OPERATIONS RESEARCH MEETS MACHINE LEARNING: BAYESIAN OPTIMIZATION FOR ACCELERATING THE PRODUCT DEVELOPMENT IN ADDITIVE MANUFACTURING AND THERMOELECTRIC MATERIALS

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Ke Wang

Alexander W. Dowling, Director

Graduate Program in Chemical and Biomolecular Engineering
Notre Dame, Indiana
March 2025

OPERATIONS RESEARCH MEETS MACHINE LEARNING: BAYESIAN OPTIMIZATION FOR ACCELERATING THE PRODUCT DEVELOPMENT IN ADDITIVE MANUFACTURING AND THERMOELECTRIC MATERIALS

Abstract

by

Ke Wang

The design of chemical-based products and functional materials is vital to modern technologies, yet remains challenging due to costly and time-consuming manufacturing processes. These processes often involve tens of variables, resulting in hundreds of thousands of possible experiments, with each experiment carrying significant resource costs. In such high-dimensional domains, traditional Edisonian, trial-and-error approaches become prohibitively expensive and inefficient at identifying optimal experimental conditions. Consequently, there is a critical need to shift from these conventional methods to more systematic, data-driven decision-making.

Machine learning (ML) and operations research (OR) offer promising approaches to address these challenges through novel optimization frameworks. By building surrogate models that capture the relationships between decision variables and targeted objectives, ML enables predictive modeling of complex manufacturing processes. OR then integrates these pre-trained models into a unified optimization framework, facilitating data-driven, rational, and scientifically grounded decisions that accelerate product development while minimizing experimental costs.

In this work, we present an ML- and OR-based framework that combines Bayesian optimization (BO) with first-principles knowledge. We demonstrate its effectiveness

in solving industrial manufacturing optimization problems across additive manufacturing and thermoelectric material domains, including flash sintering, plasma sintering, and aerosol jet printing. Our framework accelerates the identification of optimal experimental conditions while reducing both economic and labor costs. Moreover, by incorporating physical knowledge in multiple ways, it is ideally suited for data-scarce, customized, and expensive experiments. It is a general framework that is not restricted to any single application domain.

CONTENTS

FIGURES

TABLES

CHAPTER 1

INTRODUCTION

The design of chemical-based products and functional materials is vital to modern technologies. Examples include coatings, fertilizers, food additives[1], medicines, engineered gene sequences[2], and functional devices[3]. However, developing these products remains challenging due to costly and time-consuming modern manufacturing processes. For example, optimizing the sintering process for thermoelectric materials (TE)[4] involves iterating in search space in $\mathcal{O}(10^6)$, where each requires days of experimentation, specialized equipment, and costly resources. The $\mathcal{O}(10^6)$ complexity arises from the combinatorial nature of the problem: the sintering process is influenced by multiple interdependent variables, such as temperature, pressure, heating rate, and material composition. Each variable can assume a range of values, and the interactions between these variables exponentially increase the number of possible experimental configurations.[5, 6] For instance, if there are six key variables, each with ten possible values, the total number of combinations scales as 10^6. Traditional Edisonian trial-and-error approaches become prohibitively expensive and inefficient at identifying optimal experimental conditions in such high-dimensional scenarios.

This challenge requires a paradigm shift from traditional Edisonian search toward systematic, data-driven methodologies. Two fields—machine learning (ML) and operations research (OR)—offer complementary tools to address this need. However, their integration remains underexplored in industrial manufacturing scenarios, especially in the thermoelectric material and additive manufacturing domain. This

thesis sits at the intersection of these disciplines, proposing a novel framework that integrates ML's predictive capabilities with OR's rigorous optimization strategies to accelerate the discovery of optimal manufacturing conditions.

Machine learning refers to algorithms that learn patterns from data to make predictions.[7] In chemical engineering and additive manufacturing, ML's leverage its ability to model complex, nonlinear relationships between process variables and target outcomes. For example, surrogate models approximate expensive physics-based simulations or experiments, enabling rapid virtual testing of parameter combinations[8]; dimensionality reduction methods simplify high-dimensional datasets, uncovering latent variables that govern material behavior[9]. However, purely data-driven ML faces three main limitations in industrial settings. First, data scarcity is due to highly customized manufacturing processes, leaving little historical data for training. Second, physical plausibility causes models to violate fundamental laws unless constrained by domain knowledge. Third, interpretability in black-box predictions hinders model adoption in safety-critical applications.

OR focuses on optimizing complex decision-making processes using mathematical and computational methods.[10] In manufacturing, OR methods enable systematic navigation of high-dimensional search spaces to identify conditions that maximize performance while minimizing costs. For example, BO sequential balance exploration and exploitation for the global optimization of expensive black-box functions.[11] Constraint optimization can incorporate domain-specific rules and black-box ML models into optimization.[12] Multi-objective optimization balances competing objectives, such as maximizing material functionality while minimizing substrate temperature.[13] OR's strength lies in its mathematical rigor and ability to handle constraints, but it often assumes the existence of accurate models. In practice, manufacturing processes are too complex to model purely from first principles, creating a reliance on data-driven approximations.

This thesis bridges ML and OR by addressing a critical gap: optimizing high-dimensional, constrained manufacturing processes when data is scarce and experiments are expensive. While ML provides surrogate models to approximate process behavior, OR provides the optimization frameworks to navigate the search space strategically. Their integration is not merely sequential (e.g., train ML model, then optimize) but deeply synergistic. Physics-informed ML models embedded domain knowledge, allowing predictions adhere to physical laws and improving accuracy. Constraint optimization allows OR algorithms to leverage ML models while enforcing hard or soft constraints. Adaptive experimentation BO frameworks actively update ML models with new data, dynamically refining the search strategy.

The main contributions of this thesis are as follows:

- A Bayesian optimization framework to enhance experimental design efficiency for improving product performance.
- A model-based design framework to accelerate the discovery of optimal experimental designs for model identification.
- Validation of the effectiveness of combining first-principles and machine learning for predictive modeling in data-scarce scenarios.
- Demonstration of the benefits of integrating first principles, machine learning, and optimization in solving complex industrial optimization problems.

This thesis is organized to systematically address the integration of machine learning and operations research in manufacturing optimization. Chapter 2 comprehensively reviews BO and its applications in chemical product design and functional materials development, establishing foundational insights into its industrial applicability. Chapter 3 introduces a BO framework with domain expertise to maximize TE material performance in flash sintering, illustrating how expert intuition improves search efficiency in complex parameter spaces. Building on this, Chapter 4 formalizes the quantification of domain knowledge and integrates it with multi-objective optimization, proposing a unified optimization framework for plasma sintering pro-

cesses. Chapter 5 extends this methodology further by developing physics-informed machine learning models and leveraging a model-based design of experiments (MB-DoE) to identify optimal models with minimal experimental design. Finally, Chapter 6 summarizes the key contributions, discusses its implications for broad manufacturing, and outlines future research directions in integrating operations research and machine learning.

1.1 References

1. Xiang Zhang, Teng Zhou, Lei Zhang, Ka Yip Fung, and Ka Ming Ng. Food product design: a hybrid machine learning and mechanistic modeling approach. *Industrial & Engineering Chemistry Research*, 58(36):16743–16752, 2019.

2. Alex Hawkins-Hooker, Florence Depardieu, Sebastien Baur, Guillaume Couairon, Arthur Chen, and David Bikard. Generating functional protein variants with variational autoencoders. *PLoS Computational Biology*, 17(2):e1008736, 2021.

3. Koki Nakano, Yusuke Noda, Naoto Tanibata, Hayami Takeda, Masanobu Nakayama, Ryo Kobayashi, and Ichiro Takeuchi. Exhaustive and informatics-aided search for fast li-ion conductor with nasicon-type structure using material simulation and bayesian optimization. *APL Materials*, 8(4), 2020.

4. Mortaza Saeidi-Javash, Wenzheng Kuang, Chaochao Dun, and Yanliang Zhang. 3d conformal printing and photonic sintering of high-performance flexible thermoelectric films using 2d nanoplates. *Advanced Functional Materials*, 29(35): 1901930, 2019.

5. Dane Morgan and Ryan Jacobs. Opportunities and challenges for machine learning in materials science. *Annual Review of Materials Research*, 50(1):71–103, 2020.

6. Emily M Williamson and Richard L Brutchey. Using data-driven learning to predict and control the outcomes of inorganic materials synthesis. *Inorganic Chemistry*, 62(40):16251–16262, 2023.

7. Christopher M Bishop and Nasser M Nasrabadi. *Pattern Recognition and Machine Learning*, volume 4. Springer, 2006.

8. Kevin McBride and Kai Sundmacher. Overview of surrogate modeling in chemical process engineering. *Chemie Ingenieur Technik*, 91(3):228–239, 2019.

9. Chun Kit Jeffery Hou and Kamran Behdinan. Dimensionality reduction in surrogate modeling: A review of combined methods. *Data Science and Engineering*, 7(4):402–427, 2022.

10. Christian Weckenborg, Patrick Schumacher, Christian Thies, and Thomas S Spengler. Flexibility in manufacturing system design: A review of recent approaches from operations research. *European Journal of Operational Research*, 315(2):413–441, 2024.

11. Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.

12. Adejuyigbe O Fajemisin, Donato Maragno, and Dick den Hertog. Optimization with constraint learning: A framework and survey. *European Journal of Operational Research*, 314(1):1–14, 2024.

13. Nyoman Gunantara. A review of multi-objective optimization: Methods and its applications. *Cogent Engineering*, 5(1):1502242, 2018.

CHAPTER 2

BAYESIAN OPTIMIZATION FOR CHEMICAL PRODUCTS AND FUNCTIONAL MATERIALS

This chapter is based on previously published work:

Ke Wang and Alexander W Dowling. Bayesian optimization for chemical products and functional materials. *Current Opinion in Chemical Engineering*, 36:100728, 2022.

2.1 Why is materials and product design optimization so challenging?

Chemical-based products and materials with specific functionalities are ubiquitous in modern society and essential for many new technologies.[1, 2] Examples include coatings, fertilizers, food additives[3], medicines, gene sequences[4], and functional devices[5–7]. Yet the design of chemical-based products and functional materials remains challenging. Often the design space is vast; for example, there are approximately 10^{23} drug-like molecules, of which only 10^{8} have been synthesized.[8] Moreover, chemical products and functional materials cannot be designed in isolation; instead, multi-scale and multi-disciplinary design frameworks must account for technical, economic, social, and environmental factors across manufacturing, use, and end-of-life.[1] For example, engineering new functional materials (e.g., membranes) to enable enhanced separations (e.g., water treatment, CO_2 capture) should consider how the new material will be integrated into devices, systems, and infrastructures.[9] In computer-aided molecular design (CAMD), mathematical optimization is used to efficiently search through the vast molecular and materials design spaces to resolve

(some) multiscale trade-offs.[6, 10] Historically, the success of CAMD has been limited by the predictive accuracy of physical property models. Recently machine/deep learning methods have shown great promise for predicting physical properties from data and may revolutionize the computational design of chemical products and functional materials.[11]

2.2 What is Bayesian optimization (BO)?

Bayesian optimization (BO) is a family of surrogate-assisted/derivative-free optimization algorithms that use Bayesian probability theory to explicitly balance trade-offs between exploration and exploitation.[12] BO has two core components: a computationally inexpensive stochastic surrogate model that emulates expensive computational or physical experiments and an acquisition function to determine the optimal sequence of future experiments. [13] BO is typically deployed in a feedback loop with experiments using the following general steps:

1. Identify the objective function(s) and decision variables (with bounds).
2. A space filling design (e.g., Latin hypercube sample) is generated and experiments are performed.
3. The surrogate model is (re)trained using available experimental data.
4. Using the surrogate model, the acquisition function is maximized to recommend the next experiment(s).
5. Experiment(s) are performed and added to the training data.
6. If the goal has not been attained and the experimental budget has not been exhausted, GOTO step 3.

2.2.1 Gaussian process regression surrogate models

Gaussian Processes (GP) are the most popular surrogate models for BO.[14] Below we summarize GP modeling following the notation of [15]; also see [16] for an excellent introduction to GPs.

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i), |\mathbf{x}_i \in \mathbb{R}^p, y_i \in \mathbb{R}, i \in 1, \ldots, n\}$ be a collection of n samples, where the vector $\mathbf{x}_i$ represents input variables (features) correspond to the observation y_i. For convenience, we denote the data $\mathcal{D} = (\mathbf{X}, \mathbf{y})$ using matrix $\mathbf{X} = (\mathbf{x_1}, \ldots, \mathbf{x_n})$ and vector $\mathbf{y} = (y_1, \ldots, y_n)$. We assume y_i includes measurement error that is normal distributed with zero mean and variance σ^2. Mathematically, $y_i = f(\mathbf{x}_i) + \varepsilon$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ where $f(\cdot)$ is an unknown function.

A Gaussian Process can be thought of as a collection of normally distributed random variables that emulates the behavior of $f(\cdot) + \varepsilon$:

$$f \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad \mathbf{x}, \mathbf{x}' \in \mathbb{R}^p \tag{2.1}$$

Here $m(\mathbf{x})$ is the mean function; although often set to zero for computer science applications (e.g., pattern recognition), the mean function is a natural way to incorporate physical relationships into the GP model. $k(\mathbf{x}, \mathbf{x}')$ is the covariance or kernel function. Eq. (2.2) defines three popular kernel functions: Matérn (k_M), squared exponential (k_{SE}), and rational quadratic (k_{RQ}).[12] In Eq. (2.2a), $\nu = 5/2$ and $\nu = 3/2$ are two most popular Matérn kernels and K_v is a modified Bessel function.[17] Often, $\ell \in \mathbb{R}^p$ is the length-scale for the features $\mathbf{x}$. Training hyperparameter ℓ gives insights into the most important dimensions of $\mathbf{x}$ and is known as automatic relevance determination. Alternatively, ℓ is may be treated as a scalar, i.e., the same for each dimension of $\mathbf{x}$, to reduce the number of hyperparameters.

$$k_{\mathrm{M}}(\mathbf{x}, \mathbf{x}') = \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left[2\nu \sum_{j=1}^{p} \left(\frac{x_j - x_j'}{\ell_j} \right)^2 \right]^{\nu/2} K_\nu \left[2\nu \sum_{j=1}^{p} \left(\frac{x_j - x_j'}{\ell_j} \right)^2 \right]^{1/2}, \quad \theta_M = (\ell) \tag{2.2a}$$

$$k_{SE}(\mathbf{x}, \mathbf{x}') = e^{-\frac{1}{2} \sum_{j=1}^{p} \left(\frac{x_j - x_j'}{\ell_j} \right)^2}, \quad \theta_{SE} = (\ell) \tag{2.2b}$$

$$k_{RQ}(\mathbf{x}, \mathbf{x}') = \left[1 + \frac{1}{2\alpha} \sum_{j=1}^{p} \left(\frac{x_j - x_j'}{\ell_j} \right)^2 \right]^{-\alpha}, \quad \theta_{RQ} = (\alpha, \ell) \tag{2.2c}$$

Given training data $(\mathbf{X}, \mathbf{y})$ and values of the hyperparameters θ, one desires to predict $\mathbf{f}_*$ at new input values $\mathbf{X}_*$. The key mathematical property of a GP is that the outputs $\mathbf{y}$ and $\mathbf{f}_*$ follow a multivariate normal (Gaussian) distribution:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathbf{X}) \\ \mathbf{m}(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I} & \mathbf{K}(\mathbf{X},\mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*,\mathbf{X}) & \mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) \end{bmatrix} \right) \tag{2.3}$$

Here $\mathbf{K}(\cdot,\cdot)$ denotes the kernel matrix, which is the kernel function k evaluated elementwise.[15] Exploiting analytical properties of the multivariate normal distribution (see [15, 16] for details) gives the following expected value (mean) and (co)variance for the prediction $\mathbf{f}_*$ corresponding to the new input matrix $\mathbf{X}_*$:

$$\mu_*(\mathbf{X}_*) = \mathbb{E}(\mathbf{f}_*) = \mathbf{m}(\mathbf{X}_*) + \mathbf{K}(\mathbf{X}_*,\mathbf{X})\left[\mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I}\right]^{-1}(\mathbf{y} - \mathbf{m}(\mathbf{X})) \tag{2.4a}$$

$$\mathbf{\Sigma}_*(\mathbf{X}_*) = \mathbb{V}(\mathbf{f}_*) = \mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*,\mathbf{X})\left[\mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I}\right]^{-1}\mathbf{K}(\mathbf{X},\mathbf{X}_*) \tag{2.4b}$$

(For vector input $\mathbf{x}_*$, we denote the posterior prediction mean and variance as scalars $\mu_*(\mathbf{x}_*)$ and $\sigma_*^2(\mathbf{x}_*)$, respectively.) Most importantly, GP models by construction quantify prediction uncertainty via Eq. (2.4b). Furthermore, GPs are non-parametric regression models which means the training data $(\mathbf{X}, \mathbf{y})$ are directly embedded into the model. As consequence, a GP model interpolates between the training data and the kernel function describes the prediction uncertainty as a function of distance (for stationary models) from the training data.

2.2.2 GP training via hyperparameter optimization

The GP hyperparameters θ may be chosen based on modeler's intuition (e.g., σ corresponds to the random error of an experiment). In fact, Eq. (2.4) can be used to make predictions with any valid values of θ (e.g., $\ell > 0$, see [16] for details). Often,

maximum likelihood estimation (MLE) is applied to further train the hyperparameters θ.

$$\log p(\mathbf{y} \mid \mathbf{X}, \theta) = -\frac{1}{2}\mathbf{y}^T \mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{y} - \frac{1}{2}\ln|\mathbf{K}(\mathbf{X}, \mathbf{X})| - \frac{n}{2}\ln 2\pi \tag{2.5}$$

In Eq. (2.5), $p(\mathbf{y} \mid \mathbf{X}, \theta)$ is the GP likelihood function. The log of the likelihood function is nonconvex and often has many local maxima. Multi-start initialization and cross validation are strongly recommended.

2.2.3 Acquisition functions balance exploration and exploitation

The main task for acquisition function (AF) is to find the optimal next experimental conditions $\bar{\mathbf{x}}$ by balancing trade-off between exploration and exploitation.[13] When the AF is biased for exploration, it suggests $\bar{\mathbf{x}}$ near the current best known decision $\mathbf{x}^+$ to maximize $f(\bar{\mathbf{x}})$; conversely, when biased for exploitation, the AF recommends decisions with high prediction uncertainty. Expected Improvement (EI) is the most popular AF for materials and product design BO, and is defined as the expected value of $\max(f(\bar{\mathbf{x}}) - f(\mathbf{x}^+),\ 0)$. When using a GP surrogate model for $f(\cdot)$, EI is analytically calculated as follows[17]:

$$\mathrm{EI}(\bar{\mathbf{x}}) = \begin{cases} \underbrace{\left(\mu_*(\bar{\mathbf{x}}) - f\left(\mathbf{x}^+\right) - \zeta\right)\Phi(z)}_{\text{exploitation}} + \underbrace{\sigma_*(\bar{\mathbf{x}})\phi(z)}_{\text{exploration}}, & \sigma_*(\bar{\mathbf{x}}) > 0 \\ 0, \quad \text{otherwise} \end{cases} \tag{2.6a}$$

$$z(\bar{\mathbf{x}}) = \begin{cases} \frac{\mu_*(\bar{\mathbf{x}}) - f\left(\mathbf{x}^+\right) - \zeta}{\sigma_*(\bar{\mathbf{x}})}, & \sigma_*(\bar{\mathbf{x}}) > 0 \\ 0, \quad \text{otherwise} \end{cases} \tag{2.6b}$$

Here $\Phi(z)$ and $\psi(z)$ are the cumulative and probability density functions for the standard normal distribution, respectively, $f(\mathbf{x}^+)$ is the largest measured value, i.e.,

$\mathbf{x}^+ = \text{argmax}(f(\mathbf{x}_1), ..., f(\mathbf{x}_n))$, and ζ is an adjustable parameters. For large values of ζ, Eq. (2.6) favors exploration by placing less importance on the posterior prediction mean $\mu_*(\bar{\mathbf{x}})$ and thus increasing the relative importance of $\sigma_*(\bar{\mathbf{x}})$. Conversely, smaller values of ζ favor exploitation. Figure 2.1 shows BO applied to a test function using EI with $\zeta = 0$ (no bias towards exploration).

