = \sum_{x_b \in \mathcal{B}} (\hat{f}_z(x_b) - v_b)^2 \quad (19)$$

matches known boundary values v_b .

- **KL divergence terms (distributional priors)** These shape the distribution of function statistics across generated surrogates:

- *Marginal values at a point x_0 :*

$$\text{KL}(p_{\text{gen}}(\hat{f}_z(x_0)) \parallel \mathcal{N}(\mu, \sigma^2)) \quad (20)$$

- *Regional averages:*

$$\text{KL}\left(p_{\text{gen}}\left(\frac{1}{|\mathcal{R}|} \sum_{x \in \mathcal{R}} \hat{f}_z(x)\right) \parallel p_{\text{target}}\right) \quad (21)$$

- *Extremal values (max/min):*

$$\text{KL}(p_{\text{gen}}(\max_{x \in \mathcal{R}} \hat{f}_z(x)) \parallel p_{\text{target}}) \quad (22)$$

- *Gradient magnitude distribution:*

$$\text{KL}(p_{\text{gen}}(\|\nabla \hat{f}_z(x)\|) \parallel p_{\text{target}}) \quad (23)$$

– *Curvature distribution:*

$$\text{KL}(p_{\text{gen}}(\frac{\partial^2 \hat{f}_z}{\partial x_i^2}) \parallel p_{\text{target}}) \quad (24)$$

– *Functional measures (integrals):*

$$\text{KL}(p_{\text{gen}}(\int_{\Omega} g(x, \hat{f}_z, \nabla \hat{f}_z) dx) \parallel p_{\text{target}}) \quad (25)$$

Through this design, RBF-Gen provides a unifying framework for combining sparse data with rich, heterogeneous domain knowledge, ensuring that generated interpolants remain both data-consistent and knowledge-consistent.

E. Hyperparameter selection and tuning

The performance and stability of the RBF-Gen framework depend critically on the appropriate choice of several hyperparameters, most notably the number of RBF centers (K) and the scaling factor for the null-space coefficients (α). These parameters jointly control the expressiveness of the generated functions and the ability of the model to explore the admissible function space without overfitting the limited training data.

The parameter K determines the size of the overcomplete RBF basis and directly influences the flexibility of the model. A larger K increases the dimensionality of the null space, thereby enriching the diversity of generated functions; however, it can also make generator training more challenging due to higher computational cost and potential instability. In this work, we adopt a simple heuristic to select K between two and five times the number of design variables (D), which provides a practical balance between model expressiveness and numerical stability across the test cases studied.

The scaling factor for α controls the magnitude of the generator outputs projected into the null space of the RBF functions. This parameter determines the range of admissible function perturbations and, consequently, the diversity of generated surrogates. In practice, the scaling factor is selected to ensure sufficient variability in the generated ensemble while maintaining smoothness and physical consistency in the resulting function shapes.

F. 1D demonstration

We demonstrate the effectiveness of the proposed RBF-Gen framework on a 1D function interpolation problem with scarce data. The underlying function is quadratic,

$$f(x) = 20x^2 + 20x + 1,$$

but it is sampled at only 4 locations, making interpolation highly uncertain. Figure 2 illustrates the outcomes of relaxed RBF interpolation and RBF-Gen when different forms of prior knowledge are incorporated into training. In particular, we consider priors based on target function values, curvature distributions, and monotonicity trends.

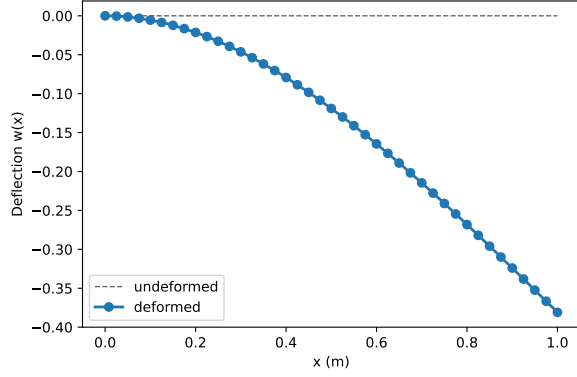
The results show that relaxing the RBF centers yields a wide variety of interpolants that exactly match the training points, reflecting the uncertainty inherent in sparse data. With RBF-Gen, however, each type of prior shapes the generated functions to better reflect the expected structural or statistical properties of the true function. These findings highlight that systematically embedding prior knowledge enables RBF-Gen to produce interpolants that not only satisfy the data constraints but also align more closely with the underlying ground truth.

IV. Numerical Results

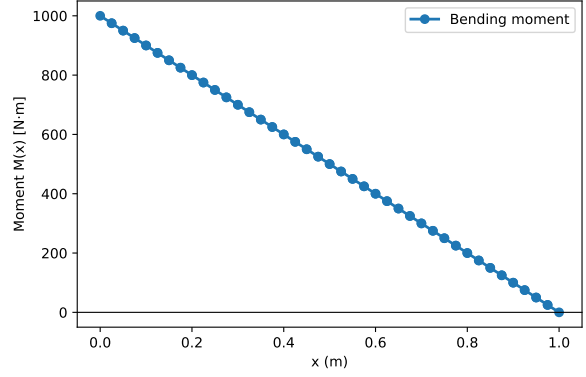
We demonstrate the effectiveness of the proposed RBF-Gen framework through three representative case studies of increasing complexity. First, we consider a **1D cantilever beam design optimization** problem, which serves as a simple yet illustrative benchmark for analyzing interpolation accuracy and design improvement under scarce data. Second, we study a **2D shell thickness optimization** problem, which introduces higher dimensionality and structural interactions, allowing us to assess the scalability of RBF-Gen to more complicated design problems. Finally, we apply RBF-Gen to a **semiconductor etching process** surrogate modeling problem, where only limited real-world experimental data is available. This case demonstrates the practicality of the proposed method in real-world manufacturing applications.