2.2.4 Extension to other surrogate models and acquisition functions

As previously discussed, GP models are non-parametric and include the training data $(\mathbf{X}, \mathbf{y})$. Unfortunately, the matrix inversions (often computed via Cholesky decomposition) in Eq. (2.4) is often too computationally expensive with more than 10,000 observations without deploying specialized modeling and numerical techniques (see Ch. 8 in [15]). Recent work focuses on adapting BO to deep learning models better suited for large datasets.[18, 19] Other AFs including portfolio allocation, entropy-based acquisition function, knowledge gradient, and mean objective cost of uncertainty.[17] The general idea of balancing the trade-off between exploration and exploitation is universal across AFs, although the relative performance is problem specific. Many practitioners start with EI and then explore other AFs as they develop BO frameworks for new applications.

2.3 How can BO accelerate materials and product design?

Traditional materials and product design harmonize empirical data, scientific intuition, and computational methods that describe the behavior of matter at the atomistic and (macro)molecular behavior[20] to iteratively discover materials that enable new products. Although successful, these methods rely on expensive and slow computation and physical experiments. Moreover, it is difficult for human-driven scientific intuition to resolve trade-offs in high-dimensional design space. BO-driven inverse design frameworks systematically guide prevailing Edisonian workflows with

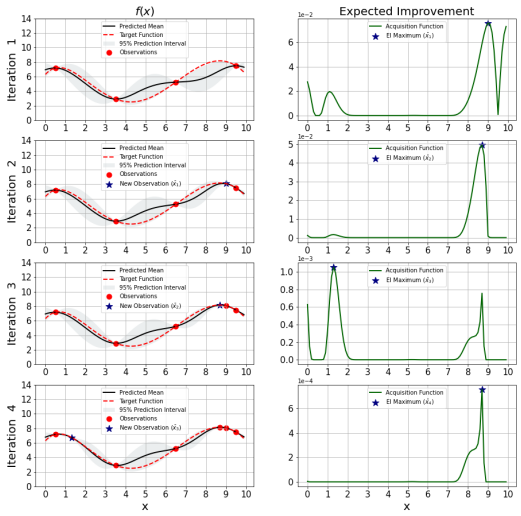

Figure 2.1. **Bayesian optimization illustrative example.** BO with EI is applied to the test problem $f(x) = 0.0015x^5 - 0.055x^4 + 0.65x^3 - 2.8x^2 + 3.0x + 6.3$ using a Squared Exponential kernel (k_{SE}) with initial hyperparameter $\theta_{SE} = 1$. The **left column** shows the GP prediction mean (black line) and 95% prediction intervals (grey regions), the unknown function $f(x)$ (red dashed line), and available training data for each iteration (red dots). The **right column** shows the recommend optimal sample (blue star) which maximizes the EI acquisition function (green line). **Iteration 1**: We randomly select 4 data points within search space $0 \leq x \leq 10$ and train the GP model via MLE. We find $\bar{x}_1 = 9.0$ maximizes EI. **Iteration 2**: We query $f(\bar{x}_1)$ and corrupt the observation with zero-mean normally distributed error with 0.1 standard deviation. We update/retrain the GP model using this new observation and find $\bar{x}_2 = 8.7$ maximizes the EI. At this iteration, EI favors exploitation. **Iteration 3**: After sampling $f(\bar{x}_2)$ and updating the GP model, we find $\bar{x}_3 = 1.3$. At this iteration, EI favors exploration. **Iteration 4**: We sample $f(\bar{x}_3)$, retrain the GP model, and find $\bar{x}_4 = 8.7$, which again favors exploitation. **Summary**: Through this example, we show BO can efficiently sample the unknown function $f(x)$ near its two local maxima at $x = 1.3$ and $x = 8.7$ by balancing exploration and exploitation. By comparing iterations 1 to 4, we see the prediction intervals (grey regions) shrink as more observations are added to the GP surrogate model.

artificial intelligence to accelerate molecular, materials, and product innovations.[21] The previously described 6-step general BO workflow is easily adapted to inverse design of new molecules, materials, or products with one or more tailored physical properties by optimizing physical or computational experiments.

2.3.1 Molecular design and discovery

Artificial Intelligence (AI) offers new paradigms to systematically design and discover (macro)molecules with specific properties and functionality.[11] For example, generative machine learning models including generative adversarial networks (GAN), reinforcement learning (RL), recurrent neural networks (RNN), and variational autoencoders (VAE) are commonly used to propose new molecular structures.[11] For brevity, we focus this review on Variational Autoencoder-based Bayesian Optimization (VAE-BO), which was first proposed by Gómez-Bombarelli et al [22]. VAE models were first proposed by Kingma et al[23] and have two main components. The VAE encoder is an RNN or a similar deep learning model that converts string-like molecular representations (e.g., SMILES strings) into a low to medium-dimensional vector in continuous latent space. Similarly, the VAE decoder converts a vector in the latent space into a string-like molecular structure. BO is then performed in the continuous latent space. Unfortunately, many points in latent space do not map to valid molecular structures. To overcome this challenge, Griffiths et al [24] proposed a constrained BO approach that uses an additional surrogate model to predict the probability at each point in latent space corresponds to a valid molecule.

Table 2.1 highlights diverse applications of VAE and BO for materials design including proteins[4], ligands (binding sites)[25], drug discovery[22, 24], and light-absorbing organic molecules[26].

BO can be integrated with atomistic or (macro)molecular physics-based models to further accelerate materials design. For example, Tamura et al[27] used BO

to estimate spin-spin interactions using magnetization curve synthesized by experimental data, successfully selecting relevant terms of interactions from redundant candidates with high accuracy. Ju et al[28] integrate atomistic Green's function and BO to maximize and minimize interfacial thermal conductance ultimately identifying a non-intuitive optimal structure with 50% improvement. Yan[29] integrate high throughput calculations with BO to discover atomistic SiGe alloy configurations with extremely low thermal conductivity. Sestito et al [30] use multiobjective Bayesian optimization to calibrate molecular dynamics force fields for polycaprolactone.

TABLE 2.1

RECENT APPLICATIONS OF BO IN MOLECULAR MODELING AND DESIGN.

Ref.	**Application**	**Finding**	**Method**
[22]	Drug-like molecular design	Seminal paper demonstrating VAE autoencoder with low reconstruction error rates (i.e., generates few invalid molecules) and proof-of-concept BO in molecular latent space.	VAE-BO
[24]	Drug-like molecular design	Seminal paper demonstrating constrained BO to reduce the number of invalid generated molecules in VAE model.	VAE-CBO
[4]	Protein sequence design	Proposes VAE framework to capture complex protein sequence-function relationships and applies BO to identify mutations that optimize specific protein functions.	VAE-BO
[26]	Organic light-absorbing molecular design	Demonstrates VAE-BO framework trained on simple molecular structures can generate complex molecular structures.	VAE-BO
[25]	Ligand design	VAE-BO identifies sub-regions of the ligand design space with improved binding sites.	VAE-BO
[27]	Magnetic material design	Leverages BO to calibrate spin-spin interactions in molecular models using experimental data.	BO
[28]	SiGe composite design	Use BO to identify a non-intuitive optimal structure for Si-Si and Si-Ge composite which delivers 50% improved thermal conductive compared to state-of-the-art.	BO
[29]	SiGe composite design	BO identifies new layers structures to minimize the thermal conductivity of SiGe alloys.	BO

2.3.2 Smart and additive manufacturing

Under the emerging smart manufacturing paradigm, intelligent systems and robotics are augmenting the synthesis and fabrication of new materials.[31] Additive manufacturing refers to layer-upon-layer fabrication processes controlled by an intelligent system such as computer-aided design (CAD) model [32] and enables the creation of new multi-functional materials that cannot be manufactured via conventional methods.[33] However, optimizing smart and additive manufacturing processes is challenging because these systems are often too complex to describe with physics-based computational models. This is especially true for human-in-the-loop cyber-physical systems. BO enables automated learning and self-optimization to address this challenge.

BO improves smart and additive manufacturing systems, often an order of magnitude fewer experiments (less data) than trial-and-error Edisonian search. [14] For example, Nakano et al[5] optimized Li, Ca, Y, and Zr composition in solid electrolytes to maximize Li-ion conductivity. They show BO with observations outperforms high throughput trail-and-error search (169 observations). Lookman et al [34] applied BO to minimize the thermal dissipation of shape memory alloys by optimizing the composition (Ni, Cu, Pd, Fe, Ti). Despite the vast search space with 800,000 possible alloy compositions, discovers a novel alloy with 42% improvement in only 9 iterations (with 4 alloy samples per iteration). Gongora et al[35] developed the Bayesian experimental autonomous researcher (BEAR) and demonstrate also a 60-fold reduction in the number of experiments to optimize additive manufacturing structures versus grid search.

TABLE 2.2

RECENT APPLICATIONS OF BO IN SMART AND ADDITIVE MANUFACTURING.

Ref.	**Application**	**Finding**	**Decision Variables**
[36]	High performance concrete formulation	Demonstrates inverse design by optimizing 7 variables and shows the prediction agree with physicochemical measurements.	Quantity of cement, water, fly ash, etc.
[5]	Solid electrolyte design	BO only requires 40 experiments to maximize Li-ion conductivity compared to 169 with trail-and-error search.	Composition of Li, Ca, Y, Zr
[37]	Mesoporous alloy design	BO only required 47 experiments to optimize catalytic activity.	Composition of $AcCl_4$, $PtCl_4$, $PdCl_4$
[34]	NiTi alloy design	BO identifies novel alloy by efficiently searches over 800,000 possible compositions using only 36 experiments.	Ratio of Ni, Cu, Pd, Fe
[35]	Additive manufacturing	5 out of 100 mechanical structures designed by BO outperformed 1,800 designs generated via grid search.	Number of hollow columns, outer radius, thickness, twist angle
[38]	3D printing	Optimized four printing variables with only 100 experiments. In contrast, full factorial response surface optimization would have required up to 10,000 experiments.	prime delay, print speed, x-position, y-position
[39]	Multi-material 3D printed composite solids	VAE-BO successfully optimizes macroscopic elastic moduli in lattice structured composite material.	Representative volume element (modeled with VAE)

2.4 What are the current challenges and future opportunities?

2.4.1 Leveraging complex data

Experiments are often expensive and involve multiple data sources with complex uncertainty structures; multi-fidelity BO is an emerging technique to systematically fuse data from multiple physical or computational experiments.[17] The general idea is to use inexpensive but less accurate models for initial exploration and transition to higher-fidelity experiments to refine the search. This is done by leveraging the correlation between low and high-fidelity experiments. For example, Herbol et al[40]

recently proposed a BO framework to fuse data from multiple experiments using the Pearson correlation coefficient. Another challenge is heteroscedastic uncertainty structures. Recall we previously assumed observations each observation is corrupted by $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ measurement error. We assumed this error was homoscedastic, i.e., it has constant (but perhaps unknown) variance σ^2. Griffiths et al[41] discuss the benefits of using heteroscedastic uncertainty for BO.

2.4.2 BO with discrete decisions

Many materials and manufacturing optimization problems involve both discrete and continuous decisions. However, a vast majority of BO frameworks only support continuous decision variables. Maximizing acquisition functions in a nonconvex bounded continuous optimization problem is already challenging due to the presence of local minima. Extending BO to include discrete decisions yields a nonconvex mixed-integer nonlinear optimization problem which is especially difficult to solve numerically. As a workaround, discrete decisions are relaxed to continuous variables, the BO optimization problem is numerically solved, and rounding is applied.[42] Although simple to implement, it is well known that rounding can yield sub-optimal results. Alternatively, Zhang et al[43] proposed a latent-variable Gaussian process (LVGP) to map discrete into numerical latent space in GP. They showed improved prediction accuracy and lower computational times compared to rounding.

2.4.3 Batch experiment optimization

Often data is collected in parallel through high throughput screening (e.g., deposition of films with continuous concentration gradients) or high-performance computing. This requires BO to recommend the next experiments in batches.[44] One approach is to assemble batches from multiple local maxima of the acquisition function.[45] Alternatively, Joy et al[46] train an ensemble of GP models with different kernels

and maximize the acquisition function for each GP model separately to assemble BO batches. Batch BO remains an active research area.

2.4.4 Open-source software platforms

Special care is required when choosing a BO software platform for molecular design, material discovery, or manufacturing optimization. For example, COMBO[47], one of the first open-source BO platforms, has not been actively developed for almost three years. Similarly, pyGPGO[48] was last updated about two years ago. Three packages most actively developed at the time of writing are ChemOS[49], BoTorch[50] and pyOpt[51]. BO is a rapidly evolving methodology with new surrogate models and acquisition functions proposed each year, and unfortunately, no single BO software platform implements every innovation.

We recommend practitioners compare the features, documentation, and tutorials for a handful of BO software platforms before starting each new project.

2.5 Acknowledgements

This work was partially supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Advanced Manufacturing Office Award Number DE-EE0009103.

2.6 References

1. Lei Zhang, Haitao Mao, Qilei Liu, and Rafiqul Gani. Chemical product design–recent advances and perspectives. *Current Opinion in Chemical Engineering*, 27: 22–34, 2020.

2. Faheem Mushtaq, Xiang Zhang, Ka Y Fung, and Ka M Ng. Computational design of structured chemical products. *Frontiers of Chemical Science and Engineering*, pages 1–17, 2021.

3. Xiang Zhang, Teng Zhou, Lei Zhang, Ka Yip Fung, and Ka Ming Ng. Food product design: a hybrid machine learning and mechanistic modeling approach. *Industrial & Engineering Chemistry Research*, 58(36):16743–16752, 2019.

4. Alex Hawkins-Hooker, Florence Depardieu, Sebastien Baur, Guillaume Couairon, Arthur Chen, and David Bikard. Generating functional protein variants with variational autoencoders. *PLoS Computational Biology*, 17(2):e1008736, 2021.

5. Koki Nakano, Yusuke Noda, Naoto Tanibata, Hayami Takeda, Masanobu Nakayama, Ryo Kobayashi, and Ichiro Takeuchi. Exhaustive and informatics-aided search for fast Li-ion conductor with NASICON-type structure using material simulation and Bayesian optimization. *APL Materials*, 8(4):041112, 2020.

6. Christopher L Hanselman, Dominic R Alfonso, Jonathan W Lekse, Christopher Matranga, David C Miller, Chrysanthos E Gounaris, et al. A framework for optimizing oxygen vacancy formation in doped perovskites. *Computers & Chemical Engineering*, 126:168–177, 2019.

7. Zhongyu Cheng, Ke Wang, Ali MN Tanvir, Wenjie Shang, Tengfei Luo, Yanliang Zhang, Alexander W Dowling, and David B Go. Bayesian optimization of low-temperature nonthermal plasma jet sintering of nanoinks. *ACS Applied Materials & Interfaces*, 16(35):46897–46908, 2024.

8. Daniel C Elton, Zois Boukouvalas, Mark D Fuge, and Peter W Chung. Deep learning for molecular design—a review of the state of the art. *Molecular Systems Design & Engineering*, 4(4):828–849, 2019.

9. Elvis A Eugene, William A Phillip, and Alexander W Dowling. Data science-enabled molecular-to-systems engineering for sustainable water treatment. *Current Opinion in Chemical Engineering*, 26:122–130, 2019.

10. Nick D Austin, Nikolaos V Sahinidis, and Daniel W Trahan. Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. *Chemical Engineering Research and Design*, 116:2–26, 2016.

11. Kristof T Schütt, Stefan Chmiela, O Anatole von Lilienfeld, Alexandre Tkatchenko, Koji Tsuda, Klaus-Robert Müller, Schütt, and Scalone. *Machine Learning Meets Quantum Physics*. Springer, 2020.

12. Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.

13. Peter I Frazier. Bayesian optimization. In *Recent Advances in Optimization and Modeling of Contemporary Problems*, pages 255–278. INFORMS, 2018.

14. Turab Lookman, Prasanna V Balachandran, Dezhen Xue, and Ruihao Yuan. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Computational Materials*, 5(1):1–17, 2019.

15. Carl Edward Rasmussen and Christopher I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

16. Robert B Gramacy. *Surrogates: Gaussian Process Modeling, Design, and Optimization for The Applied Sciences*. CRC Press, 2020.

17. Francesco Archetti and Antonio Candelieri. *Bayesian Optimization and Data Science*. Springer, 2019.

18. Keith T Butler, Daniel W Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine learning for molecular and materials science. *Nature*, 559(7715): 547–555, 2018.

19. Tian Wang, Mingqi Shao, Rong Guo, Fei Tao, Gang Zhang, Hichem Snoussi, and Xingling Tang. Surrogate model via artificial intelligence method for accelerating screening materials and performance prediction. *Advanced Functional Materials*, 31(8):2006245, 2021.

20. Edward J Maginn. From discovery to data: What must happen for molecular simulation to become a mainstream chemical engineering tool. *AIChE Journal*, 55(6):1304–1310, 2009.

21. Alex Zunger. Inverse design in search of materials with target functionalities. *Nature Reviews Chemistry*, 2(4):1–16, 2018.

22. Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 4(2):268–276, 2018.

23. Diederik P. Kingma and Max Welling. Auto-encoding variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

24. Ryan-Rhys Griffiths and José Miguel Hernández-Lobato. Constrained Bayesian optimization for automatic chemical design using variational autoencoders. *Chemical Science*, 11(2):577–586, 2020.

25. Vincent Mallet. *Leveraging Binding-site Structure for Drug Discovery with Point-cloud Methods*. McGill University (Canada), 2019.

26. Kyungdoc Kim, Seokho Kang, Jiho Yoo, Youngchun Kwon, Youngmin Nam, Dongseon Lee, Inkoo Kim, Youn-Suk Choi, Yongsik Jung, Sangmo Kim, et al. Deep-learning-based inverse design model for intelligent discovery of organic molecules. *npj Computational Materials*, 4(1):1–7, 2018.

27. Ryo Tamura and Koji Hukushima. Method for estimating spin-spin interactions from magnetization curves. *Physical Review B*, 95(6):064407, 2017.

28. Shenghong Ju, Takuma Shiga, Lei Feng, Zhufeng Hou, Koji Tsuda, and Junichiro Shiomi. Designing nanostructures for phonon transport via Bayesian optimization. *Physical Review X*, 7(2):021024, 2017.

29. Jiahao Yan, Han Wei, Han Xie, Xiaokun Gu, and Hua Bao. Seeking for low thermal conductivity atomic configurations in SiGe alloys with Bayesian optimization. *ES Energy & Environment*, 8(2):56–64, 2020.

30. Jesse M Sestito, Mary L Thatcher, Leshi Shu, Tequila AL Harris, and Yan Wang. Coarse-grained force field calibration based on multiobjective Bayesian optimization to simulate water diffusion in poly-ε-caprolactone. *J. Phys. Chem. A*, 124 (24):5042–5052, 2020.

31. Chao Shang and Fengqi You. Data analytics and machine learning for smart process manufacturing: recent advances and perspectives in the big data era. *Engineering*, 5(6):1010–1016, 2019.

32. Lingbin Meng, Brandon McWilliams, William Jarosinski, Hye-Yeong Park, Yeon-Gil Jung, Jehyun Lee, and Jing Zhang. Machine learning in additive manufacturing: A review. *The Journal of The Minerals, Metals Materials Society*, 72 (6):2363–2377, 2020.

33. Zeqing Jin, Zhizhou Zhang, Kahraman Demir, and Grace X Gu. Machine learning for advanced additive manufacturing. *Matter*, 3(5):1541–1556, 2020.

34. Turab Lookman, Prasanna V Balachandran, Dezhen Xue, John Hogden, and James Theiler. Statistical inference and adaptive design for materials discovery. *Current Opinion in Solid State and Materials Science*, 21(3):121–128, 2017.

35. Aldair E Gongora, Bowen Xu, Wyatt Perry, Chika Okoye, Patrick Riley, Kristofer G Reyes, Elise F Morgan, and Keith A Brown. A Bayesian experimental autonomous researcher for mechanical design. *Science Advances*, 6(15): eaaz1708, 2020.

36. Xinyuan Ke and Yu Duan. A Bayesian machine learning approach for inverse prediction of high-performance concrete ingredients with targeted performance. *Construction and Building Materials*, 270:121424, 2021.

37. Asep Sugih Nugraha, Guillaume Lambard, Jongbeom Na, Md Shahriar A Hossain, Toru Asahi, Watcharop Chaikittisilp, and Yusuke Yamauchi. Mesoporous trimetallic ptpdau alloy films toward enhanced electrocatalytic activity in methanol oxidation: unexpected chemical compositions discovered by Bayesian optimization. *Journal of Materials Chemistry A*, 8(27):13532–13540, 2020.

38. James R Deneault, Jorge Chang, Jay Myung, Daylond Hooper, Andrew Armstrong, Mark Pitt, and Benji Maruyama. Toward autonomous additive manufacturing: Bayesian optimization on a 3D printer. *MRS Bulletin*, pages 1–10, 2021.

39. Tianju Xue, Thomas J Wallin, Yigit Menguc, Sigrid Adriaenssens, and Maurizio Chiaramonte. Machine learning generative models for automatic design of multi-material 3D printed composite solids. *Extreme Mechanics Letters*, 41:100992, 2020.

40. Henry C Herbol, Matthias Poloczek, and Paulette Clancy. Cost-effective materials discovery: Bayesian optimization across multiple information sources. *Materials Horizons*, 7(8):2113–2123, 2020.

41. Ryan-Rhys Griffiths, Alexander A Aldrick, Miguel Garcia-Ortegon, Vidhi Lalchand, et al. Achieving robustness to aleatoric uncertainty with heteroscedastic bayesian optimisation. *Machine Learning: Science and Technology*, 3(1):015004, 2021.

42. Anh Tran, Minh Tran, and Yan Wang. Constrained mixed-integer gaussian mixture Bayesian optimization and its applications in designing fractal and auxetic metamaterials. *Structural and Multidisciplinary Optimization*, 59(6):2131–2154, 2019.

43. Yichi Zhang, Daniel W Apley, and Wei Chen. Bayesian optimization for materials design with mixed quantitative and qualitative variables. *Scientific Reports*, 10 (1):1–13, 2020.

44. Richard Couperthwaite, Abhilash Molkeri, Danial Khatamsaz, Ankit Srivastava, Douglas Allaire, and Raymundo Arròyave. Materials design through batch Bayesian optimization with multisource information fusion. *The Journal of The Minerals, Metals Materials Society*, 72(12):4431–4443, 2020.

45. Javier González, Zhenwen Dai, Philipp Hennig, and Neil Lawrence. Batch Bayesian optimization via local penalization. In *Artificial Intelligence and Statistics*, pages 648–657. PMLR, 2016.

46. Tinu Theckel Joy, Santu Rana, Sunil Gupta, and Svetha Venkatesh. Batch Bayesian optimization using multi-scale search. *Knowledge-Based Systems*, 187: 104818, 2020.

47. Tsuyoshi Ueno, Trevor David Rhone, Zhufeng Hou, Teruyasu Mizoguchi, and Koji Tsuda. COMBO: an efficient Bayesian optimization library for materials science. *Materials Discovery*, 4:18–21, 2016.

48. José Jiménez and Josep Ginebra. pyGPGO: Bayesian optimization for python. *Journal of Open Source Software*, 2(19):431, 2017.

49. Loïc M Roch, Florian Häse, Christoph Kreisbeck, Teresa Tamayo-Mendoza, Lars PE Yunker, Jason E Hein, and Alán Aspuru-Guzik. ChemOS: An orchestration software to democratize autonomous discovery. *PLoS One*, 15(4):e0229862, 2020.

50. Maximilian Balandat, Brian Karrer, Daniel Jiang, Samuel Daulton, Ben Letham, Andrew G Wilson, and Eytan Bakshy. Botorch: A framework for efficient monte-carlo bayesian optimization. *Advances in Neural Information Processing Systems*, 33:21524–21538, 2020.

51. Ruben E Perez, Peter W Jansen, and Joaquim RRA Martins. pyOpt: a python-based object-oriented framework for nonlinear constrained optimization. *Structural and Multidisciplinary Optimization*, 45(1):101–118, 2012.

CHAPTER 3

MACHINE LEARNING-ASSISTED ULTRAFAST FLASH SINTERING OF HIGH-PERFORMANCE AND FLEXIBLE SILVER–SELENIDE THERMOELECTRIC DEVICES.

This chapter is based on previously published work:

Ke Wang, Mortaza Saeidi-Javash, Minxiang Zeng, Zeyu Liu, Yanliang Zhang, Tengfei Luo, and Alexander W Dowling. Gaussian process regression machine learning models for photonic sintering. *In Computer Aided Chemical Engineering*, volume 49, pages 1819–1824. Elsevier, 2022.

Mortaza Saeidi-Javash, Ke Wang, Minxiang Zeng, Tengfei Luo, Alexander Dowling, and Yanliang Zhang. Machine learning-assisted ultrafast flash sintering of high-performance and flexible silver-selenide thermoelectric devices. *Energy & Environmental Science*, 15:5093–5104, 2022.

3.1 Introduction

Flexible Thermoelectric Generators (TEGs) are promising candidates for developing self-powered wearable devices and industrial Internet of Things.[1–6] Flexible TEGs are lightweight, compact, and maintenance-free solid-state energy convertors with no moving parts that directly convert heat into electricity, and they can easily conform to a variety of heat sources with curved surfaces (e.g., body heat). The efficiency of thermoelectric (TE) materials largely depends on the dimensionless figure of merit (zT) defined as:

$$zT = \frac{\sigma \cdot S^2}{\kappa} \cdot T \tag{3.1}$$

where S, σ, κ, and T denote the Seebeck coefficient, electrical conductivity, thermal conductivity, and absolute temperature, respectively.[7–9] Despite the significant progress in thermoelectric materials to date, Bi_2Te_3-based alloys remain as dominant materials for thermoelectric applications near room temperature, and the zT for n-type TE materials still remains below or around unity at room temperature.[10, 11] In addition, the scarcity of tellurium (Te) necessitates the development of new tellurium-free thermoelectric materials for use in widespread industrial and wearable applications. Silver-selenide (Ag_2Se) is a narrow band gap n-type chalcogenide and an ideal candidate for room temperature applications owing to its high power factor (PF) and low intrinsic thermal conductivity.[12–14]

Sintering is an essential step in materials processing to improve transport properties. In printed TE films, sintering transforms TE particles into a dense structure with improved thermoelectric properties. Conventional thermal sintering requires hours of processing time at elevated temperatures which hinders the widespread development of flexible TEGs on organic substrates of low melting point (e.g., polymers, fabrics). In addition, it hampers the high-throughput discovery and energy-efficient manufacturing of high-performance TE materials with optimized compositions. Substantial effort has been devoted to the development of innovative sintering methods, such as microwave-assisted sintering, spark plasma sintering (SPS), chemical sintering, and intense pulsed light (flash) sintering.[15, 16] Among these techniques, flash sintering using intense pulsed light is uniquely advantageous. For example, it is ultrafast, energy-efficient, and can sinter the TE films at elevated temperatures on low melting point substrates without damaging the underlying substrate. Although flash sintering has been used for a variety of conductive materials such as silver, copper, and graphene, it remains relatively underexplored on semiconducting nanomaterials, par-

ticularly TE materials.[17, 18] Sintering of TE nanoparticles constitutes a complex process involving solvent evaporation, decomposition of organic ingredients, formation of inter-particle conduction pathways, and densification, which highlights the imperative role of optimized flash sintering variables on the resulting TE properties.

Previous efforts to discover the optimum flash sintering variables relied on expert-driven Edisonian trial-and-error search which is time and labor-intensive.[19] Enabled by recent advances in machine learning, data-driven approaches such as Bayesian Optimization (BO) have rapidly permeated many fields including TE materials,[20, 21] smart manufacturing,[20, 22] and molecular modeling of chemical products[23, 24]. Novel artificial intelligence (AI) systems enable automated prediction and optimization of materials and additive manufacturing processes.[20–22] Moreover, machine learning algorithms can both help intelligently maximize specific performance metrics as well as aid in revealing the underlying physical mechanisms. Despite the renewed interest and recent success of AI and machine learning, there are often significant barriers in translating these methods into new application domains.

In this work, we integrate, for the first time, flash sintering with a Gaussian process regression (GPR) machine learning model and BO to predict the optimum flash sintering variables for n-type silver-selenide TE films that lead to maximum PF at room temperature. The proposed methodology successfully optimized four sintering variables – voltage, pulse duration, number of pulses, and pulse delay time – resulting in a PF of 2205 μW/mK^2, and a corresponding zT of 1.1 at room temperature (among the highest in the reported flexible TE films) with a sintering time less than 1.0 second after only 32 experiment-machine learning iterations. This methodology could be easily generalized to ultrafast and high-throughput flash sintering of a diverse range of energy and electronic materials, as well as other manufacturing processes in general.

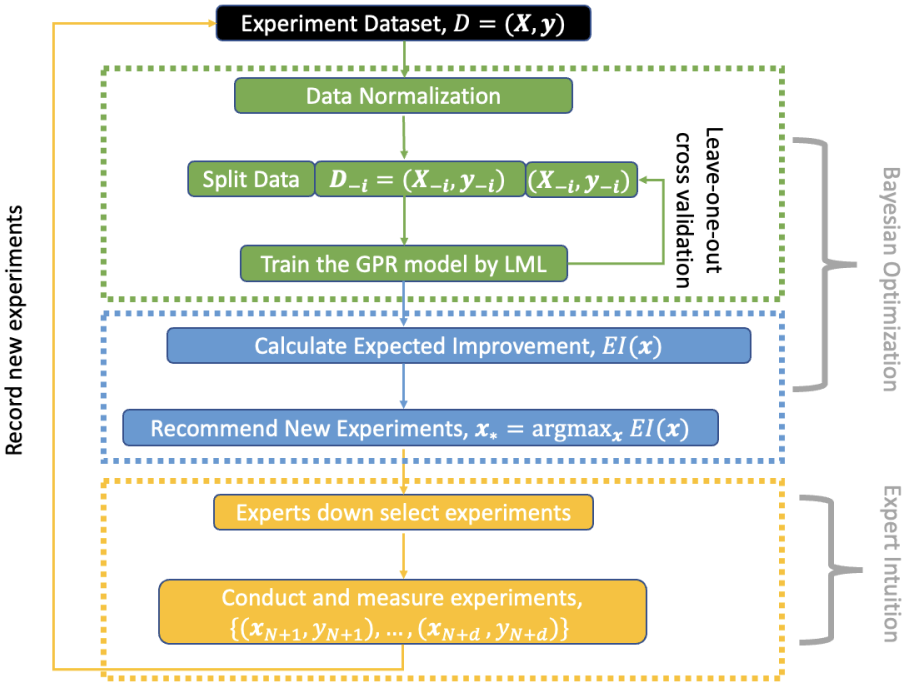

Figure 3.1. The proposed workflow integrates Bayesian optimization (BO) and human intuition. The overall procedure contains three steps: Gaussian process regression (GPR) model training (green box), expected improvement (EI) calculation (blue box), and experimentalist down selection and fabrication (yellow box). The BO is implemented in the first two steps, and expert intuition is incorporated in the last one. The dataset, $D = (\mathbf{X}, \mathbf{y})$, contains N samples of recorded sintering variables $\mathbf{x}_i$, which consist of voltage (x_{i_1}), pulse duration (x_{i_2}), pulse delay (x_{i_3}), number of pulses (x_{i_4}), and thickness (x_{i_5}), and the corresponding power factor y_i. In each iteration, dataset D is provided to BO, and d new experiments, $\{(\mathbf{x}_{N+1}, y_{N+1}), \ldots, (\mathbf{x}_{N+d}, y_{N+d})\}$, are selected informed by human intuition, performed, and added to dataset D; the procedure terminates when the expected improvement approaches zero or the experimental budget is exhausted.

3.2 Method

Figure 3.1 summarizes the Bayesian optimization framework we designed for optimizing thermoelectric (TE) materials via Gaussian process regression (GPR) and

Bayesian optimization. We begin by defining the flash sintering dataset, which contains five decision variables (voltage, pulse duration, number of pulses, pulse delay, and thickness), all standardized alongside the power factor (the objective). A GPR model is then used to capture the relationship between these experimental conditions and their corresponding power factors. The hyperparameters of the GPR model, specifically, the length scales for each feature and the noise variance are tuned by maximizing either the log marginal likelihood or leave-one-out cross-validation criterion. Finally, the expected improvement (EI) acquisition function is used to balance exploration and exploitation when recommending new experiments that are most likely to improve the power factor.

3.2.1 Decision variables, target, and data pre-processing

Flash sintering is a well-known technique for fabricating high-performance thermoelectric material that contains multiple controllable experimental conditions, details in A.1 in SI. Previous studies have indicated that the voltage (x_1), pulse duration (x_2), number of pulses (x_3), pulse delay (x_4), and thickness (x_5) are the most valuable experimental conditions, which we define as decision variables represented as $\mathbf{x}_i$, where i denotes the sample order. The experimentalist starts with 7 experiments to randomly fabricate TE devices and, based on the initial results, performs single-variable control to optimize the experimental conditions to achieve the target (y_i), the power factor.

Let $D = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^5, y_i \in \mathbb{R}, i \in \{1, \ldots, n\}\}$ be a collection of flash sintering datasets, where the vector $\mathbf{x}_i$ represents decision variables corresponding to the experimental result y_i. For convenience, we denote the data $D = (\mathbf{X}, \mathbf{y})$ using matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ and vector $\mathbf{y} = (y_1, \ldots, y_n)^T$. Each $\mathbf{x}_i$ and y_i has different units; directly applying them into GPR will degrade the model performance. The simplest way to address this issue is by standardizing. Here, we use standardization to trans-

form each decision variable and target to have a mean of zero and a unit standard deviation:

$$z = \frac{\mathbf{y} - \mathbb{E}(\mathbf{y})}{\sqrt{\mathrm{Var}(\mathbf{y})}}. \tag{3.2}$$

3.2.2 Gaussian process regression

During the experiment, randomness is unavoidable; it is natural to consider the observed error (ϵ), which we assume is normally distributed with zero mean and variance σ^2, i.e., $\epsilon \sim N(0, \sigma^2)$. We incorporate this into the model as:

$$y_i = f(\mathbf{x}_i) + \epsilon, \tag{3.3}$$

where $f(\cdot)$ is the objective function representing the behavior of different experimental conditions. This can be defined as a Gaussian process (GP):

$$f \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad \mathbf{x}, \mathbf{x}' \in \mathbb{R}^d. \tag{3.4}$$

For computational simplicity, we set $m(\mathbf{x})$ to zero. The kernel function $k(\mathbf{x}, \mathbf{x}')$ determines the behavior of the GP. We use the radial basis function (RBF) kernel for model fitting:

$$k_{\mathrm{RBF}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2}\sum_{j=1}^{d}\frac{(x_j - x_j')^2}{\ell_j^2}\right), \tag{3.5}$$

where ℓ_j is the length-scale for the j-th feature. To simplify, ℓ_j can be set as a scalar for human intuition, assuming equal importance across all features. However, real-world scenarios often require varying importance for features. A smaller ℓ_j indicates higher importance of the corresponding feature. To find the optimal ℓ, log marginal likelihood and cross-validation are used. The tuned results provide insights into the most important features of $\mathbf{x}$.

Given new input values $\mathbf{X}^*$ with corresponding predictions $\mathbf{f}^*$, and training data $(\mathbf{X}, \mathbf{y})$, the outputs $\mathbf{y}$ and $\mathbf{f}^*$ follow a multivariate normal distribution:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{X}^*) \end{bmatrix}, \begin{bmatrix} K(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I} & K(\mathbf{X},\mathbf{X}^*) \\ K(\mathbf{X}^*,\mathbf{X}) & K(\mathbf{X}^*,\mathbf{X}^*) \end{bmatrix} \right), \tag{3.6}$$

where $K(\cdot,\cdot)$ is the kernel function evaluated elementwise. The conjugacy properties of the Gaussian distribution yield the predictive results:

$$\mathbb{E}(\mathbf{f}^*) = m(\mathbf{X}^*) + K(\mathbf{X}^*,\mathbf{X})\left[K(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I}\right]^{-1}(\mathbf{y} - m(\mathbf{X})), \tag{3.7}$$

$$\mathrm{Var}(\mathbf{f}^*) = K(\mathbf{X}^*,\mathbf{X}^*) - K(\mathbf{X}^*,\mathbf{X})\left[K(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I}\right]^{-1}K(\mathbf{X},\mathbf{X}^*). \tag{3.8}$$

3.2.3 Hyperparameter tuning

In this chapter, we consider exploring the performance of log marginal likelihood (LML) and cross-validation (CV) for determining GPR hyperparameters θ in flash sintering applications. LML uses all the training data $D = (\mathbf{X}, \mathbf{y})$ to find the θ that maximizes the function, given by:

$$\log p(\mathbf{y}|\mathbf{X},\theta) = -\frac{1}{2}\mathbf{y}^T[K(\mathbf{X},\mathbf{X}|\theta) + \sigma^2\mathbf{I}]^{-1}\mathbf{y} - \frac{1}{2}\log|K(\mathbf{X},\mathbf{X}|\theta) + \sigma^2\mathbf{I}| - \frac{n}{2}\log 2\pi. \tag{3.9}$$

Cross-validation (CV) reduces the variance of the prediction evaluation. The conjugacy property of GPR largely reduces computational cost; therefore, leave-one-out cross-validation (Loo-CV) is adopted for evaluating the optimal θ that maximizes the objective function L_{CV}:

$$\log p(y_i|\mathbf{X}_{-i}, \mathbf{y}_{-i}, \theta) = -\frac{1}{2}\log \sigma_i^2 - \frac{1}{2\sigma_i^2}(y_i - \mu_i)^2 - \frac{1}{2}\log 2\pi, \tag{3.10}$$

$$L_{\mathrm{CV}}(\mathbf{X}, \mathbf{y}, \theta) = \frac{1}{n}\sum_{i=1}^{n} \log p(y_i|\mathbf{X}_{-i}, \mathbf{y}_{-i}, \theta), \tag{3.11}$$

where $\mathbf{X}_{-i}$ and $\mathbf{y}_{-i}$ denote all data except sample i.