A. 1D cantilever beam design optimization

The first problem is a structural design optimization problem, which minimizes the compliance of a cantilever beam under a point load, subject to a volume constraint. This problem provides a flexible benchmark with a known

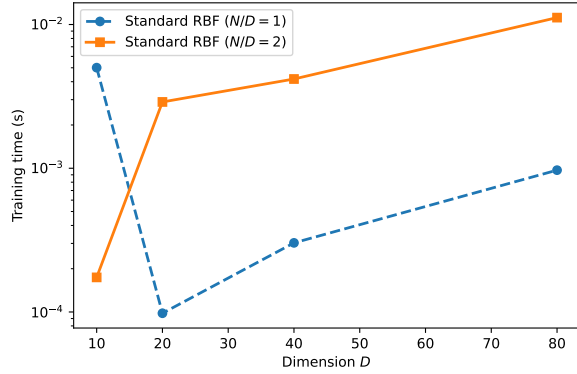


(a) Deflected shape of the beam.

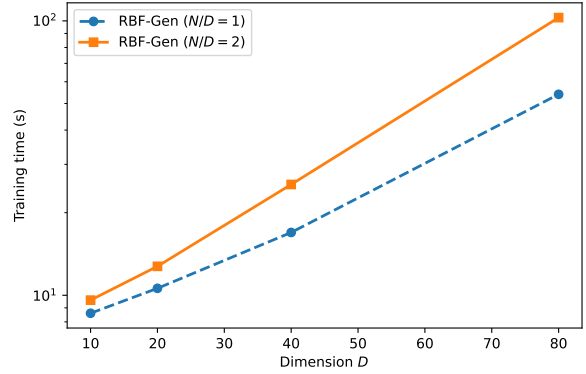


(b) Bending moment distribution.

Fig. 3 Structural response of a uniform cantilever beam.



(a) Standard RBF model.



(b) Proposed RBF-Gen model.

Fig. 4 Comparison of training time versus problem dimension D for (a) the standard RBF model and (b) the proposed RBF-Gen model under two data ratios ($N/D = 1$ and $N/D = 2$).

physics-based model, allowing us to evaluate the framework's ability to improve surrogate-driven optimization under data scarcity.

1. Problem description

In this problem, we have a cantilever beam that is discretized into D finite elements along its length, and modeled using Euler-Bernoulli beam elements. Each element i has a rectangular cross-section with fixed width b and adjustable height h_i . The vector of element heights $\mathbf{h} = [h_1, \dots, h_D]^\top$ defines the design variables. The objective is to minimize the beam's tip compliance:

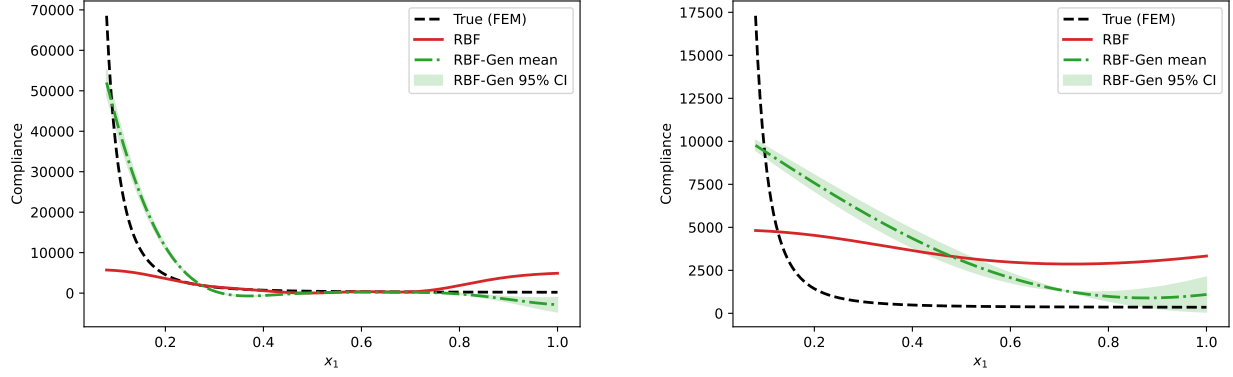
$$C(\mathbf{h}) = \mathbf{f}^\top \mathbf{d}, \quad (26)$$

where $\mathbf{f}$ is the nodal load vector and $\mathbf{d}$ is the displacement vector obtained by solving the finite element equilibrium equations:

$$\mathbf{K}(\mathbf{h}) \mathbf{d} = \mathbf{f}. \quad (27)$$

The optimization problem contains a volume constraint,

$$V(\mathbf{h}) = b \cdot \frac{L}{N} \sum_{i=1}^N h_i \leq 0.2, \quad (28)$$



(a) Slice plot for $D = 2, N = 4$.

(b) Slice plot for $D = 10, N = 20$.

Fig. 5 Slice plots of the cantilever compliance along x_1 comparing the true FEM response, standard RBF surrogate, and RBF-Gen surrogate (with 95% CI).

together with bound constraints on each element height,

$$0.05 \leq h_i \leq 0.5. \quad (29)$$

Figure 3 illustrates the structural response of a uniform cantilever beam in this problem.

2. RBF-Gen implementation details

To emulate realistic data scarcity, training samples of compliance values were generated from a small neighborhood of the design space, $0.05 \leq h_i \leq 0.1$, rather than spanning the entire feasible region. This reflects the practical setting where experimental data are often concentrated around nominal recipes.

Regarding the domain knowledge, we incorporated three sources of information when training RBF-Gen:

- **Monotonicity:** Compliance is expected to decrease as element thickness increases. For each design variable x_j , we constructed paired samples along its range while holding other variables fixed. Violations of the non-increasing trend were penalized.
- **Positivity:** Compliance values must remain strictly non-negative. A positivity penalty was introduced to ensure that all generated surrogates remain above zero across randomly probed points in the design domain.
- **Slice-based distributional priors:** For each design dimension, we selected one-dimensional slices of the design space (with other variables fixed at nominal values). At these slice points, compliance values were perturbed by $\pm 30\%$ around the finite-element "true" values to form target Gaussian distributions. The generated surrogate functions were then trained to match these target distributions by minimizing a KL divergence term.

The complete loss function for training the generator is

$$\begin{aligned} \mathcal{L} = & \underbrace{\lambda_{\text{mono}} \sum_{j=1}^N \sum_{k=1}^{G-1} \max(0, \Delta f_j^{(k)})}_{\mathcal{L}_{\text{mono}}} \\ & + \underbrace{\lambda_{\text{pos}} \sum_{x \in \mathcal{G}} \max(0, -f(x))}_{\mathcal{L}_{\text{pos}}} \\ & + \underbrace{\lambda_{\text{KL}} \text{KL}\left(p_{\text{gen}}(f(x)) \parallel \mathcal{N}(\mu_t, \sigma_t^2)\right)}_{\mathcal{L}_{\text{dist}}}, \end{aligned} \quad (30)$$

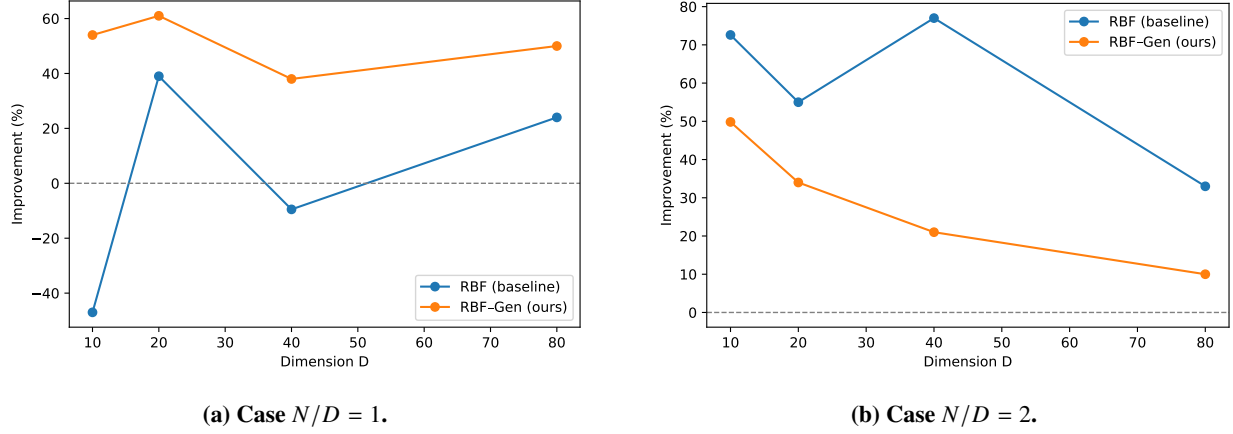


Fig. 6 Measured improvement (%) in the 1D cantilever beam optimization problem as a function of design dimension D . Here, N denotes the number of training samples used to fit the surrogate models and D denotes the number of design variables (beam elements). Results are shown for (a) $N/D = 1$ and (b) $N/D = 2$.

where μ_t, σ_t are the prescribed slice-based statistics derived from the true finite element evaluations, and p_{gen} is the empirical distribution induced by the generator across the sampled slices.

3. Results and discussion

The training time comparison between the standard RBF and the proposed RBF-Gen models is shown in Figure 4. As expected, the RBF-Gen model requires considerably higher computational time than the standard RBF model due to the additional steps of null-space computation and generator training. Moreover, the training time of RBF-Gen increases superlinearly with the problem dimension, reflecting the higher computational cost associated with exploring the admissible function space in higher dimensions. Nevertheless, in the context of practical mechanical and manufacturing process optimization problems, this additional computational overhead is typically negligible. When each experimental measurement or high-fidelity simulation can take hours or even days to complete, the training time of RBF-Gen, on the order of seconds to minutes, becomes trivial relative to the total cost of data generation.

The slice plots in Figure 5 illustrate the predicted compliance profiles along the first design variable, comparing the true finite element (FEM) response, the standard RBF surrogate, and the proposed RBF-Gen model. Each slice represents the predicted response when varying a single variable while fixing the others at their mean values, thereby providing an interpretable view of how each surrogate captures the underlying functional dependence. The results show that the functions generated by the RBF-Gen model follow the true FEM response much more closely than those obtained from the standard RBF surrogate, demonstrating that the inclusion of domain knowledge leads to more physically consistent predictions. However, as the dimensionality increases from $D = 2$ to $D = 10$, the generated functions deviate further from the true response, reflecting the increasing difficulty of accurately modeling high-dimensional problems with limited data. Despite this degradation, the RBF-Gen model continues to outperform the standard RBF approach, maintaining better alignment with the true physical trend.

Regarding the effectiveness of the surrogate models in design optimization, we use the *measured true improvement* as the performance metric. The measured true improvement quantifies the actual reduction in the objective function achieved through optimization based on each surrogate model, evaluated against a common initial design point. Specifically, for each case, surrogate-based optimization is performed using either the standard RBF or the proposed RBF-Gen model, and the resulting optimal design is re-evaluated using the true finite element model. The percentage reduction in the true objective function relative to the initial design is then reported as the measured true improvement. Figure 6 summarizes the measured true improvement (%) across different design dimensions D for two levels of data availability: (i) $N/D = 1$, where the number of training samples (N) equals the number of design variables (D), representing a highly data-scarce regime, and (ii) $N/D = 2$, where twice as many training samples as design variables are available.