The length scales of the five decision variables are denoted by $\ell_1, \ell_2, \ell_3, \ell_4, \ell_5$, respectively, and the observation error σ is set as a hyperparameter. We are leveraging the existing dataset from prior study[19], see Table A.1 in SI, for an exploration that excludes thickness as a feature; therefore, ℓ_5, is temporarily omitted. Each of these has a search region between $(0, 1)$. The low-dimensional optimization (e.g., 5 variables) makes it computationally acceptable to apply a grid search. For higher feature spaces, advanced optimization methods (e.g., gradient descent) are preferred.

3.2.4 Expected improvement

Expected improvement (EI) is a popular acquisition function to recommend an optimal $\mathbf{x}_*$ in a Bayesian Optimization (BO) framework. EI balances the trade-offs between exploration, i.e., choosing $\mathbf{x}_*$ in regions with high uncertainty, and exploitation, i.e., choosing $\mathbf{x}_*$ in regions that will maximize $f(\mathbf{x}_*)$. EI achieves this balance by computing the expected value of the improvement between $f(\mathbf{x}_*)$ and $f(\mathbf{x}^+)$, where $\mathbf{x}^+$ is the experimental condition in dataset D that has the highest power factor. Thus, mathematically:

$$EI(\mathbf{x}_*) = \mathbb{E}[\max(f(\mathbf{x}_*) - f(\mathbf{x}^+), 0)]. \tag{3.12}$$

By exploiting mathematical properties of the normal distribution, EI($\mathbf{x}$) has the following analytic formula:

$$EI(\mathbf{x}_*) = \begin{cases} (\mu_*(\mathbf{x}_*) - f(\mathbf{x}^+))\Phi(z) + \sigma_*(\mathbf{x}_*)\phi(z), & \text{if } \sigma_*(\mathbf{x}_*) > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{3.13}$$

where

$$z(\mathbf{x}_*) = \begin{cases} \frac{\mu_*(\mathbf{x}_*) - f(\mathbf{x}^+)}{\sigma_*(\mathbf{x}_*)}, & \text{if } \sigma_*(\mathbf{x}_*) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.14}$$

Here, $\Phi(\cdot)$ is the cumulative distribution function, and $\phi(\cdot)$ is the probability density function, respectively, for the standard normal distribution. The GPR and BO workflows were implemented in Scikit-learn. The entire workflow, including the interaction between BO and human experts (experimentalists), is illustrated in Figure 3.1. The entire procedure, including hyperparameter tuning, training, and EI optimization, requires less than 2 minutes on a MacBook with a 2.6 GHz Intel Core i7 CPU.

3.3 Results

3.3.1 Log marginal likelihood (LML) and leave-one-out cross-validation (Loo-CV) identify similar hyperparameter values

Table 3.1 shows that we reached the same hyperparameter tuning results with either LML or Loo-CV. The first two rows give the optimal hyperparameters of ℓ when setting $\sigma = 0.1$ as a fixed parameter. The optimal ℓ in both frameworks are exactly the same, which indicates that under the scenario of fixing observation error, there is not much difference between the two methods for hyperparameter tuning, at least for the flash sintering dataset. Conversely, the last two rows illustrate that when σ is added as a hyperparameter, there is still no difference between the two

TABLE 3.1

LML AND LOO-CV COMPARISON USING DATASET FROM TABLE A.1 IN SI.

Method	ℓ_1	ℓ_2	ℓ_3	ℓ_4	σ
σ fixed: LML	1	0.635	0.322	1	0.1
σ fixed: Loo-CV	1	0.635	0.322	1	0.1
σ tuned: LML	1	0.687	0.322	1	0.2
σ tuned: Loo-CV	1	0.687	0.322	1	0.2

methodologies.

One notable observation is that when adding σ as a tunable hyperparameter, ℓ_2 increases from 0.635 in the σ-fixed model to 0.687, and σ itself increases from the pre-fixed 0.1 to 0.2. This behavior reflects the trade-off between bias and variance, corresponding to the conclusion that relatively complex models (e.g., $\ell_2 = 0.635$) usually obtain lower observation errors (e.g., $\sigma = 0.1$), while simpler models (e.g., $\ell_2 = 0.687$) have higher observation errors (e.g., $\sigma = 0.2$).

3.3.2 Retrospective and sensitivity analysis

Figure 3.2 demonstrates and explains the efficacy of the GPR machine learning model to predict the PF of flash sintered films as a function of four sintering variables (voltage, pulse duration, number of pulses, pulse delay) and the thickness of the silver-selenide films.

The predicted power factors in Figure 3.2(a) are generated iteratively (with constant hyperparameters) using data from the prior experiments. For example, the GPR prediction for experiment 7 uses data from the six prior experiments for train-

ing. Out of the five variables (including four sintering variables, plus the thickness of the film), the film thickness is the most challenging to control during the vacuum filtration process but most influential on GPR PF predictions. Table A.2 in SI further divides the two groups of samples shown in Figure 3.2(a) with average thicknesses of 2.6 μm to 12.5 μm into eight subgroups. Analysis of Figure 3.2(a) and Table A.2 in SI shows the GPR rapidly learns the process-property relationship and only needs one or two experiments in each thickness subgroup to make confident predictions on experiment with close distance (each prediction is a "weighted sum" of prior experiments). For example, experiments 14 to 22 are in the 9 μm thickness subgroup; experiment 14 has high prediction uncertainty of ± 545 μW/mK^2 as there is no prior experimental data for a sample with 9 μm. Experiment 15 has close distance with experiment 14; thus, after incorporating experiment 14 in the GPR model, the prediction uncertainty of experiment 15 reduced to ± 192 μW/mK^2. Experiment 16 is far away in distance compared with both experiment 14 and 15, leading to high prediction uncertainty of ± 508 μW/mK^2; comparing experiment 16 with experiment 14, the prediction uncertainty decreased 6.7% because the prior experiments 14 and 15 enhance the confidence of the GPR model. As more data are added to each thickness subgroup near the optimal sintering conditions, one expects the GPR prediction uncertainty to converge to the experimental measurement uncertainty. In addition, Table A.2 in SI includes data for 8 experiments in which the thin film burned, due to the excessive energy input, and the PF was measured to be zero. These experiments are not shown in Figure 3.2(a) for clarity but were included in the GPR analysis. Inspecting the GPR hyperparameters, $l_1 = 0.625$ (voltage), $l_2 = 0.459$ (pulse duration), $l_3 = 5$ (number of pulses), $l_4 = 2.36$ (pulse delay), and $l_5 = 0.0477$ (thickness), reveals that thickness is the most influential parameter for the flash sintering process (the importance of a feature is inversely proportional to its length scale l) and the thinner films typically have higher PF comparing similar thicker films. This is a re-

markable finding, as thickness is controlled when preparing the thin films via vacuum filtration and was not originally considered as a sintering optimization variable. As aforementioned, the films with reduced thicknesses experience more uniform heating and sintering across the entire thickness, leading to dense microstructure, and greater PF. These GPR results motivated the team to prepare the second group of thinner films shown in Figure 3.2(a), which underscores the synergies between experiments and machine learning modeling. Figures 3.2(b)-(e) show the sensitivity of the expected improvement metric (objective for BO) as a function of thickness and the other four flash sintering variables. Figure A.2 in the SI shows similar heatmaps for the prediction mean, prediction uncertainty, and expected improvement over a wider thickness range (1 to 16 μm). These heatmaps confirm that the GPR model predicts a narrow thickness range, 2.2 μm to 2.6 μm, which maximizes the expected improvement. Moreover, Figures 3.2(b)-(e) and Figure A.2 show voltage and pulse duration are the second and third most influential factors on PF. Similarly, there is a wide range of pulse delay time and number of pulses that give a high expected improvement. This finding is consistent with importance of features indicated by the length scales as well as the Pearson correlation matrix shown in Figure A.3. Furthermore, the gradual improvement in PF in each thickness group emphasizes the importance of optimizing all flash sintering variables. In this application, the film thickness was determined by the vacuum filtration process. The GPR model was then used to optimize the remaining four sintering variables with the thickness held constant.

A key contribution of this work is the integration of BO recommendations and expert intuition to maximize the PF of flash sintered silver-selenide TE films. To illustrate this integration, consider the results from experiments 23 to 30 in the second thickness group (2.6 $\pm$ 0.5 μm). Experiment 23 is chosen by intuition from previous experiments 1 to 22 as it is the first experiment in 2.35 μm thickness sub-

group. The GPR was then updated to incorporate the result from experiment 23 and BO recommend up to five optimal conditions for the next experiment, which were then downselected by the experimental expert. Following this same procedure, the conditions for the next eight experiments (24 to 32) were chosen, resulting in the steady increase in power factor. The maximum PF was achieved at experiment 32 (see details of sintering conditions in Table A.2 in SI) which is the final experiment in the sixth thickness subgroup (2.45 μm). We observed that the PF decreases for all subsequent five sintering experiments which correspond to seventh and eighth thickness subgroups (2.7 and 3.8 μm, respectively). One possibility, suggested by Figure 3.2(b)-3.2(e), is that there is a narrow range of thickness values, approximately 2.3 to 2.6 μm for which the PF is maximized. The final five experiments (and two thickness subgroups) are outside this range.

3.4 Conclusions

In summary, we report the first machine learning-assisted ultrafast flash sintering of flexible silver-selenide TE devices for energy harvesting applications. BO significantly accelerated our findings of a set of intense pulsed light (flash) sintering variables leading to an ultrahigh power factor of 2205 μW/mK^2, and zT of 1.1 at room temperature realized with sintering time less than 1.0 second. Flash sintered films demonstrate outstanding flexibility with 92% retention of the PF after 10^3 bending cycles. The maximum power density of a six-leg TEG is 0.5 and 26.6 mW/cm^2 at ΔT of 10 K and 70 K, respectively. The ultrahigh-performance, low-cost, and highly flexible silver-selenide TE films show great potential for energy harvesting and wearables. Although this study focuses on the optimization of flash sintering for silver-selenide TE devices, this machine learning-assisted experimentation strategy possesses the potential for ultrafast sintering of other TE material systems (e.g., Bi_2Te_3, and Sb_2Te_3) and roll-to-roll manufacturing of a broad range of energy, thermal, and electronic

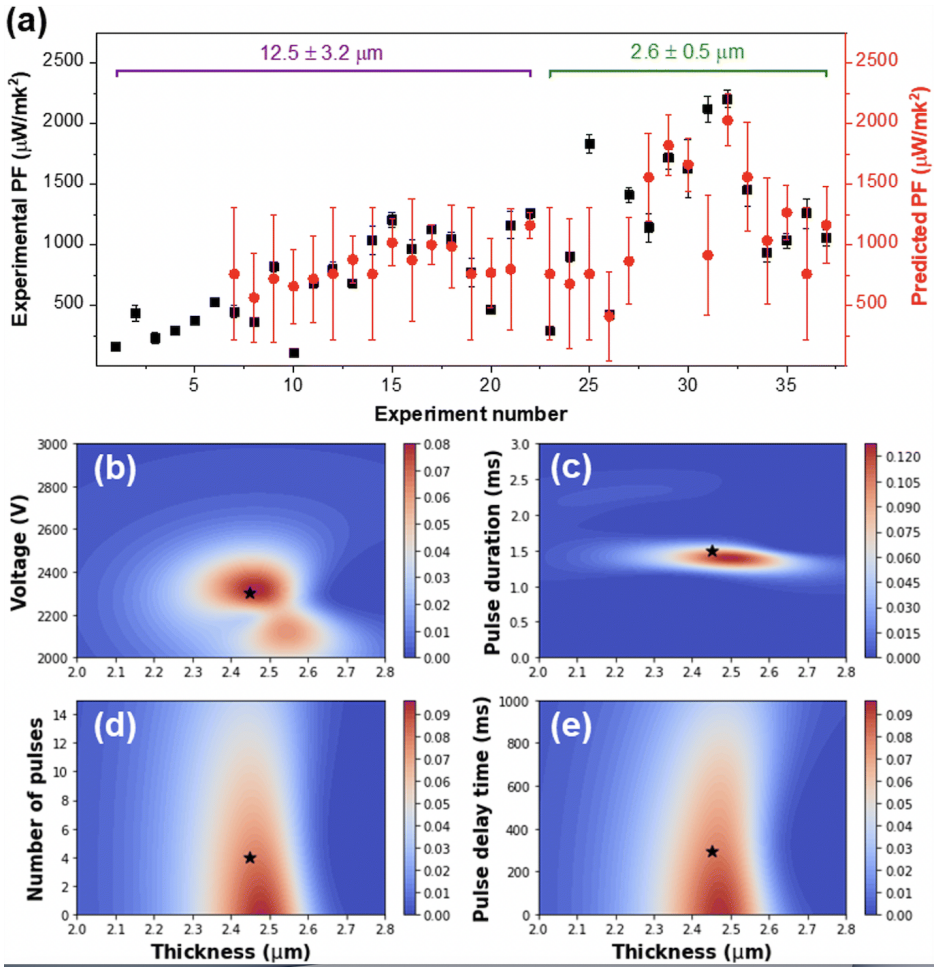

Figure 3.2. (a) Comparison of the measured and machine learning predicted power factors for sintered films. The red dots and error bars correspond to the GPR prediction mean and standard deviation. The black squares and error bars show the measured power factor. (b)-(e) Heatmaps show the sensitivity of the expected improvement (BO objective) as a function of thickness and (b) voltage, (c) pulse duration, (d) number of pulses, and (e) pulse delay time. The color scale from blue to red shows the expected improvement, where the red region indicates the range of optimal sintering variables. The black star marks the conditions of experiment 32 which had the maximum measured power factor.

devices.

Besides, in this work, we successfully developed a GPR model for flash sintering of TE material that can predict the majority of power factors around the experimental results. We demonstrated that the two most popular frameworks are not consequential for identifying the optimal hyperparameter θ, at least for our flash sintering dataset. Specifically, the retrospective analysis supports using GPR as a surrogate model in the Bayesian optimization framework, as GPR precisely predicts the behavior of the objective function $f(\cdot)$. This capability enables combining GPR with acquisition functions to choose the next experimental conditions that balance the trade-off between exploration and exploitation.

As demonstrated by these results, this study highlights the synergies between machine learning-enabled Bayesian optimization and expert-driven Edisonian search. Human intuition is critical to defining the BO problem by identifying the experimental decision variables and their bounds. GPR is especially well-suited for sparse noisy data arising from expensive experiments as GPR "intelligently interpolates" from prior experiments. Early in the experimental campaign, we purposefully explored a mix of BO and human recommended sintering conditions. The latter helped bias the search to consider unexplored regions of the decision space based on prior knowledge. Late in the campaign, we used expert intuition to down select recommended experimental conditions with similar EI scores. We found these strategies to be less cumbersome than designing custom GPR kernels to incorporate said prior knowledge. Moreover, this study demonstrates the robustness and flexibility of the GPR strategy, as we successfully extended the GPR input space to include thickness partway through the experimental campaign. While GPR models do not offer full mechanistic insights, analysis of the kernel length scales provides a relative importance of each input variable. We used this information to design one-dimensionally sensitivity analyses near the optimal sintering conditions and perform the correspond-

ing materials characterization to develop a mechanistic understanding of the results. Moreover, emerging physics-based machine learning models can be incorporated into the proposed framework. In our opinions, these synergies between machine learning and expert intuition are key factors to success.

3.5 Acknowledgements

We would like to acknowledge support from the U.S. Department of Energy under awards DE-EE0009103. Y. Z. would like to acknowledge funding support from the National Science Foundation under award CMMI-1747685 and the U.S. Department of Energy under award DE-NE0008812.

3.6 References

1. Jiasheng Liang, Tuo Wang, Pengfei Qiu, Shiqi Yang, Chen Ming, Hongyi Chen, Qingfeng Song, Kunpeng Zhao, Tian-Ran Wei, Dudi Ren, et al. Flexible thermoelectrics: from silver chalcogenides to full-inorganic devices. *Energy & Environmental Science*, 12(10):2983–2990, 2019.

2. Byeongmoon Lee, Hyeon Cho, Kyung Tae Park, Jin-Sang Kim, Min Park, Heesuk Kim, Yongtaek Hong, and Seungjun Chung. High-performance compliant thermoelectric generators with magnetically self-assembled soft heat conductors for self-powered wearable electronics. *Nature Communications*, 11(1):5948, 2020.

3. Amin Nozariasbmarz, Henry Collins, Kelvin Dsouza, Mobarak Hossain Polash, Mahshid Hosseini, Melissa Hyland, Jie Liu, Abhishek Malhotra, Francisco Matos Ortiz, Farzad Mohaddes, et al. Review of wearable thermoelectric energy harvesting: From body temperature to electronic systems. *Applied Energy*, 258: 114069, 2020.

4. Matthew Dargusch, Wei-Di Liu, and Zhi-Gang Chen. Thermoelectric generators: alternative power supply for wearable electrocardiographic systems. *Advanced Science*, 7(18):2001362, 2020.

5. Saeed Masoumi, Seamus O'Shaughnessy, and Amir Pakdel. Organic-based flexible thermoelectric generators: From materials to devices. *Nano Energy*, 92: 106774, 2022.

6. Xiaoqiang Xu, Yongjia Wu, Lei Zuo, and Shikui Chen. Multimaterial topology optimization of thermoelectric generators. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 59186, page V02AT03A064. American Society of Mechanical Engineers, 2019.

7. G Jeffrey Snyder and Eric S Toberer. Complex thermoelectric materials. *Nature Materials*, 7(2):105–114, 2008.

8. Ming He, Feng Qiu, and Zhiqun Lin. Towards high-performance polymer-based thermoelectric materials. *Energy & Environmental Science*, 6(5):1352–1361, 2013.

9. Yong Du, Shirley Z Shen, Kefeng Cai, and Philip S Casey. Research progress on polymer–inorganic thermoelectric nanocomposite materials. *Progress in Polymer Science*, 37(6):820–841, 2012.

10. H Julian Goldsmid. Bismuth telluride and its alloys as materials for thermoelectric generation. *Materials*, 7(4):2577–2592, 2014.

11. Ian T Witting, Thomas C Chasapis, Francesco Ricci, Matthew Peters, Nicholas A Heinz, Geoffroy Hautier, and G Jeffrey Snyder. The thermoelectric properties of bismuth telluride. *Advanced Electronic Materials*, 5(6):1800904, 2019.