In the scarce-data setting ($N/D = 1$, Fig. 6a), the proposed RBF-Gen framework consistently outperforms the standard RBF surrogate. Across dimensions ranging from $D = 10$ to $D = 80$, RBF-Gen achieves substantially higher

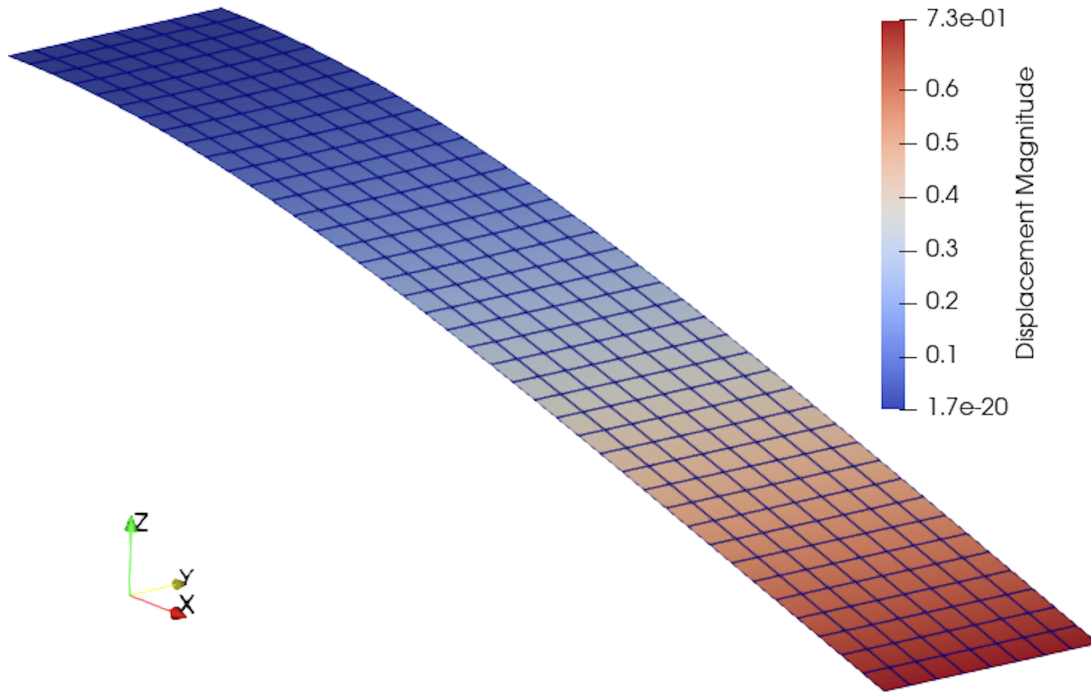


Fig. 7 Structural response of a cantilever thin plate subject to a distributed vertical load.

design improvements. These results highlight the strength of RBF-Gen in leveraging prior knowledge to guide surrogate training when data are extremely limited.

In contrast, when more training samples are available ($N/D = 2$, Fig. 6b), the performance of RBF-Gen deteriorates relative to the baseline RBF. Here, the baseline RBF achieves stronger improvements on average, while RBF-Gen tends to underperform. This behavior reflects the fact that, with richer data, in this problem, the standard RBF surrogate can already capture the input–output mapping effectively, while the additional priors imposed by RBF-Gen may introduce unnecessary biases.

Overall, these results demonstrate that, in this problem, RBF-Gen is particularly beneficial in data-scarce regimes, where classical surrogates suffer from poor generalization. However, as the data-to-dimension ratio improves, conventional RBF interpolation regains its advantage due to the direct availability of sufficient data to constrain the surrogate model.

B. 2D cantilever shell design optimization

The second problem is a shell thickness optimization problem. We model the steady-state deformation of a cantilever thin plate with the Reissner–Mindlin (RM) shell model. This test case allows us to demonstrate RBF-Gen’s performance in optimization problems with more than one spatial dimension and high-dimensional optimization variable spaces.

1. Problem description

We model the steady-state deformation of a $2\text{ m} \times 10\text{ m}$ cantilever plate using the Reissner–Mindlin (RM) shell model using FEniCS. Previously, FEniCS-based [29] steady-state and time-dependent implementations of the RM shell model have been used for design optimization in [30–32]. We refer the interested reader to [33] for a detailed description of the RM shell model.

For this problem, the plate is clamped on one edge and subjected to a distributed vertical load on the opposite edge. The shell thickness is taken as the design variable, discretized over the mesh elements of the domain Ω as a

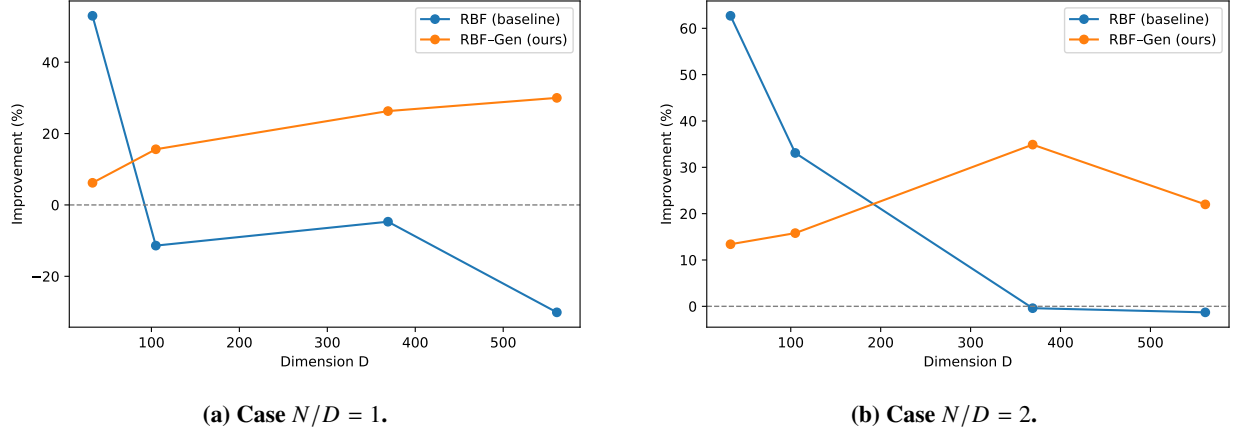


Fig. 8 Measured improvement (%) in the 2D cantilever shell optimization problem as a function of design dimension D . Here, N denotes the number of training samples used to fit the surrogate models and D denotes the number of design variables (shell elements). Results are shown for (a) $N/D = 1$ and (b) $N/D = 2$.