12. Yao Lu, Yang Qiu, Kefeng Cai, Yufei Ding, Mengdi Wang, Cong Jiang, Qin Yao, Changjun Huang, Lidong Chen, and Jiaqing He. Ultrahigh power factor and flexible silver selenide-based composite film for thermoelectric devices. *Energy & Environmental Science*, 13(4):1240–1249, 2020.

13. Cong Jiang, Ping Wei, Yufei Ding, Kefeng Cai, Liang Tong, Qi Gao, Yao Lu, Wenyu Zhao, and Song Chen. Ultrahigh performance polyvinylpyrrolidone/ag2se composite thermoelectric film for flexible energy harvesting. *Nano Energy*, 80: 105488, 2021.

14. Cong Jiang, Yufei Ding, Kefeng Cai, Liang Tong, Yao Lu, Wenyu Zhao, and Ping Wei. Ultrahigh performance of n-type ag2se films for flexible thermoelectric power generators. *ACS Applied Materials & Interfaces*, 12(8):9646–9655, 2020.

15. Morteza Oghbaei and Omid Mirzaee. Microwave versus conventional sintering: A review of fundamentals, advantages and applications. *Journal of Alloys and Compounds*, 494(1-2):175–189, 2010.

16. Jun Xi, Kai Xi, Aditya Sadhanala, Kelvin HL Zhang, Guangru Li, Hua Dong, Ting Lei, Fang Yuan, Chenxin Ran, Bo Jiao, et al. Chemical sintering reduced grain boundary defects for stable planar perovskite solar cells. *Nano Energy*, 56: 741–750, 2019.

17. Ethan B Secor, Theodore Z Gao, Manuel H Dos Santos, Shay G Wallace, Karl W Putz, and Mark C Hersam. Combustion-assisted photonic annealing of printable graphene inks via exothermic binders. *ACS Applied Materials & Interfaces*, 9 (35):29418–29423, 2017.

18. Markus Hösel and Frederik C Krebs. Large-scale roll-to-roll photonic sintering of flexo printed silver nanoparticle electrodes. *Journal of Materials Chemistry*, 22(31):15683–15688, 2012.

19. Mortaza Saeidi-Javash, Wenzheng Kuang, Chaochao Dun, and Yanliang Zhang. 3d conformal printing and photonic sintering of high-performance flexible thermoelectric films using 2d nanoplates. *Advanced Functional Materials*, 29(35): 1901930, 2019.

20. Chengcheng Wang, Xipeng P Tan, Shu Beng Tor, and CS Lim. Machine learning in additive manufacturing: State-of-the-art and perspectives. *Additive Manufacturing*, 36:101538, 2020.

21. Zhufeng Hou, Yoshiki Takagiwa, Yoshikazu Shinohara, Yibin Xu, and Koji Tsuda. Machine-learning-assisted development and theoretical consideration for the al2fe3si3 thermoelectric material. *ACS Applied Materials & Interfaces*, 11 (12):11545–11554, 2019.

22. Chenxi Tian, Tianjiao Li, Jenniffer Bustillos, Shonak Bhattacharya, Talia Turnham, Jingjie Yeo, and Atieh Moridi. Data-driven approaches toward smarter additive manufacturing. *Advanced Intelligent Systems*, 3(12):2100014, 2021.

23. Ryan-Rhys Griffiths and José Miguel Hernández-Lobato. Constrained bayesian optimization for automatic chemical design using variational autoencoders. *Chemical Science*, 11(2):577–586, 2020.

24. Ke Wang and Alexander W Dowling. Bayesian optimization for chemical products and functional materials. *Current Opinion in Chemical Engineering*, 36: 100728, 2022.

CHAPTER 4

BAYESIAN OPTIMIZATION OF LOW-TEMPERATURE NONTHERMAL PLASMA JET SINTERING OF NANOINKS

This chapter is based on previously published work:

Zhongyu Cheng, Ke Wang, Ali MN Tanvir, Wenjie Shang, Tengfei Luo, Yanliang Zhang, Alexan- der W Dowling, and David B Go. Bayesian optimization of low-temperature nonthermal plasma jet sintering of nanoinks. *ACS Applied Materials & Interfaces*, 16(35):46897–46908, 2024.

4.1 Introduction

Flexible devices are widely used as wearables, sensors, and touch screens due to their stretchability, transparency, and biocompatibility.[1–4] Such devices are typically developed using deposited inks containing active nanomaterials (nanoinks). While electronic metals such as gold[5], silver[6], and copper[7] are the most common nanoinks, semiconductor materials increasingly play a more critical role in these devices[8]. Once deposited, the nanoink films must be sintered before being fabricated into functional devices[9]. Conventional sintering methods require high temperatures, high pressures, and long sintering times[10–13], making them unsuitable for delicate and thermally sensitive substrates[10, 14] or for emerging printing methods that can conformally deposit nanoinks on complex surfaces[15]. Sintering methods that operate at low temperatures and atmospheric pressure can be conducted in place, making them attractive alternatives to conventional furnace-based techniques.

Non-thermal plasma-based sintering is an emerging technique that has been demonstrated for both silver (Ag)[10, 16–18] and copper (Cu)[19] nanoink thin films. A non-thermal plasma is a non-equilibrium ionized gas, where free electrons are at high energies (1-10 eV) while ions and neutrons remain cooler (0.01 eV)[20], providing a very reactive environment of ions, electrons, excited neutrals, and photons. Importantly, under proper conditions, the plasma itself can stay at or near room temperature, making it suitable for processing on thermally sensitive surfaces. To that end, there have been a few examples of plasma jet-based sintering on surfaces such as polymers[16–18] and the flesh of fruit[10]. Turan et al.'s work, for example, demonstrated that pulsing an atmospheric-pressure plasma jet on and off with a relatively slow frequency (10^{-2} Hz) maintained the substrate temperature close to room temperature[10].

While plasma jet sintering has been demonstrated for Ag and Cu nanoinks, it is not clear a priori that the same conditions would be optimal for semiconductor nanoinks. Moreover, plasma jet sintering has many degrees of freedom, and the relationships among experimental variables are non-linear, making Edisonian (trial-and-error) or heuristic (rule-of-thumb, one-factor-at-a-time) optimization insufficient. Other researchers have employed machine learning methodologies[21, 22] and related modeling strategies to plasma processing techniques, including applications to plasma medicine[23, 24], plasma etching[25], simulation of plasmas, and plasma deposition[26]. Plasma jet sintering is another plasma processing application that could benefit from these more systematic optimization approaches.

Three main categories of optimization methodologies are commonly applied to manufacturing: experimental empirical investigations[27], physics-informed glass-box optimization[28–30], and data-driven black-box machine learning (ML) optimization[31–33]. Empirical investigations typically rely on one-dimensional sensitivity analyses to experimental conditions, which often cannot result in complex multivariate inter-

actions and can be highly biased towards expert intuition. Glass-box optimization relies on predictive physics-informed models, which are often difficult to postulate, train, and validate, especially for complex systems where the physical phenomena are not fully understood or computationally expensive to model. This motivates data-driven optimization, which relies on ML models to predict the outcome of experiments. Many data-driven black-box models require vast datasets (e.g., deep neural networks) and lack physical interpretability. Gaussian process regression (GPR) [34], physics-informed ML [35, 36], and hybrid models [37] are popular because of their reduced data requirements for optimization [38–40].

This work develops and demonstrates a systematic optimization method leveraging scientific knowledge and ML for non-thermal plasma sintering, using indium tin oxide (ITO) nanoink as a model N-type semiconductor material. Starting with a full factorial experiment design, we identified the constraint envelope for stable plasma jet operation. These constraints are subsequently utilized in a constrained Bayesian Optimization (BO) scheme to maximize the specific energy input (SEI) of the plasma jet. Finally, single- and multi-objective optimization schemes maximized the sintered electrical conductivity while maintaining the substrate temperature below 50 C by manipulating seven sintering decision variables. Ultimately, these results establish the utility of BO to emerging plasma processing techniques for new technology areas.

4.2 Experimental methods

4.2.1 Preparation of ITO thin films

We used a commercial nanoink consisting of an ITO nanoparticle water dispersion (US Research Nanomaterials Inc, ITO, In_2O_3:SnO_2 = 90:10, 20–70 nm, 20wt%) as the raw material for fabricating the ITO thin films. The ITO nanoink was deposited onto glass substrates (25 × 25 mm^2, 1 mm thick) by blade coating with a micrometer-

adjustable film applicator (MTI, EQ-Se-KTQ-50) and using a PDMS mask (20 μm thickness). Before applying blade coating, the glass substrates were cleaned with acetone (MilliporeSigma, acetone ACS reagent, $\geq$ 99.5%) and dried by air jet. To ensure the films' thicknesses were uniform, we also used an injector to deposit 0.01 mL of the liquid ITO dispersion onto the substrate for each sample. The dimensions of the ITO films were 3 $\times$ 3 mm^2. After being coated, the ITO nanoink was in a liquid state and dried at 60 °C for 1 min in open air using a hot plate (VWR, 97042-634 STD) prior to sintering. Note that this process did not influence the resistances of ITO samples prior to sintering, and they all had what we term non-measurable electrical conductivity, with magnitudes much greater than MΩ.

4.2.2 Non-Thermal plasma jet sintering apparatus

A similar plasma jet sintering apparatus as developed in prior work[10] was used here, as shown in Figure 4.1. A cylindrical volume dielectric barrier discharge (DBD) plasma jet configuration was employed with a quartz tube with an inner diameter of 4.7 mm and an outer diameter of 7.0 mm. An external concentric copper tape was used as the ground electrode, and a 1.2 mm diameter stainless-steel rod secured axially down the center of the quartz tube was used as the high-powered electrode. Argon gas (Ar, 99.999%) was used as the working plasma jet gas and controlled by a mass flow controller (Omega FMA5500A). The plasma jet was expelled into a 4 $\times$ 4 $\times$ 4 cm^3 transparent acrylic chamber, which served as a controlled environment to avoid the influence of air during the sintering process. Prior to each experiment, the chamber was flushed with Ar for five minutes at a flow rate of 800 sccm. The ITO sample was affixed to the opposite side of the acrylic chamber, and the distance from the outlet of the quartz tube to the substrate was variable. To monitor the temperature of the substrate, an infrared (IR) camera (FLIR T420) was directed at the back of the substrate[10], which was covered with 0.2 mm thick black tape with

an emissivity of 0.92 and recorded the substrate temperature as a function of the experiment running time.

The DBD plasma jet was operated using a sinusoidal waveform over frequencies ranging from 35 to 70 kHz. The applied high voltage was regulated through a transformer (CMI-5533) and a power amplifier (Powertron, model 50A, RF Amplifier), with the input frequency adjusted via a function generator (Agilent 33220A). The applied voltage was measured using a high-voltage probe (Tektronix P6015A) connected to a digital oscilloscope (Tektronix TBS 2000). As in previous work[10], the plasma jet was pulsed on and off to control the substrate temperature (Figure 4.1b), with the pulsing automated by a MATLAB program connected to the function generator.

As a benchmark, ITO samples were also sintered in a tube furnace (Thermcraft, XST-2-0-12-1V1-F04) at 300 °C with air as the gas atmosphere. We tested furnace sintering at temperatures of 47 °C, 100 °C, 200 °C, 300 °C, and 400 °C, with sintering durations of 18 min (same as the active sintering time of our best plasma jet sintering result), 1 h (same as the total experimental time of our best plasma jet sintering result), 2 h, 3 h, and 4 h. It took 30 seconds to 2 minutes for the furnace's temperature to heat up from room temperature to the specified sintering temperature. When the furnace sintering process was done, we opened the furnace and immediately took the ITO samples out of the furnace. Then, we let the furnace cool down naturally to room temperature before conducting a new furnace sintering test. We found that at 47 °C, the sample electrical conductivity was non-measurable, no matter how long the sintering duration was. At 100 °C, the sample electrical conductivity showed no appreciable increase unless the sintering duration was longer than three hours. At 400 °C, the ITO samples were damaged, especially after sintering for over three hours. We found that 300 °C for a sintering duration of three hours resulted in the highest electrical conductivity without damaging the samples; we used this as our

furnace sintering benchmark. ITO samples were fabricated using the same method and dimensions and underwent the same pre-sintering drying process. We sintered three ITO samples under the same sintering condition each time. They were set in a quartz boat (MTI, H-EQ-QB-1042) and put into the tube furnace for furnace sintering.

Sintered ITO samples were evaluated by measuring the electrical conductivity, σ, based on their resistance and thicknesses, which can be expressed as:

$$\sigma = \frac{1}{\rho} = \frac{a}{R \cdot A}, \tag{4.1}$$

where ρ ($\Omega \cdot m$) is the resistivity of the sample, a (mm) is the length of the sample's edge in the direction of electrical current, A is the cross-sectional area through which the electrical current flows, and R (Ω) is the resistance of the sample, measured using a 2-probe resistance measurement with an AMPROBE digital multimeter (CAT III). The cross-sectional area is based on the thin film thickness l (μm), measured with an optical profilometer (Filmetrics Profilm3D).

Typically, for thin square films, the 4-probe Van der Pauw method is used to measure the resistance of the thin film. In this work, we utilized a 2-probe approach, described in detail in Section B.1 of the Supporting Information (SI), which has been used before for aluminum-doped zinc oxide thin films[41] and ITO thin films fabricated by electron beam evaporation[42]. While 2-probe resistance measurements enable rapid measurements to accelerate training data for ML and BO, a major drawback of the technique relative to 4-probe measurements is reduced accuracy caused by contact resistance. To minimize contact resistance, we used conductive silver paint (Spi, 05001-AB) to create contact pads at the four corners of the ITO sample after the sintering process. The untreated blade-coated ITO samples had resistances in the MΩ range, and after plasma jet sintering, they were in the kΩ range, such that the impact of contact resistance was deemed relatively small. This trade-off was deemed

acceptable for this study, but the limitations of the approach were acknowledged. Therefore, we only compare conductivity values relative to others within our study rather than comparing absolute values in the literature. The surface morphology of the samples was characterized by a scanning electron microscope (Thermo Scientific Helios G4 UX dual beam microscope).

4.2.3 Plasma jet sintering variables

As Figure 4.1 shows, seven decision variables most significantly affect plasma jet sintering performance: the plasma jet flow rate (Q, sccm), the applied peak-to-peak voltage (U_a, kV), the frequency (f, kHz), the gap distance between the outlet of the dielectric tube and the sample surface (d, mm), the pulse-on time (t_{on}, s), the pulse-off time (t_{off}, s), and the number of cycles of plasma jet operation (n). Once in operation, the transient plasma current and voltage can be used to determine the plasma power P via:

$$P = f \cdot \int_0^f U_a(t) \cdot I(t)\, dt, \tag{4.2}$$

where $U_a(t)$ and $I(t)$ are the applied peak-to-peak voltage and measured plasma current, respectively.

Three key output observables were monitored to determine the sintering performance: the electrical conductivity (σ, S m^{-1}), the peak substrate temperature measured during the sintering process (T_{max}, °C), and the total experimental time (t_{total}, s), which is the product of the number of cycles and pulse period:

$$t_{\text{total}} = n \cdot (t_{\text{on}} + t_{\text{off}}). \tag{4.3}$$

The specific energy input (SEI, eV atom^{-1}) is often used to quantitatively characterize a plasma process [43], and can be calculated by the ratio of the plasma power

to the flow rate[43]:

$$\text{SEI} = \frac{P}{Q}. \tag{4.4}$$

Physically, the SEI describes the energy deposited by the plasma per particle in the plasma, and prior work has shown that plasma jet sintering performance follows an Arrhenius-like relationship with the SEI. Three decision variables, Q, U_a, and f, determine the SEI[44], and therefore, it is reasonable to start by optimizing these three variables to maximize SEI. These variables can then be set to fixed values when further optimizing the other four variables in the plasma jet sintering process. We hypothesize this heuristic decomposition makes our entire optimization process efficient, with fewer experiments required overall.

4.2.4 Uncertainty analyses for the repeatability of plasma jet sintering experiments

As discussed in the ensuing sections, the optimization process occurred in three phases. When determining the operating envelope for the onset voltage (U_{on}) and maximum voltage (U_{max}) as a function of jet flow rate (Q) and input frequency (f), the systematic uncertainty in U_{on} and U_{max} of each experimental condition was based on the resolution of the Tektronix TBS 2000 oscilloscope, which is 3%. For the optimization of SEI, the systematic uncertainties in current and voltage were again based on the resolution of the Tektronix TBS 2000 oscilloscope, and the systematic uncertainty in jet flow rate was based on the accuracy of the Omega FMA5500A mass flow controller, which is 1%. Precision uncertainty was quantified at 95% confidence for $N = 3$ experiments under each condition. The combined uncertainty of SEI was $\leq 4\%$ for each condition.

Initial plasma jet sintering experiments consisted of nine different sintering conditions (see Table B.1 in the SI). Each sintering condition was repeated in triplicate to

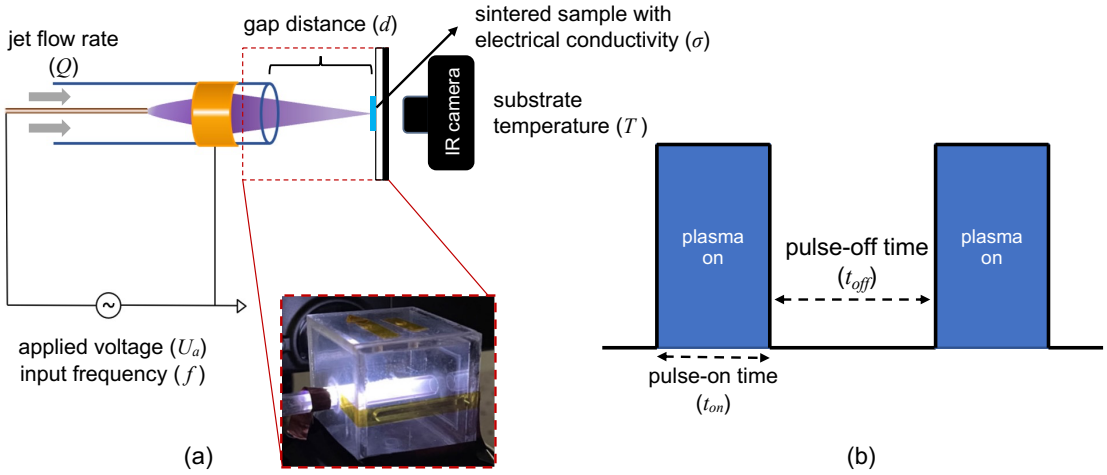

Figure 4.1. Schematic illustration of experimental apparatus. (a) Under the given applied voltage (U_a), input frequency (f) to a stainless-steel powered electrode (brown rod) and a copper grounded electrode, and fixed jet flow rate (Q), the DBD plasma jet is generated and propagates through the dielectric tube, impinging on the sample surface with gap distance (d). The substrate temperature is measured during the sintering process with an IR camera. The inset is a photo of the reaction chamber with a plasma jet impinging on the sample's surface. (b) The on/off operating mechanism with pulse-on time (t_{on}) and pulse-off time (t_{off}) for each cycle determines the peak substrate temperature ($T_{\max}$). The active sintering time t_{active} = number of cycles(n) × pulse-on time(t_{on}) for each experiment affects the film's electrical conductivity (σ).

quantify experimental variability. According to Eq. (4.1), the electrical conductivity of a sintered ITO sample depends on the film's resistance and thickness. Therefore, the error primarily came from the measured 2-probe resistance value and the measured film thickness, considering both precision and systematic contributions using standard error propagation approaches[45]. Precision uncertainty was quantified at 95% confidence for $N = 3$ experiments. The uncertainty range across all nine sintering conditions was 4.55% to 9.36% (see Table B.1 in the SI), revealing the overall repeatability of the process under a wide variety of conditions. The maximum

uncertainty of 9.36% was assumed for subsequent rounds of plasma jet sintering experiments, where only one experiment was conducted for each set of conditions. The uncertainty in peak substrate temperature ($T_{\max}$) measurements was based on the 2% accuracy of the FLIR T420 infrared camera.

4.3 Machine learning and optimization methods

4.3.1 Optimization workflow

Delicate and thermally sensitive substrates are more susceptible to damage as the temperature increases during sintering. On the other hand, a higher specific energy input (SEI) indicates more plasma power deposited per gas particle in a plasma jet. Consequently, when a high-SEI plasma jet interacts with the surface of a sample, it results in an increased transfer of energy from the plasma jet to the sample, potentially leading to an increase in the substrate temperature. Therefore, it is necessary to effectively manage the balance between achieving optimal electrical conductivity through plasma jet sintering and maintaining the lowest maximum substrate temperature.

Our overall goal was to maximize the electrical conductivity of sintered ITO samples while controlling the substrate temperature below a threshold of 50 °C. The seven decision variables shown in Fig. 4.1 led to a challenging high-dimensional optimization problem with complex physical interactions between decision variables that are hard to disentangle.

We leveraged our experience with the plasma jet sintering process to decompose the optimization problem into three phases to reduce the total number of required experiments. In the proposed workflow (Fig. 4.2), we gradually increased the complexity of the optimization problem in each phase from two to three and ultimately to seven decision variables. Phase 1 started with a full factorial design to explore the

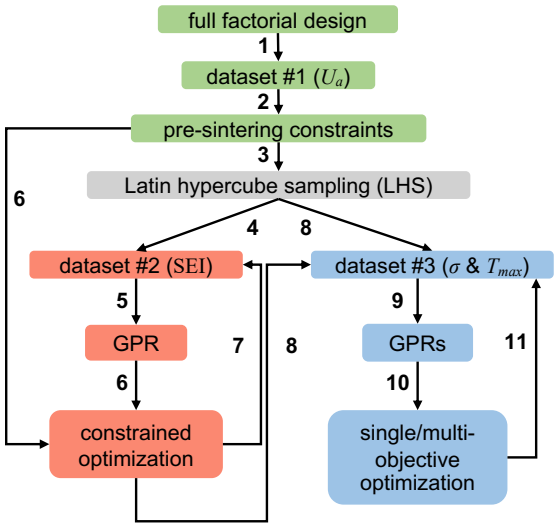

Figure 4.2. Optimization, machine learning, and data science tools are integrated into an iterative workflow to identify constraints and improve objectives. **Phase 1 (green)**: The workflow started with a full factorial design to collect dataset #1 (arrow 1) to identify how the onset voltage and maximum voltage constraints depend on jet flow rate and input frequency (arrow 2). **Phase 2 (red)**: Then, the constraints were considered in Latin hypercube sampling (arrow 3) to form the initial nine ITO samples and provide dataset #2 (arrow 4). The GPR model was trained (arrow 5) and combined with pre-sintering constraints (arrow 6) to conduct an iterative optimization procedure (arrow 7). **Phase 3 (blue)**: After identifying the variable values that produce the highest SEI value (arrow 8), dataset #3 was used for two GPR models (arrow 9) to maximize the electrical conductivity and minimize peak substrate temperature (arrow 10) in an iterative optimization procedure (arrow 11). The workflow terminates after successfully quantifying the Pareto optimal trade-offs between electrical conductivity and peak substrate temperature, or the experimental budget is exhausted.

operating limitations of our plasma jet system and identify the constraint envelope, considering the interactions between jet flow rate (Q) and input frequency (f) on the onset (U_{on}) and maximum (U_{max}) voltages. Phase 2 applied the envelope constraint to a constrained Bayesian Optimization (CBO) procedure to identify the plasma jet flow rate (Q), applied voltage (U_a), and frequency (f), achieving the highest SEI

value. Phase 3 consisted of two steps. In Phase 3a, we fixed Q, U_a, and f to the values that provided us with the highest SEI in Phase 2 and optimized the remaining four decision variables, gap distance (d), pulse-on time (t_{on}), pulse-off time (t_{off}), and number of cycles (n), to maximize the electrical conductivity (σ). In Phase 3b, we optimized all seven variables, considering both objectives: maximizing electrical conductivity and minimizing peak substrate temperature. Figure B.2 in the SI visualizes the model structure, including all the decision variables in our optimization process, which is consistent with the workflow shown in Fig. 4.2.

Phase 1: Establish Plasma Jet Operating Envelope. To maintain a stable plasma jet, the applied voltage (U_a) must be between the onset voltage (U_{on}), the minimum voltage required to induce gas breakdown, and the maximum voltage (U_{max}), which is the upper limit beyond which the plasma jet transitions into attached streamers to the substrate. U_{on} and U_{max} depend on Q and f via mechanisms that are challenging to model with first principles. Instead, we trained two data-driven Gaussian Process Regression (GPR) models. We selected seven uniformly spaced levels for Q and f and applied a full factorial design[46]. For each experiment (Q and f value), it took one minute to measure U_{on} and U_{max}. We denoted this dataset as $D_1 = \{(\mathbf{x}_i^{(1)}, \mathbf{y}_i^{(1)}) \,|\, \mathbf{x}_i^{(1)} \in \mathbb{R}^2, \mathbf{y}_i^{(1)} \in \mathbb{R}^2\}$, where $\mathbf{x}_i^{(1)} = (Q, f)$ and $\mathbf{y}_i^{(1)} = (U_{\text{on}}, U_{\text{max}})$, which are summarized in Table B.2 of the SI. For simplicity, we denoted $D_1 = (\mathbf{X}^{(1)}, \mathbf{y}^{(1)})$ using matrix $\mathbf{X}^{(1)} = (\mathbf{x}_1, \ldots, x_n)$ and vector $\mathbf{Y}^{(1)} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$.

Phase 2: Maximize Specific Energy Input. Turan et al.[43] proposed an Arrhenius-like dependence between the electrical conductivity of silver nanoink thin films and the SEI, which can be expressed as:

$$\sigma \propto \exp\left(-\frac{1}{\text{SEI}}\right). \tag{4.5}$$

Thus, maximizing SEI is a good proxy for maximizing σ. The SEI value is primarily determined by three decision variables: Q, U_a, and f[43]. We began with

nine experiments identified through Latin hypercube sampling (LHS) to populate the search space with starting values for all seven decision variables and initial sintered electrical conductivities. Then, we utilized the Q, U_a, and f information from this initial LHS dataset to start the SEI optimization process using (constrained) expected improvement, where each SEI experiment only took five minutes. We denoted the dataset as $D_2 = \{(\mathbf{x}_i^{(2)}, y_i^{(2)}) \,|\, \mathbf{x}_i^{(2)} \in \mathbb{R}^3, y_i^{(2)} \in \mathbb{R}\}$, where $\mathbf{x}_i^{(2)} = (\mathbf{x}_i^{(1)}, U_a)$ and $y_i^{(2)} = \text{SEI}$. The corresponding dataset is summarized in Table B.3 in the SI and denoted $D_2 = (\mathbf{X}^{(2)}, \mathbf{y}^{(2)})$.

Phase 3a: Maximize Electrical Conductivity. Direct optimization of both σ and peak substrate temperature ($T_{\max}$) is a complicated task. The absence of prior studies on non-thermal plasma jet sintering's effectiveness for ITO thin films presented a considerable risk to the simultaneous optimization of all decision variables. Thus, we first maximized σ without considering $T_{\max}$. Using the findings from Phase 2, we fixed Q, U_a, and f, $x_*^{(2)}$, to the values obtained in Phase 2, which maximize SEI. This reduces the dimensionality of Phase 3a to only four decision variables (d, t_{on}, t_{off}, n). We started with the nine LHS experiments from Phase 2. The dataset used in Phase 3a is denoted as $D_3 = \{(\mathbf{x}_i^{(3)}, \mathbf{y}_i^{(3)}) \,|\, \mathbf{x}_i^{(3)} \in \mathbb{R}^7, \mathbf{y}_i^{(3)} \in \mathbb{R}^2\}$, where $\mathbf{x}_i^{(3)} = (\mathbf{x}_*^{(2)}, d, t_{\text{on}}, t_{\text{off}}, n)$ and $\mathbf{y}_i^{(3)} = (\sigma, T_{\max})$. Thus for all Phase 3a experiments, while we only maximized σ, we also recorded the peak substrate temperature $T_{\max}$. The corresponding dataset is summarized in Table B.4 in the SI and denoted $D_3 = (\mathbf{X}^{(3)}, \mathbf{y}^{(3)})$.

Phase 3b Maximize Electrical Conductivity and Minimize Peak Substrate Temperature. After three successful rounds of maximizing σ (Phase 3a, single objective), we continued two rounds of bi-objective optimization considering all seven decision variables, where $\mathbf{x}_i^{(3)} = (\mathbf{x}_i^{(2)}, d, t_{\text{on}}, t_{\text{off}}, n)$. Phase 3b established the Pareto trade-offs between electrical conductivity and peak substrate temperature constrained by the total experimental time. The corresponding results are summarized in Table B.4.

4.3.2 Gaussian process regression

A Gaussian Process Regression (GPR) model is a non-parametric model that learns from data directly[47]. We assumed y_i includes measurement error that is independent and identically normally distributed with zero mean and variance σ; mathematically, $_i = f(\mathbf{x}_i) + \epsilon$ and $\epsilon \sim \mathcal{N}(0, \sigma^2)$, where $f : \mathcal{X} \to \mathbb{R}$ is an unknown function. Assuming f follows a Gaussian process, i.e., $f \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}))$, where the mean function $m : \mathcal{X} \to \mathbb{R}$ is the expected value of the output variable given the input variables, and the kernel (covariance) function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ measures the degree of similarity between different output values given different input feature values. To predict the output variable of a new experiment, $f(\mathbf{x}_*)$, at conditions $\mathbf{x}_*$, given dataset $\mathcal{D}$, we have:

$$\begin{bmatrix} y \\ f \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{x}_*) \end{bmatrix}, \begin{bmatrix} K(\mathbf{X}) + \sigma^2 I & k(\mathbf{X}, \mathbf{x}_*) \\ k(\mathbf{x}_*, \mathbf{X}) & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix} \right). \tag{4.6}$$

Applying Bayes' rule of probability[48], $f(\mathbf{x}_*) \sim \mathcal{N}(m(\mathbf{x}_*), k(\mathbf{x}_*, \mathbf{x}_*))$, the analytical formula for the prediction mean $\mu_*(\mathbf{x}_*)$ and prediction variance (uncertainty) $\sigma_*^2(\mathbf{x}_*)$ are given by:

$$\mu_*(\mathbf{x}_*) = m(\mathbf{x}_*) + \mathbf{k}^\top(\mathbf{x}_*, \mathbf{X})[\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I}]^{-1}(\mathbf{y} - m(\mathbf{X})), \tag{4.7}$$

$$\sigma_*^2(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}^\top(\mathbf{x}_*, \mathbf{X})[\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I}]^{-1}\mathbf{k}(\mathbf{X}, \mathbf{x}_*). \tag{4.8}$$

The kernel function $k(\cdot, \cdot)$ with hyperparameters h describes the similarity between input variables. Here, we applied the commonly used radial basis kernel function, $k(x, x') = \exp\left(-\frac{|x-x'|^2}{2 \cdot h^2}\right)$, where p is the degree of freedom for the input variables. The hyperparameters h, also called length scales, determine the relevance of the input variables; if the length scale has a relatively large value, the kernel

(covariance) function $k(x, x')$ will depend less on the input variables. Researchers often remove irrelevant input variables[49] using automatic relevance determination methods[50]. The optimal length scales h^* are determined by maximizing the log marginal likelihood (LML):

$$\log P(\mathbf{y} \mid \mathbf{X}, h) = -\frac{1}{2}\mathbf{y}^\top [K(\mathbf{X}, \mathbf{X} \mid h) + \sigma^2 \mathbf{I}]^{-1}\mathbf{y} - \frac{1}{2}\log \left| K(\mathbf{X}, \mathbf{X} \mid h) + \sigma^2 \mathbf{I} \right| - \frac{n}{2}\log 2\pi. \tag{4.9}$$

The typical approach for hyperparameter h^* optimization is the multi-start strategy with the quasi-Newton method[51], and the mean function is set to zero, i.e., $m : \mathcal{X} \to \{0\}$. Here, we developed the GPR model using scikit-learn (version 1.0.2)[52], and optimized the hyperparameters using LBFGS[53]. The entire procedure, including model building and hyperparameter training, required less than 3 minutes. After the model training, the performance of the GPR model was measured by leave-one-out prediction. All calculations were performed on a MacBook with a 2.6 GHz Intel Core i7-6700HQ CPU and 8 GB of RAM.

4.3.3 Constrained Bayesian optimization

Bayesian Optimization (BO) has been used in many emerging applications in materials science[54, 55], advanced manufacturing[56], and molecular modeling of chemical products[57, 58], among others, during the past five years. BO is a family of surrogate-assisted, derivative-free optimization algorithms that use Bayesian probability theory to explicitly balance trade-offs between exploration and exploitation[59]. BO has two core components: a computationally inexpensive stochastic surrogate model that emulates expensive computational or physical experiments and an acquisition function to determine the optimal sequence of future experiments. BO is typically deployed in a feedback loop with rounds following the general steps, as de-

scribed in Fig. 4.2. Standard BO, also called efficient global optimization (EGO)[60], uses the GPR as the surrogate model and expected improvement (EI):

$$\mathrm{EI}(\mathbf{x}) = \mathbb{E}[\max(0, f(\mathbf{x}) - f(\mathbf{x}^+))], \tag{4.10}$$

as the acquisition function, where $f(\mathbf{x}^+)$ is the current best objective value in dataset $\mathcal{D}$ and $f(\mathbf{x})$ is the prediction from the GPR. The next recommended experiment is $\mathbf{x}_* = \arg\max \mathrm{EI}(\mathbf{x})$. EI automatically balances the trade-offs between exploration, i.e., choosing $\mathbf{x}_*$ in regions with high uncertainty, and exploitation, i.e., choosing $\mathbf{x}_*$ in regions that will maximize $f(\mathbf{x}_*)$. By exploiting the mathematical properties of the normal distribution, $\mathrm{EI}(\mathbf{x}_*)$ has the following analytic formula[61]:

$$\mathrm{EI}(\mathbf{x}_*) = \begin{cases} (\mu_*(\mathbf{x}_*) - f(x^+))\Phi(z) + \sigma_*(\mathbf{x}_*)\phi(z), & \sigma_*(\mathbf{x}_*) > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{4.11}$$

where:

$$z(\mathbf{x}_*) = \begin{cases} \frac{\mu_*(\mathbf{x}_*) - f(x^+)}{\sigma_*(\mathbf{x}_*)}, & \sigma_*(\mathbf{x}_*) > 0, \\ 0, & \text{otherwise}. \end{cases} \tag{4.12}$$

Here, $\Phi(\cdot)$ is the cumulative distribution function, and $\phi(\cdot)$ is the probability density function for the standard normal distribution. z is the improvement standardized by the prediction standard deviation σ_*.

Constrained BO restricts the search to the feasible region $\mathcal{F} = \{\mathbf{x} \mid c_i(\mathbf{x}) \leq 0\}$, where constraints $\{c_i(\mathbf{x}) \leq 0\}$ are inferred from data, e.g., modeled with GPRs. Computational studies have shown that CBO can be more efficient than BO. The expected improvement constraint (EIC) acquisition function uses GPR or similar probabilistic models (which are assumed to be independent[62]) to consider the probability the constraints are feasible[63]:

$$\mathrm{EIC}(\mathbf{x}) = \mathrm{EI}(\mathbf{x}) \cdot \prod_i P(c_i(\mathbf{x}) \leq 0). \tag{4.13}$$

Both EI and EIC functions are non-convex with many possible local maxima. We used a local quasi-Newton method with the multi-start strategy. The entire procedure, including hyperparameter training and EIC optimization, requires less than 4 minutes.

4.3.4 Multi-objective optimization

Multi-objective optimization systematically determines the Pareto optimal trade-offs between conflicting objectives. A point is Pareto optimal if improving one objective requires sacrificing at least one other objective[64]. Mathematically, $\min\{f_i(\mathbf{x})\}$ denotes the multi-objective problem where $i \in \mathcal{I}$ denotes the index of the objectives. Multi-objective optimization problems are often solved via the epsilon-constrained methods by maximizing one objective (denoted 1 without loss of generality) while constraining all other objectives to be less than the tolerance ϵ_i:

$$\begin{aligned} &\min\ f_1(\mathbf{x}), \\ &\text{s.t.}\quad f_i(\mathbf{x}) \leq \epsilon_i,\ \forall i \in \mathcal{I} \setminus \{1\}. \end{aligned} \tag{4.14}$$

Problem (4.14) was solved for many ϵ_i values, identifying solutions within the Pareto set. For our proposed workflow, problem (4.13) was implemented and solved in Scipy[65] (version 1.11.4) with a trust-region solver, and the optimization procedure requires one minute for each point. We note that establishing Pareto optimality requires solving Eq. (4.14) to provable global optimality[66]. We solved Eq. (4.14) with a multi-start search as a global optimization heuristic. This means our reported results are likely in the Pareto set.

4.4 Results and discussion

4.4.1 Data-driven determination of the plasma jet operating envelope (Phase 1)

We first identified the operating envelope for the U_{on} and U_{max} of the U_a as a function of Q and f, using Gaussian Process Regression (GPR):

$$U_{\text{on}} \sim \mathcal{GP}(\mathbf{x}_i^{(1)}), \quad U_{\text{max}} \sim \mathcal{GP}(\mathbf{x}_i^{(1)}), \quad \mathbf{x}_i^{(1)} = (Q, f). \tag{4.15}$$

We employed a full factorial design with seven levels for Q from 100 to 1900 sccm and f from 20 to 80 kHz, which correspond to the limitations of the experimental apparatus. We used the data from these 49 experiments (see Table B.2 in the SI) to train the GPR models. The prediction results are shown in Fig. 4.3, which show that overall $U_{\text{on}} \geq 2.2$ kV and $U_{\text{max}} \leq 8.7$ kV, with the exact values depending on Q and f. For most flow rates and frequencies, the plasma jet was uniform between $U_{\text{on}} \approx 2.5$ kV and $U_{\text{max}} \approx 5.5$ kV. Although the relationship among Q, U, and f is complicated and not well understood, it is clear that Q and f have a much more pronounced effect on the maximum operating voltage than the onset voltage, and these effects are particularly more pronounced at lower flow rates ($Q < 1000$ sccm in Fig. 4.3). Qualitatively, we observed that at higher flow rates, the plasma jet propagates out of the dielectric tube more easily. This is likely because of the balance of heating by the plasma with heat transfer through the quartz tube affecting the formation of surface ionization waves along the tube surface, though there are likely other relevant behaviors as well. More work is needed to better understand these underlying relationships. For this study, the GPR models accurately predicted the onset and maximum voltages, respectively, and as such, we utilized them as envelope constraints for SEI optimization in Phase 2.