linearly-varying function. The objective is to minimize the compliance,

$$C(\mathbf{t}) = \mathbf{f}^\top \mathbf{d}, \quad (31)$$

where $\mathbf{f}$ is the applied nodal force vector representing a distributed vertical load, and $\mathbf{d}$ is the displacement vector obtained from the RM equilibrium equations. As in the beam case, the optimization problem includes a volume constraint,

$$V(\mathbf{t}) = \sum_{i=1}^N \int_{\Omega} t_i \phi_i d\Omega \leq V_{\max}, \quad (32)$$

where $t_i \in [0.01, 0.2]$ is the i^{th} thickness degree of freedom and ϕ_i its corresponding finite element basis function. $V_{\max}$ is set equal to the volume of a plate with uniform thickness of 0.1 m, so $V_{\max} = 2$.

2. RBF-Gen implementation details

The RBF-Gen setup follows the same formulation as in the cantilever beam problem. Training data are sampled from a small region of the design space, and prior knowledge is incorporated through a positivity penalty on compliance and monotonicity priors on the thickness variables. These additions ensure the generated surrogates remain physically consistent while interpolating sparse training data.

3. Results and discussion

We again compare the optimization performance of a standard RBF surrogate approach with the proposed RBF-Gen. We start by looking at the optimization improvements that are obtained when we have a single data point per design variable ($N/D = 1$). Figure 8a shows that the RBF surrogate drastically degrades in performance as the number of design variables is increased: The RBF surrogate finds an optimal design that performs worse than the initial design for all but the smallest considered case. This in contrast to RBF-Gen, which outperforms the RBF surrogate for all but the lowest-dimensional optimization problem considered, and consistently finds optimized designs that outperform the initial design.

Figure 8b shows how both models perform when we double the number of available data points ($N/D = 2$). While the performance of the RBF approach has improved relative to the results shown in Figure 8a, the RBF surrogate still fails to obtain any performance improvement in the higher-dimensional optimization cases. In contrast, RBF-Gen consistently finds optimal design points that improve upon the initial design. These results show that, in this problem using RBF-Gen has clear advantages for surrogate-based optimization over using a traditional RBF approach.

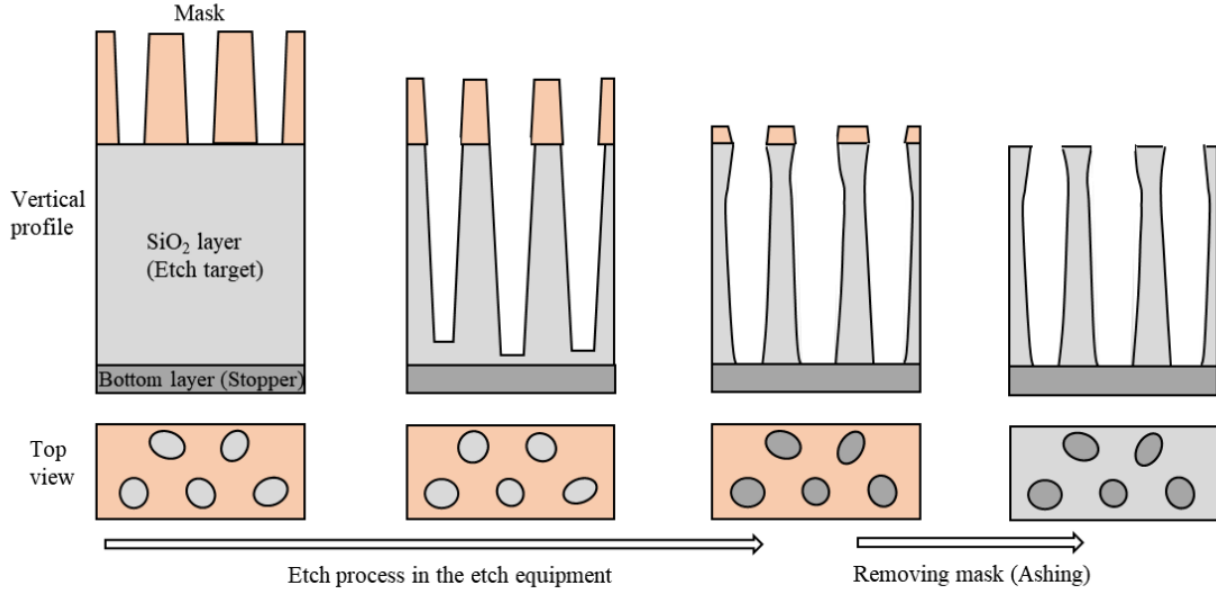


Fig. 9 Vertical profiles and top views during the etch and ash processes

C. Etching process data from semiconductor manufacturing

We evaluate the proposed RBF-Gen on a real-world dataset from the development of a High Aspect Ratio Contact (HARC) etching process in semiconductor manufacturing. The HARC etch has become a critical step as the design rule (DR) in DRAM continues to shrink and the string height in 3D NAND devices increases [34], requiring deep and narrow holes to be etched with uniform diameters and vertical profiles. Figure 9 illustrates a typical hole-etching process, in which a wafer coated with an SiO_2 layer and patterned photoresist (PR) mask is etched to produce high-aspect-ratio holes. Achieving the desired uniformity is extremely challenging due to the *depth-loading effect*, where the etch rate decreases as the aspect ratio increases [35], and the mask-induced *neck effect*, which causes a bowing region near the top of the hole profile [36]. These coupled effects lead to significant shape distortion, where the upper region becomes wider and the lower region narrower, making it difficult to maintain a vertical profile throughout the depth. To compensate for these nonlinear and competing phenomena, process engineers must iteratively adjust multiple recipe parameters—including process time, power, temperature, pressure, and gas flow rate—across several sequential steps. However, each experimental trial is expensive and time-consuming, resulting in very few measured data points relative to the number of controllable variables. This imbalance creates a severely scarce-data setting in a high-dimensional input space, posing a major challenge for constructing reliable surrogate models capable of assisting in process recipe optimization.

The dataset used in this study comprises 34 real fabrication experiments, each corresponding to a complete multi-step etch recipe. Each recipe is defined by 17 controllable input variables that characterize the physical conditions inside the chamber. These inputs include the process times of each recipe step, the RF power and temperature setpoints, the chamber pressure, and the flow rates of multiple process gases. Together, these variables capture the knobs available to process engineers for manipulating plasma-material interactions during the etch. Figure 10 shows the schematic of the etching process equipment.

The process outputs are measured as 5 quantities of interest (QoIs), which summarize critical aspects of etch quality. The primary objective QoI is the average critical dimension (CD) difference, defined as the difference between the largest and smallest diameters along the vertical profile of an etched hole. Minimizing this difference ensures uniformity and reduces the risk of electrical malfunction in the fabricated device. The remaining four QoIs act as constraints that evaluate different aspects of the etch profile, including sidewall verticality, hole spacing, and structural distortion. For confidentiality reasons, the exact variable names and values cannot be disclosed, and thus we present the input and output variables in anonymized form (Variable 1, Variable 2, ..., QoI 1, ...).

In this problem, in addition to the available experimental data, field experts provided partial information about the monotonic relationships between the input variables and the quantities of interest (QoIs). Table 1 summarizes

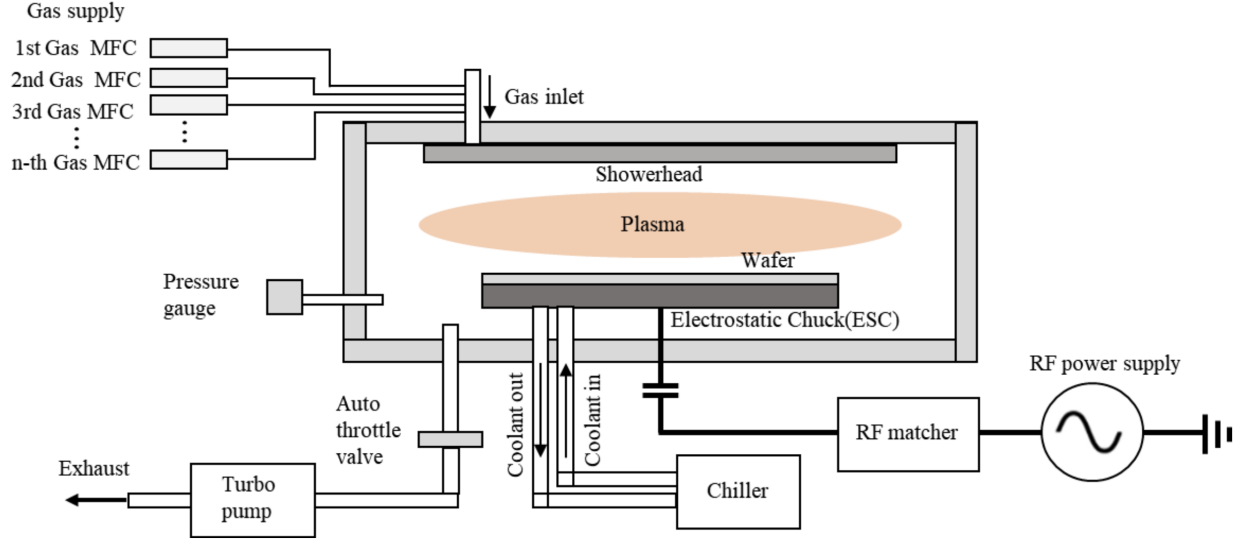


Fig. 10 Etching process equipment schematic

the redacted version of this expert-provided monotonicity information, where rows correspond to *Variable 1–17* and columns correspond to *QoI 1–5*. Each entry is marked as increasing (+1), decreasing (−1), or unknown (0). The objective of this problem is to construct accurate surrogate models for all five QoIs under severe data scarcity by integrating the 34 experimental measurements with the monotonicity knowledge provided by field experts.

To test the effectiveness of our proposed RBF-Gen method, we compared the standard RBF method which was trained purely from the dataset with the RBF-Gen method that uses both the dataset and the monotonicity information to train the surrogate model. For both methods, we combined with the standard partial least square method to reduce the input space to 5 dimensions to improve their effectiveness.

In this problem, we employ a Leave-Two-Out (L2O) cross-validation strategy to evaluate the predictive accuracy of the surrogate models. For $n = 34$ experimental samples, there are $\binom{34}{2} = 561$ folds. In each fold, the model is trained on $n - 2 = 32$ samples and evaluated on the two held-out samples, yielding a total of $2 \times 561 = 1122$ predictions per quantity of interest (QoI). We report the *Average Relative Error* (ARE) and *Average Absolute Error* (AAE), computed by aggregating the 1122 predictions across all folds. Table 2 summarizes the L2O errors for the two surrogate modeling approaches.

The results show that with only 34 experiments, purely data-driven surrogates exhibit limited predictive accuracy, while incorporating domain knowledge through the proposed RBF-Gen framework guides the surrogate toward physically consistent trends. Under the L2O evaluation, clear accuracy improvements are observed for three QoIs (with reductions in both ARE and AAE), while performance on the remaining QoIs remains comparable to the baseline. These results demonstrate that expert-informed priors can provide an effective means of enhancing surrogate reliability in data-scarce settings.

V. Conclusion

In this work, we introduced **RBF-Gen**, a knowledge-guided surrogate modeling framework that integrates sparse data with engineering priors. By leveraging the null space of an overparameterized RBF system and training a generator network inspired by InfoGAN, RBF-Gen enables the generation of function ensembles that simultaneously interpolate scarce training data and conform to expert-specified physical properties. Numerical experiments on cantilever beam and shell optimization problems, as well as a real-world semiconductor manufacturing case, demonstrated that RBF-Gen can significantly improve predictive accuracy and design optimization performance compared to standard RBF surrogates, particularly in data-scarce regimes.