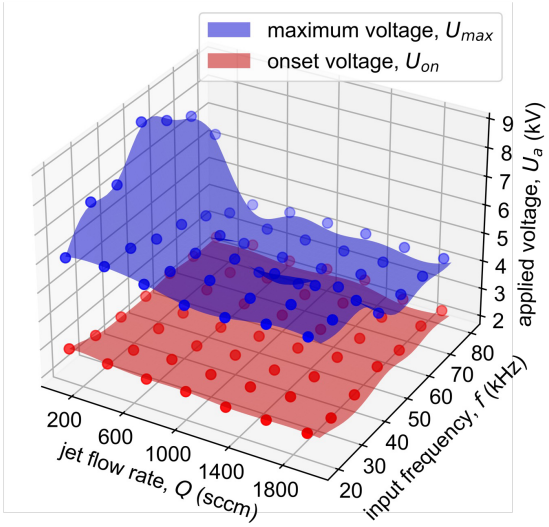

Figure 4.3. Operating envelope for U_{on} and U_{max} as a function of Q and f. Blue and red dots represent experimental data, and the surfaces are the prediction means of the GPR models for U_{on} and U_{max}, respectively.

4.4.2 Bayesian optimization maximizes specific energy input (Phase 2)

Three decision variables, Q, U_a, and f, determine SEI, and importantly, these three variables have synergistic (and likely nonlinear) roles in influencing plasma jet sintering results, which motivates the following optimization problem:

$$\begin{aligned} &\max_{\mathbf{x}_i^{(2)}} \text{SEI} \sim \mathcal{GP}(\mathbf{x}_i^{(2)}) \\ &\text{s.t.} \quad U_{\text{on}}(\mathbf{x}_i^{(1)}) \leq U_a \leq U_{\text{max}}(\mathbf{x}_i^{(1)}). \end{aligned} \tag{4.16}$$

where $\mathbf{x}_i^{(1)} = \mathbf{x}_i^{(2)} = (Q, f)$. Latin-hypercube sampling (LHS) with nine levels generated nine initial data samples for round 0 of SEI optimization (see Table B.3 in the SI). Subsequently, we employed Bayesian Optimization (BO) to iteratively refine the SEI prediction, conducting eight rounds, each consisting of five individual

experiments, to maximize the SEI.

Figure 4.4(a) summarizes the optimization results, showing the maximum SEI value achieved for each round (with full results in Table B.3 in the SI). For the first four rounds, we utilized constrained BO with the operating envelope constraints. After round 4, we determined that the constraints on U_{on} and U_{max} were too restrictive, such that none of the recommended experiments had surpassed the peak SEI result in round 0. In round 5, we considered both constrained BO (3 experiments) and unconstrained BO (2 experiments). Surprisingly, the unconstrained BO experimental conditions yielded SEI values comparable to the maximum SEI value obtained from the round 0 dataset. This result indicated that the envelope constraint GPR over-restricted the high applied voltage search space, i.e., $U_a \geq 6.0$ kV, for the majority of possible Q and f conditions. By inspecting the leave-one-out prediction shown in the parity plots in Figure B.3 in the SI, we observed that the GPR model for U_{max} is less accurate in high voltage regions, $U_a \geq 7.0$ kV.

We, therefore, relaxed the upper operating envelope constraint to have a global bound of $U_{\text{max}} = 6.5$ kV for all Q and f, and for the subsequent three rounds (6, 7, and 8), we utilized unconstrained BO with this constant bound. In round 8, we achieved a maximum SEI = 0.132 eV atom^{-1} at experimental conditions of $U_a = 6.5$ kV, $f = 45$ kHz, and $Q = 800$ sccm, denoted as $\mathbf{x}_*^{(2)} = (800 \text{ sccm}, 6.5 \text{ kV}, 45 \text{ kHz})$. This highest SEI (0.132 eV atom^{-1}) is a 136% improvement compared to round 0 (0.056 eV atom^{-1}). In hindsight, rounds 1 to 5 showed how BO can fail when over-restrictive constraints, in this case for U_{on} and U_{max}, appear in the GPR model. Rounds 6 to 8, on the other hand, showed how BO efficiently identified experimental conditions that lead to high SEIs under relaxed bounds.

In order to understand the SEI optimization, we conducted sensitivity analyses of the GPR models around $\mathbf{x}_*^{(2)} = (800 \text{ sccm}, 6.5 \text{ kV}, 45 \text{ kHz})$, as shown in Figs. 4.4b-4.4d, to characterize the behavior around the experimental conditions that yielded

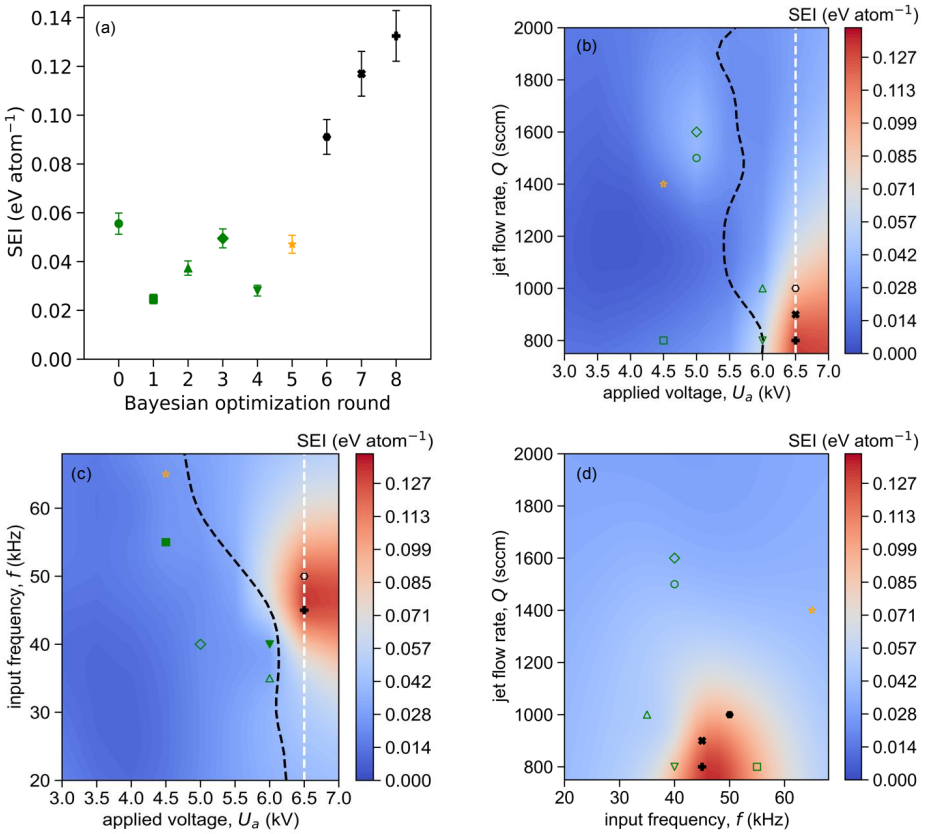

Figure 4.4. BO improves SEI. (a) Plot of the trajectory of SEI optimization over 8 rounds, with the peak experimental SEI value shown for each round. (b)-(d) Plots of 2D sensitivity analyses for SEI, each with a fixed optimal experimental condition: (b) $f = 45$ kHz, (c) $Q = 800$ sccm, and (d) $U_a = 6.5$ kV. The green, yellow, and black markers show the experimental results from rounds 0 to 4, 5, and 7 to 8, respectively. The closed markers represent experimental data, and the open markers represent projections of experimental data to the 2D plane in the plot. The black dotted lines show the envelope constraints $U_{\max}(x_i^{(1)})$ and the white dotted lines show the global bound on $U_{\max} = 6.5$ kV, which is active in rounds 6-8. The error bars represent an uncertainty of 4% for the SEI.

the highest SEI. Figure 4.4b illustrates the relationship between Q and U_a when the input frequency is fixed at $f = 45$ kHz. We observe that an increase in applied voltage increases the SEI value, consistent with Eq. (4.4). Hence, we expected to obtain the

maximum SEI at the maximum value of U_a. We also observe that decreasing the jet flow rate increases the SEI, given that the jet flow rate is inversely proportional to SEI, as indicated in Eq. (4.4).

Figure 4.4c depicts the relationship between U_a and f when the jet flow rate is fixed at $Q = 800$ sccm. Similar trends are shown with applied voltage, while an optimal region exists for input frequency, revealing the non-linear dependence of power on frequency. Figure 4.4d demonstrates the relationship between f and Q when the applied voltage is fixed at $U_a = 6.5$ kV. Again, we observed that input frequency has an optimal operating window and that lower jet flow rates will result in higher SEI values. Overall, these plots show that the optimized point of $\mathbf{x}_*^{(2)} = (800\ \text{sccm}, 6.5\ \text{kV}, 45\ \text{kHz})$ sits within a relatively small window where high SEI values are achievable, with a steep drop-off when moving away from $\mathbf{x}_*^{(2)}$. The prediction of the GPR model also suggested that increasing U_a beyond the global bound, $U_{\max} = 6.5$ kV, may likely achieve higher SEI; but due to experimental apparatus limitations, we decided to terminate the optimization. However, this prediction can be used to inform the design of future experimental apparatuses that have a higher applied voltage operating envelope to increase SEI. Detailed optimization iterations with two-dimensional sensitivity analyses for each round are shown in Figures B.4-B.13 in the SI, and the leave-one-out prediction of SEI is shown in Figure B.14 in the SI.

4.4.3 Bayesian optimization for maximizing electrical conductivity (Phase 3a)

From Phase 2, we obtained the optimal SEI experimental condition $\mathbf{x}_*^{(2)} = (800\ \text{sccm}, 6.5\ \text{kV}, 45\ \text{kHz})$, and with these three decision variables fixed, we then varied the other four decision variables, d, t_{on}, t_{off}, and n, to optimize the electrical conductivity (σ) of sintered ITO samples. Mathematically,

$$\max_{\mathbf{x}_i^{(3)}} \sigma \sim \mathcal{GP}(\mathbf{x}_i^{(3)}),$$
$$\text{s.t.} \quad \mathbf{x}_i^{(2)} = \mathbf{x}_*^{(2)}. \tag{4.17}$$

where $\mathbf{x}_i^{(3)} = (\mathbf{x}_*^{(2)}, d, t_{\text{on}}, t_{\text{off}}, n)$. We used the same round 0 dataset as Phase 2 (shown in Table B.4 in the SI) consisting of nine experiments to initialize Phase 3a optimization. In round 1, we combined $\mathbf{x}_*^{(2)}$ with d, t_{on}, t_{off}, and n values from the round 0 experiments that yielded the highest σ and conducted a single experiment. As shown in Fig. 4.5a, round 1 increased the experimentally measured electrical conductivity by 40% relative to the best σ value in round 0.

In rounds 2 and 3, we used BO to select five experimental conditions in each round, resulting in a 4% improvement in electrical conductivity compared with that in round 1 and a 3% improvement relative to that in round 2, respectively. Although in round 3 we obtained an appreciably high electrical conductivity of $\sigma = 7.421$ S m^{-1} under the experimental condition of $Q = 800$ sccm, $U_a = 6.5$ kV, $f = 45$ kHz, $d = 1.5$ mm, $t_{\text{on}} = 60$ s, $t_{\text{off}} = 90$ s, and $n = 50$, we observed that it resulted in a high substrate temperature of $T_{\text{max}} = 70.1$ °C, surpassing the temperature goal of 50 °C. It also required a long sintering time of approximately two hours. Hence, in Phase 3b, we sought to balance this increase in electrical conductivity with a lower substrate temperature. Detailed optimization iterations with two-dimensional sensitivity analyses for rounds 2 and 3 are shown in Figure B.15 and B.16 in the SI.

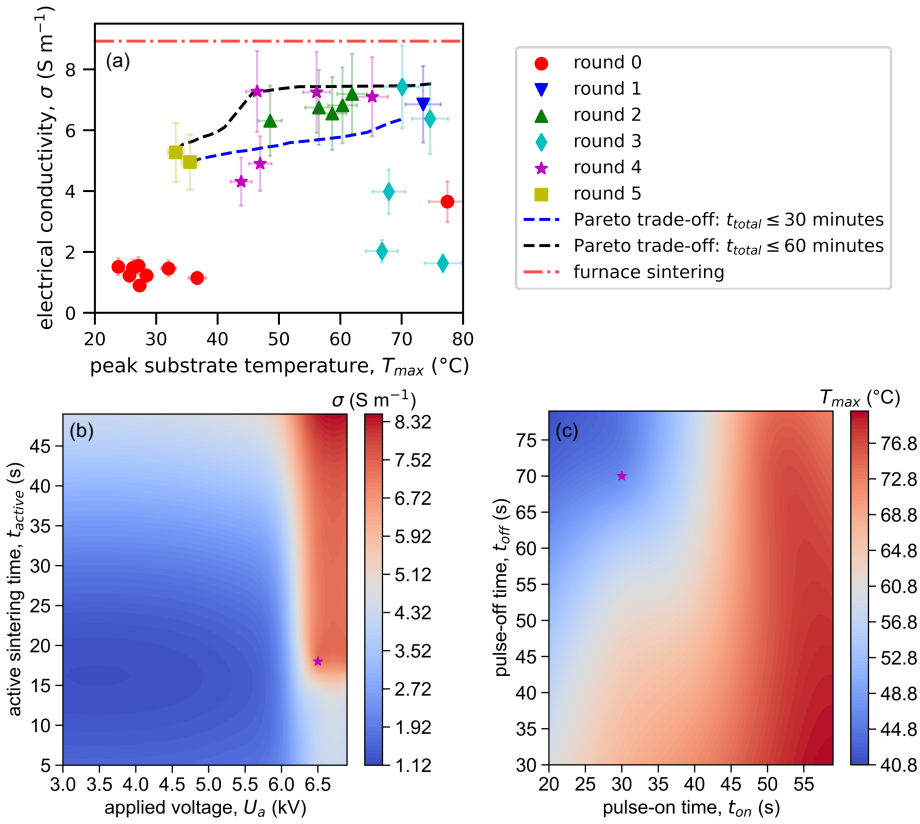

Figure 4.5. Multi-objective optimization of electrical conductivity and peak substrate temperature. (a) Plot of the experimental data from the electrical conductivity optimization, with the red dash-dotted line representing the experimental conductivity obtained via furnace sintering at 300 °C as a reference. The black and blue dashed lines show Pareto trade-offs between electrical conductivity and peak substrate temperature for a total sintering time of 60 min (black) and a projected total sintering time of 30 min (blue). (b) Plot of a 2D sensitivity analysis for σ with fixed $Q = 800$ sccm, $f = 45$ kHz, $d = 2.5$ mm, and $t_{\text{off}} = 70$ s. (c) Plot of a 2D sensitivity analysis for T_{max} with fixed $Q = 800$ sccm, $U_a = 6.5$ kV, $f = 45$ kHz, $d = 2.5$ mm, and $n = 36$. The error bars represent an uncertainty of 9.36% for electrical conductivity and 2% for the peak substrate temperature.

4.4.4 Multi-objective optimization maximizes electrical conductivity and minimizes substrate temperature (Phase 3b)

Based on the single-objective optimization results from Phase 3a (rounds 0 to 3), we formulated a bi-objective optimization problem (Phase 3b):

$$\begin{aligned} &\max_{\mathbf{x}_i^{(3)}} \{\sigma \sim \mathcal{GP}(\mathbf{x}_i^{(3)}), -(T \sim \mathcal{GP}(\mathbf{x}_i^{(3)}))\}, \\ &\text{s.t.} \quad t_{\text{total}} = (t_{\text{on}} + t_{\text{off}}) \cdot n \leq c. \end{aligned} \tag{4.18}$$

where $\mathbf{x}_i^{(3)} = (Q, U_a, f, d, t_{\text{on}}, t_{\text{off}}, n)$ and c is an upper bound for the total experimental time t_{total}. Phase 3b started at round 4, where we conducted sensitivity analyses (shown in Figure B.16 in the SI) using the Phase 2 optimum for SEI of $\mathbf{x}_*^{(2)} = (800 \text{ sccm}, 6.5 \text{ kV}, 45 \text{ kHz})$. The results reveal that the active sintering time is the most important decision variable for electrical conductivity; meanwhile, pulse-on time is the most important decision variable for peak substrate temperature. As such, we fixed the gap distance at $d = 2.5$ mm, which was expected to produce a relatively high electrical conductivity, as shown in Figure B.16(a) in the SI. We also fixed the pulse-on time at $t_{\text{on}} = 30$ s, which was expected to produce a relatively low peak substrate temperature, as shown in Figure B.16(b) in the SI, and fixed the total experimental time as $t_{\text{total}} = 60$ min. We changed the number of cycles with five different values between 24 and 50 and pulse-off time with five different values between 40 s and 120 s, resulting in a 33.7% reduction in the experimentally observed T_{max} from 70.1 °C (round 3) to 46.5 °C (round 4), with only a 2% decrease in the experimentally measured σ from 7.421 S m^{-1} (round 3) to 7.272 S m^{-1} (round 4). Thus, we were able to meet the 50 °C temperature constraint while only having a minimal effect on electrical conductivity. This optimal experimental condition is $x_i^{(3)} = (x_*^{(2)}, d = 2.5 \text{ mm}, t_{\text{on}} = 0.5 \text{ min}, t_{\text{off}} = 1.2 \text{ min}, n = 36)$ and is shown as

experiment 23 in Table B.4 in the SI.

Figures 4.5b and 4.5c show the 2D sensitivity analyses for σ and $T_{\max}$ around the optimal experimental condition. Figure 4.5b shows the influence of U_a and t_{active} on σ; we observe that increasing U_a increases σ, which is consistent with our earlier conclusion that a higher voltage could yield a higher SEI and, thus, a higher electrical conductivity. Meanwhile, we also observe that σ may increase with t_{active}. However, according to additional tests, we found that if the active sintering time was extremely long (three hours or more), the ITO sample's electrical conductivity would not keep increasing, but rather, the ITO sample would be damaged. Given these observations, we understand that further improvement of σ requires higher U_a, i.e., $U_a \geq U_{\max}$; however, this improvement is unavailable in the current experimental apparatus, and we decided to stop further optimizing σ. Figure 4.5c shows the influence of t_{on} and t_{off} on $T_{\max}$. We observed that increasing t_{on} and decreasing t_{off} could increase $T_{\max}$ and vice versa. This is reasonable, as for a fixed number of cycles, a longer pulse-on time means a longer active sintering time. Hence, the energy transfer process between the plasma jet and the sample surface will be longer, leading to an increase in the substrate temperature.