A key limitation of the current framework lies in its sensitivity to parameter settings. The effectiveness of RBF-Gen depends on careful tuning of hyperparameters such as the number of RBF centers, null-space scaling, and penalty weights, which may limit its usability in practice. Another limitation is the absence of a mechanism to incorporate other

Table 1 Monotonicity information provided by field experts. A “+” indicates increasing, “-” indicates decreasing, and “0” indicates no prior information. Input and output names are redacted for confidentiality.

Variable	QoI 1	QoI 2	QoI 3	QoI 4	QoI 5
Variable 1	+1	-1	-1	+1	-1
Variable 2	0	0	0	0	0
Variable 3	0	0	0	0	0
Variable 4	-1	+1	+1	-1	-1
Variable 5	0	0	0	0	0
Variable 6	-1	+1	+1	-1	+1
Variable 7	-1	+1	+1	-1	0
Variable 8	-1	+1	+1	-1	0
Variable 9	-1	+1	+1	-1	+1
Variable 10	0	-1	-1	+1	0
Variable 11	0	-1	-1	+1	0
Variable 12	0	-1	-1	+1	+1
Variable 13	0	0	0	0	0
Variable 14	-1	+1	+1	-1	-1
Variable 15	-1	+1	+1	-1	-1
Variable 16	0	0	0	0	0
Variable 17	-1	+1	+1	-1	-1

forms of domain knowledge, such as periodic or oscillatory priors commonly encountered in mechanical and control systems. In addition, the accuracy of the proposed method depends on the quality of the domain knowledge provided, and its sensitivity to inaccuracies or inconsistencies in this knowledge has not yet been investigated.

For future work, it is essential to develop more principled heuristics or adaptive strategies for selecting and controlling these parameters to reduce reliance on manual tuning. In addition, extending the framework to encode a broader range of functional priors, including oscillatory and symmetric behaviors, represents an important step toward enhancing the generality and applicability of RBF-Gen. Another important direction is extending RBF-Gen beyond surrogate construction to directly enable uncertainty prediction of optimal design’s performance. In addition, an interesting and promising direction for future research is to explore the incorporation of physics-informed or variational discriminators, thereby extending the proposed RBF-Gen framework into a generative adversarial formulation to further enhance the flexibility and expressiveness of the model.

Overall, this study highlights the potential of combining scarce data with engineering knowledge to construct reliable surrogates for complex mechanical and manufacturing process design problems, and points the way toward more automated and robust knowledge-guided surrogate modeling frameworks.

Funding Data

- This work was supported by Samsung Electronics Co., Ltd. (Fund number: IO241018-11024-01)

References

- [1] Buhmann, M. D., “Radial basis functions,” *Acta numerica*, Vol. 9, 2000, pp. 1–38.
- [2] Krige, D. G., “A statistical approach to some basic mine valuation problems on the Witwatersrand,” *Journal of the Southern African Institute of Mining and Metallurgy*, Vol. 52, No. 6, 1951, pp. 119–139.
- [3] Matheron, G., “Principles of geostatistics,” *Economic geology*, Vol. 58, No. 8, 1963, pp. 1246–1266.

Table 2 Leave-Two-Out (L2O) errors per QoI and overall means. Reported are average relative error (ARE) and average absolute error (AAE).

QoI	RBF (baseline)		RBF+InfoGAN (ours)	
	ARE	AAE	ARE	AAE
QoI 1	0.4140	6.9710	0.3220	5.4200
QoI 2	0.1678	19.7025	0.3885	20.3100
QoI 3	0.0940	7.5008	0.4505	8.9500
QoI 4	1.4881	18.6276	0.6021	10.8200
QoI 5	0.6544	5.8356	0.2500	4.1100
Overall mean	0.5638	11.7275	0.5030	9.9220

- [4] Shepard, D., “A two-dimensional interpolation function for irregularly-spaced data,” *Proceedings of the 1968 23rd ACM national conference*, 1968, pp. 517–524.
- [5] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., “Learning representations by back-propagating errors,” *nature*, Vol. 323, No. 6088, 1986, pp. 533–536.
- [6] Wang, B., Kim, M. S., Yoon, T., Lee, D., Kim, B.-S., Sung, D., and Hwang, J. T., “Design optimization of semiconductor manufacturing equipment using a novel multi-fidelity surrogate modeling approach,” *Optimization and Engineering*, 2025, pp. 1–29.
- [7] Deenen, P. C., Middelhuis, J., Akcay, A., and Adan, I. J., “Data-driven aggregate modeling of a semiconductor wafer fab to predict WIP levels and cycle time distributions,” *Flexible services and manufacturing Journal*, Vol. 36, No. 2, 2024, pp. 567–596.
- [8] Cardoso, I., Dubreuil, S., Bartoli, N., Gogu, C., and Salaün, M., “Constrained efficient global multidisciplinary design optimization using adaptive disciplinary surrogate enrichment,” *Structural and Multidisciplinary Optimization*, Vol. 67, No. 2, 2024, p. 23.
- [9] Han, Z.-H., Görtz, S., and Zimmermann, R., “Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function,” *AIAA Journal*, Vol. 50, No. 9, 2012, pp. 1885–1896.
- [10] Shen, Y., and Alonso, J., “Graph Neural Network-Guided Aerodynamic Shape Optimization for Conceptual Design of Supersonic Transport Wings,” *AIAA AVIATION FORUM AND ASCEND 2025*, 2025, p. 3228.
- [11] Gaul, L., Kögl, M., and Wagner, M., *Boundary element methods for engineers and scientists: An introductory course with advanced topics*, Springer, 2009.
- [12] Haeri, A., and Fadaee, M. J., “Efficient reliability analysis of laminated composites using advanced Kriging surrogate model,” *Composite structures*, Vol. 149, 2016, pp. 26–32.
- [13] Zhang, Z., Poudel, L., Sha, Z., Zhou, W., and Wu, D., “Data-driven predictive modeling of tensile behavior of parts fabricated by cooperative 3d printing,” *Journal of Computing and Information Science in Engineering*, Vol. 20, No. 2, 2020, p. 021002.
- [14] Koo, G.-W., Lee, S.-M., and Kim, K.-Y., “Shape optimization of inlet part of a printed circuit heat exchanger using surrogate modeling,” *Applied thermal engineering*, Vol. 72, No. 1, 2014, pp. 90–96.
- [15] Ganti, H., and Khare, P., “Data-driven surrogate modeling of multiphase flows using machine learning techniques,” *Computers & Fluids*, Vol. 211, 2020, p. 104626.
- [16] Li, X., Wang, Y., and Sha, Z., “Deep learning methods of cross-modal tasks for conceptual design of product shapes: A review,” *Journal of Mechanical Design*, Vol. 145, No. 4, 2023.
- [17] Rao, H. N., Nair, S. K., and Karimi, I. A., “Operational optimization of processes with multistream heat exchangers using data-driven predictive modeling,” *Industrial & Engineering Chemistry Research*, Vol. 58, No. 15, 2019, pp. 5838–5850.

- [18] Misener, R., and Biegler, L., “Formulating data-driven surrogate models for process optimization,” *Computers & Chemical Engineering*, Vol. 179, 2023, p. 108411.
- [19] Raissi, M., Perdikaris, P., and Karniadakis, G. E., “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *Journal of Computational physics*, Vol. 378, 2019, pp. 686–707.
- [20] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., “Generative adversarial nets,” *Advances in neural information processing systems*, Vol. 27, 2014.
- [21] Linsker, R., “Self-organization in a perceptual network,” *Computer*, Vol. 21, No. 3, 1988, pp. 105–117.
- [22] Bell, A. J., and Sejnowski, T. J., “An information-maximization approach to blind separation and blind deconvolution,” *Neural Computation*, Vol. 7, No. 6, 1995, pp. 1129–1159.
- [23] Tishby, N., Pereira, F. C., and Bialek, W., “The information bottleneck method,” *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, 2000, pp. 368–377.
- [24] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., and Abbeel, P., “InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets,” *Advances in neural information processing systems*, Vol. 29, 2016.
- [25] Hjelm, R. D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., and Bengio, Y., “Learning deep representations by mutual information estimation and maximization,” *International Conference on Learning Representations (ICLR)*, 2019.
- [26] van den Oord, A., Li, Y., and Vinyals, O., “Representation learning with contrastive predictive coding,” *arXiv preprint arXiv:1807.03748*, 2018.
- [27] Chen, W., Lee, D., Balogun, O., and Chen, W., “GAN-DUF: hierarchical deep generative models for design under free-form geometric uncertainty,” *Journal of Mechanical Design*, Vol. 145, No. 1, 2023, p. 011703.
- [28] Ruh, M. L., Liu, X., Yu, R., and Hwang, J. T., “Airfoil shape parameterization using reconstruction-error-minimizing generative adversarial networks,” *AIAA AVIATION 2023 Forum*, 2023, p. 3722.
- [29] Baratta, I. A., Dean, J. P., Dokken, J. S., Habera, M., Hale, J. S., Richardson, C. N., Rognes, M. E., Scroggs, M. W., Sime, N., and Wells, G. N., “DOLFINx: the next generation FEniCS problem solving environment,” preprint, 2023. <https://doi.org/10.5281/zenodo.10447666>.
- [30] Xiang, R., van Schie, S. P. C., Scotzniovsky, L., Yan, J., Kamensky, D., and Hwang, J. T., “Automating adjoint sensitivity analysis for multidisciplinary models involving partial differential equations,” *Structural and Multidisciplinary Optimization*, Vol. 67, No. 146, 2024. <https://doi.org/10.1007/s00158-024-03847-2>.
- [31] van Schie, S. P., Ruh, M. L., Fletcher, A. H., Warner, M., Sperry, M., Scotzniovsky, L., Orndorff, N. C., Xiang, R., Yan, J., Zhao, H., Krokowski, J., Chen, J.-S., Sarojini, D., Gill, H., Lee, S., Tagg, A. C., Anderson, R., Green, E., Joseph, C., Ning, A., Cheng, Z., Cao, Z., Mi, C., Guibert, A. T., Cronk, A., Kim, A. A., Meng, S., Silva, C., and Hwang, J. T., “Large-Scale Distributed Multidisciplinary Design Optimization of the NASA Lift-Plus-Cruise Air Taxi Concept,” *AIAA SCITECH 2025 Forum*, 2025. <https://doi.org/10.2514/6.2025-0362>.
- [32] van Schie, S. P. C., Kramer, B., and Hwang, J. T., “Weighted Proper Orthogonal Decomposition for High-Dimensional Optimization,” 2025. URL <https://arxiv.org/abs/2508.09084>.
- [33] Campello, E. M. B., Pimenta, P. M., and Wriggers, P., “A triangular finite shell element based on a fully nonlinear shell formulation,” *Computational Mechanics*, Vol. 31, 2003, pp. 505–518. <https://doi.org/10.1007/s00466-003-0458-8>.
- [34] Choe, J., “Memory technology 2021: Trend & challenges,” *2021 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, IEEE, 2021, pp. 111–115.
- [35] Yen, T. F., Chang, K. J., and Chiu, K. F., “Profile control in high aspect ratio contact hole etching by a capacitively coupled plasma source,” *Microelectronic Engineering*, Vol. 82, No. 2, 2005, pp. 129–135.
- [36] Lee, J.-K., Jang, I.-Y., Lee, S.-H., Kim, C.-K., and Moon, S. H., “Mechanism of sidewall necking and bowing in the plasma etching of high aspect-ratio contact holes,” *Journal of The Electrochemical Society*, Vol. 157, No. 3, 2010, p. D142.