A Pareto set is a set of Pareto optimal solutions, i.e., points where it is not possible to improve one objective without sacrificing one or more of the remaining objectives. The goal of multi-objective optimization is to characterize the Pareto set, which helps understand the trade-off between objectives. To characterize the trade-off between σ and $T_{\max}$ in our system, we decided to explore this relationship in the region of $T_{\max} < 40$ °C by conducting one additional round of experiments (round 5) that considered all seven decision variables. The purpose of this round 5 was not to further increase σ, but to understand how σ might be affected if we restricted $T_{\max} < 40$ °C rather than 50 °C and constrained the total experimental time. We conducted two BO analyses, one for a total experiment time of 60 min

and one for 30 min, and we conducted a single experiment that corresponded to the lowest predicted peak substrate temperature for each.

The results of round 5 are plotted as the two yellow squares in Figure 4.5a, and the corresponding predicted Pareto sets are the solid black ($t_{\text{total}} < 60$ min) and blue ($t_{\text{total}} < 30$ min) lines, respectively. Note that the black line is the GPR prediction mean, which accurately captures the best experimental points from rounds 2-4; it also shows that an improvement in electrical conductivity is constrained by the substrate temperature beyond roughly 50 °C. That is, allowing the substrate temperature to increase substantially would result in only a minor improvement in σ. If we were to reduce the total experimental time to $t_{\text{total}} = 30$ min, we would observe similar behavior, but the electrical conductivity would be generally lower. That is, electrical conductivity is highly sensitive to substrate temperature below 50 °C but more sensitive to total experimental time above 50 °C. Detailed optimization iterations with two-dimensional sensitivity analyses for rounds 4 and 5 are shown in Figure B.18 and B.19 in the SI, and the leave-one-out prediction of σ and T_{max} are shown in Figure B.20 in the SI.

4.4.5 Characterization and comparison between plasma jet sintering and furnace sintering

The optimal plasma jet sintering condition $\mathbf{x}_i^{(3)} = (\mathbf{x}_*^{(2)}, t_{\text{on}} = 0.5 \text{ min}, t_{\text{off}} = 1.2 \text{ min}, n = 36, d = 2.5 \text{ mm})$ yielded an experimental electrical conductivity of 7.272 S m^{-1} in only a 60-minute total experimental time and with the peak substrate temperature never exceeding 46.5 °C. In order to understand this value relative to thermal sintering, we sintered three ITO samples in a tube furnace at 300 °C for a duration of three hours, yielding an electrical conductivity of 8.93 ± 0.05 S m^{-1}. Notably, the plasma jet sintering process achieved 81.4% of the electrical conductivity achieved through furnace sintering while reducing the processing time by a factor of

three and at a temperature of more than 250 °C cooler.

Figure 4.6 shows scanning electron microscopy (SEM) images of the surface structure of ITO samples before (Fig. 4.6a) and after (Figs. 4.6b, c) plasma jet sintering. While there is some densification of the film for the best experiment in round 0 (Fig. 6b), the degree of densification is much greater for the optimal condition in round 4 (Fig. 4.6c). Qualitatively, the densification is similar to that observed for thermal sintering, and thus it is unsurprising that the electrical conductivities were similar.

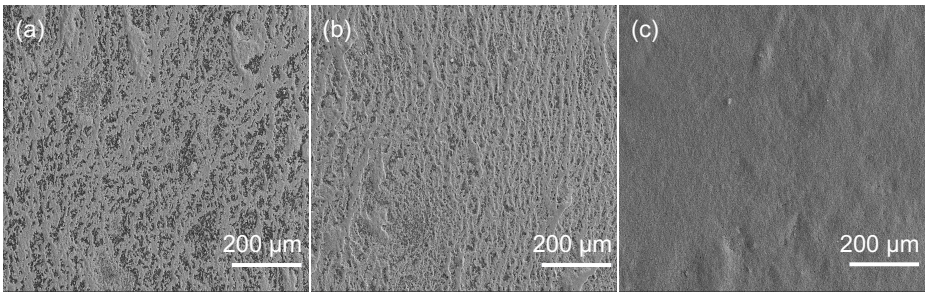

Figure 4.6. Top-view SEM images of ITO thin films: (a) after pre-sintering but before plasma jet sintering, (b) with the highest electrical conductivity after plasma sintering in round 0 (Phase 3a), and (c) with the highest electrical conductivity after plasma sintering in round 4 (Phase 3b). The black regions in (a) and (b) are voids in the thin film.

4.5 Conclusions

This work demonstrates the potential of non-thermal plasma jet sintering as a viable and efficient method for enhancing the electrical conductivity of ITO thin films while maintaining a low substrate temperature. Moreover, Bayesian Optimization

(BO) and Machine Learning (ML) techniques enable the systematic and effective optimization of sintering conditions across many dimensions, resulting in a 99.2% improvement in measured sintering performance within five rounds of experiments. Our findings indicate that non-thermal plasma jet sintering can achieve 81.4% of the electrical conductivity attained by traditional furnace sintering, and this achievement is particularly noteworthy given that non-thermal plasma jet sintering accomplishes this within a significantly reduced time frame and at a far lower temperature. These outcomes signify an advancement in sintering technology compatible with emerging nanoink-based devices and underscore the potential for non-thermal plasma jet sintering to impact the field of materials processing, offering faster and more energy-efficient solutions for enhancing material properties.

Looking forward, the application of machine learning, especially BO, in plasma jet processes shows promising avenues for addressing complicated, high-dimensional problems inherent in the treatments of various materials. Adapting this approach for other plasma jet applications, such as processing alternative semiconducting materials like silver selenide (Ag_2Se), molybdenum disulfide (MoS_2), and tin dioxide (SnO_2), may require more predictive and sophisticated hybrid surrogate models. These models should synergistically integrate physics-informed parameters with data-driven insights to further diminish the experimental burden[67]. This advancement in optimization techniques could lead to universally applicable results, transcending the specific conditions of a single material system. By leveraging the inherent parallelizability and iterative learning capability of BO, we can extend this framework to tackle the increased complexity of multi-material systems, ultimately streamlining the discovery and refinement of plasma jet processes across various domains.

4.6 Acknowledgments

This research was supported by the U.S. Department of Energy by the Center for Sustainable Energy at the University of Notre Dame under Award Number DOE DE-EE0009103. The authors also acknowledge ND Energy and the Patrick and Jana Eilers Graduate Student Fellowship for Energy-Related Research for providing financial support to advance this research.

4.7 References

1. Siya Huang, Yuan Liu, Yue Zhao, Zhifeng Ren, and Chuan Fei Guo. Flexible electronics: stretchable electrodes and their future. *Advanced Functional Materials*, 29(6):1805924, 2019.

2. Tran Quang Trung and Nae-Eung Lee. Flexible and stretchable physical sensor integrated platforms for wearable human-activity monitoringand personal healthcare. *Advanced Materials*, 28(22):4338–4372, 2016.

3. Kyunghun Kim, Bongjoong Kim, and Chi Hwan Lee. Printing flexible and hybrid electronics for human skin and eye-interfaced health monitoring systems. *Advanced Materials*, 32(15):1902051, 2020.

4. Lirong Wang, Tailin Xu, and Xueji Zhang. Multifunctional conductive hydrogel-based flexible wearable sensors. *TrAC Trends in Analytical Chemistry*, 134: 116130, 2021.

5. A Hu, R Li, D Bridges, W Zhou, S Bai, D Ma, and P Peng. Photonic nanomanufacturing of high performance energy devices on flexible substrates. *Journal of Laser Applications*, 28(2), 2016.

6. Najwa Ibrahim, John O Akindoyo, and M Mariatti. Recent development in silver-based ink for flexible electronics. *Journal of Science: Advanced Materials and Devices*, 7(1):100395, 2022.

7. Xianghui Zeng, Pei He, Minglu Hu, Weikai Zhao, Huitong Chen, Longhui Liu, Jia Sun, and Junliang Yang. Copper inks for printed electronics: a review. *Nanoscale*, 14(43):16003–16032, 2022.

8. Elvira Fortunato, Pedro Barquinha, and Rodrigo Martins. Oxide semiconductor thin-film transistors: a review of recent advances. *Advanced Materials*, 24(22): 2945–2986, 2012.

9. Richard Robinson, Virginia Krause, Shan Wang, Shan Yan, Guojun Shang, Justine Gordon, Serena Tycko, and Chuan-Jian Zhong. Silver–copper alloy nanoinks for ambient temperature sintering. *Langmuir*, 38(18):5633–5644, 2022.

10. Nazli Turan, Mortaza Saeidi-Javash, Jiahao Chen, Minxiang Zeng, Yanliang Zhang, and David B Go. Atmospheric pressure and ambient temperature plasma jet sintering of aerosol jet printed silver nanoparticles. *ACS Applied Materials & Interfaces*, 13(39):47244–47251, 2021.

11. Morteza Oghbaei and Omid Mirzaee. Microwave versus conventional sintering: A review of fundamentals, advantages and applications. *Journal of Alloys and Compounds*, 494(1-2):175–189, 2010.

12. Juha Niittynen, Robert Abbel, Matti Mäntysalo, Jolke Perelaer, Ulrich S Schubert, and Donald Lupo. Alternative sintering methods compared to conventional thermal sintering for inkjet printed silver nanoparticle ink. *Thin Solid Films*, 556:452–459, 2014.

13. Wei Liu, Rong An, Chunqing Wang, Zhen Zheng, Yanhong Tian, Ronglin Xu, and Zhongtao Wang. Recent progress in rapid sintering of nanosilver for electronics applications. *Micromachines*, 9(7):346, 2018.

14. Markus Hösel and Frederik C Krebs. Large-scale roll-to-roll photonic sintering of flexo printed silver nanoparticle electrodes. *Journal of Materials Chemistry*, 22(31):15683–15688, 2012.

15. Mortaza Saeidi-Javash, Ke Wang, Minxiang Zeng, Tengfei Luo, Alexander W Dowling, and Yanliang Zhang. Machine learning-assisted ultrafast flash sintering of high-performance and flexible silver–selenide thermoelectric devices. *Energy & Environmental Science*, 15(12):5093–5104, 2022.

16. Ingo Reinhold, Chris E Hendriks, Rebecca Eckardt, Johannes M Kranenburg, Jolke Perelaer, Reinhard R Baumann, and Ulrich S Schubert. Argon plasma sintering of inkjet printed silver tracks on polymer substrates. *Journal of Materials Chemistry*, 19(21):3384–3388, 2009.

17. Sebastian Wünscher, Steffi Stumpf, Anke Teichler, Oliver Pabst, Jolke Perelaer, Erik Beckert, and Ulrich S Schubert. Localized atmospheric plasma sintering of inkjet printed silver nanoparticles. *Journal of Materials Chemistry*, 22(47): 24569–24576, 2012.

18. Sebastian Wünscher, Steffi Stumpf, Jolke Perelaer, and Ulrich S Schubert. Towards single-pass plasma sintering: temperature influence of atmospheric pressure plasma sintering of silver nanoparticle ink. *Journal of Materials Chemistry C*, 2(9):1642–1649, 2014.

19. Kwang-Seok Kim, Jae-Oh Bang, Yong-Ho Choa, and Seung-Boo Jung. The characteristics of cu nanopaste sintered by atmospheric-pressure plasma. *Microelectronic Engineering*, 107:121–124, 2013.

20. Peter Bruggeman and Ronny Brandenburg. Atmospheric pressure discharge filaments and microplasmas: physics, chemistry and diagnostics. *Journal of physics D: Applied Physics*, 46(46):464001, 2013.

21. Dogan Gidon, Xuekai Pei, Angelo D Bonzanini, David B Graves, and Ali Mesbah. Machine learning for real-time diagnostics of cold atmospheric plasma sources. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(5):597–605, 2019.

22. Ali Mesbah and David B Graves. Machine learning for modeling, diagnostics, and control of non-equilibrium plasmas. *Journal of Physics D: Applied Physics*, 52(30):30LT02, 2019.

23. Angelo D Bonzanini, David B Graves, and Ali Mesbah. Learning-based smpc for reference tracking under state-dependent uncertainty: An application to atmospheric pressure plasma jets for plasma medicine. *IEEE Transactions on Control Systems Technology*, 30(2):611–624, 2021.

24. Angelo D Bonzanini, Ketong Shao, Augusto Stancampiano, David B Graves, and Ali Mesbah. Perspectives on machine learning-assisted plasma medicine: Toward automated plasma treatment. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 6(1):16–32, 2021.

25. Shane A Lynn, Niall MacGearailt, and John V Ringwood. Real-time virtual metrology and control for plasma etch. *Journal of Process Control*, 22(4):666–676, 2012.

26. Florian Krüger, Tobias Gergs, and Jan Trieschmann. Machine learning plasma-surface interface for coupling sputtering and gas-phase transport simulations. *Plasma Sources Science and Technology*, 28(3):035002, 2019.

27. Mortaza Saeidi-Javash, Wenzheng Kuang, Chaochao Dun, and Yanliang Zhang. 3d conformal printing and photonic sintering of high-performance flexible thermoelectric films using 2d nanoplates. *Advanced Functional Materials*, 29(35): 1901930, 2019.

28. Ke Wang, Minxiang Zeng, Jialu Wang, Wenjie Shang, Yanliang Zhang, Tengfei Luo, and Alexander W Dowling. When physics-informed data analytics outperforms black-box machine learning: A case study in thickness control for additive manufacturing. *Digital Chemical Engineering*, 6:100076, 2023.

29. Bridgette J Befort, Alejandro Garciadiego, Jialu Wang, Ke Wang, Gabriela Franco, Edward J Maginn, and Alexander W Dowling. Data science for thermodynamic modeling: Case study for ionic liquid and hydrofluorocarbon refrigerant mixtures. *Fluid Phase Equilibria*, 572:113833, 2023.

30. Xiaoqiang Xu, Xianfeng David Gu, and Shikui Chen. Topology optimization of thermal cloaks in euclidean spaces and manifolds using an extended level set method. *International Journal of Heat and Mass Transfer*, 202:123720, 2023.

31. Ketong Shao, Xuekai Pei, David B Graves, and Ali Mesbah. Active learning-guided exploration of parameter space of air plasmas to enhance the energy efficiency of no x production. *Plasma Sources Science and Technology*, 31(5): 055018, 2022.

32. Minxiang Zeng, Shuai Yuan, Dali Huang, and Zhengdong Cheng. Accelerated design of catalytic water-cleaning nanomotors via machine learning. *ACS Applied Materials & Interfaces*, 11(43):40099–40106, 2019.

33. Wenjie Shang, Minxiang Zeng, ANM Tanvir, Ke Wang, Mortaza Saeidi-Javash, Alexander Dowling, Tengfei Luo, and Yanliang Zhang. Hybrid data-driven discovery of high-performance silver selenide-based thermoelectric composites. *Advanced Materials*, 35(47):2212230, 2023.

34. Robert B Gramacy. *Surrogates: Gaussian Process Modeling, Design, and Optimization for The Applied Sciences*. Chapman and Hall/CRC, 2020.

35. George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

36. Jiahang Zhou, Ruiyang Li, and Tengfei Luo. Physics-informed neural networks for solving time-dependent mode-resolved phonon boltzmann transport equation. *npj Computational Materials*, 9(1):212, 2023.

37. Elvis A Eugene, Kyla D Jones, Xian Gao, Jialu Wang, and Alexander W Dowling. Learning and optimization under epistemic uncertainty with bayesian hybrid models. *Computers & Chemical Engineering*, 179:108430, 2023.

38. Ali Newaz Mohammad Tanvir, Md Omarsany Bappy, Minxiang Zeng, Wenjie Shang, Ke Wang, Kaidong Song, Yukun Liu, Eleonora Isotta, Mercouri G Kanatzidis, G Jeffrey Snyder, et al. High-performance thermoelectric composites via scalable and low-cost ink processing. *Energy & Environmental Science*, 2024.

39. Sixian Jia, Jieliyue Sun, Andrew Howes, Michelle R Dawson, Kimani C Toussaint Jr, and Chenhui Shao. Hybrid physics-guided data-driven modeling for generalizable geometric accuracy prediction and improvement in two-photon lithography. *Journal of Manufacturing Processes*, 110:202–210, 2024.

40. Deepak Akhare, Zeping Chen, Richard Gulotty, Tengfei Luo, and Jian-Xun Wang. Probabilistic physics-integrated neural differentiable modeling for isothermal chemical vapor infiltration process. *npj Computational Materials*, 10(1):120, 2024.

41. ND Md Sin, M Fuad Kamel, Rosalena Irma Alip, Zulfakri Mohamad, and M Rusop. The electrical characteristics of aluminium doped zinc oxide thin film for humidity sensor applications. *Advances in Materials Science and Engineering*, 2011(1):974906, 2011.

42. T Neubert, F Neumann, K Schiffmann, P Willich, and A Hangleiter. Investigations on oxygen diffusion in annealing processes of non-stoichiometric amorphous indium tin oxide thin films. *Thin Solid Films*, 513(1-2):319–324, 2006.

43. Nazli Turan, Mortaza Saeidi-Javash, Yanliang Zhang, and David B Go. Does plasma jet sintering follow an arrhenius-type expression? *Plasma Processes and Polymers*, 19(8):2200011, 2022.

44. James A Svoboda and Richard C Dorf. *Introduction to Electric Circuits*. John Wiley & Sons, 2013.

45. Patrick F Dunn and Michael P Davis. *Measurement and Data Analysis for Engineering and Science*. CRC press, 2017.

46. Hassan Jafari, Mohd Hsbullah Idris, and Amirreza Shayganpour. Evaluation of significant manufacturing parameters in lost foam casting of thin-wall al–si–cu alloy using full factorial design of experiment. *Transactions of Nonferrous Metals Society of China*, 23(10):2843–2851, 2013.

47. Christopher KI Williams and Carl Edward Rasmussen. *Gaussian Processes for Machine Learning*, volume 2. MIT press Cambridge, MA, 2006.

48. Ke Wang, Mortaza Saeidi-Javash, Minxiang Zeng, Zeyu Liu, Yanliang Zhang, Tengfei Luo, and Alexander W Dowling. Gaussian process regression machine learning models for photonic sintering. In *Computer Aided Chemical Engineering*, volume 49, pages 1819–1824. Elsevier, 2022.

49. Christopher M Bishop and Nasser M Nasrabadi. *Pattern Recognition and Machine Learning*, volume 4. Springer, 2006.

50. Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.

51. Lieven Vandenberghe and Stephen Boyd. *Convex Optimization*, volume 1. Cambridge university press Cambridge, 2004.

52. Oliver Kramer. *Machine Learning for Evolution Strategies*, volume 20. Springer, 2016.

53. Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, 23(4):550–560, 1997.

54. Tian Wang, Cheng Zhang, Hichem Snoussi, and Gang Zhang. Machine learning approaches for thermoelectric materials research. *Advanced Functional Materials*, 30(5):1906041, 2020.

55. Zhufeng Hou, Yoshiki Takagiwa, Yoshikazu Shinohara, Yibin Xu, and Koji Tsuda. Machine-learning-assisted development and theoretical consideration for the al2fe3si3 thermoelectric material. *ACS Applied Materials & Interfaces*, 11 (12):11545–11554, 2019.

56. Chenxi Tian, Tianjiao Li, Jenniffer Bustillos, Shonak Bhattacharya, Talia Turnham, Jingjie Yeo, and Atieh Moridi. Data-driven approaches toward smarter additive manufacturing. *Advanced Intelligent Systems*, 3(12):2100014, 2021.

57. Ryan-Rhys Griffiths and José Miguel Hernández-Lobato. Constrained bayesian optimization for automatic chemical design using variational autoencoders. *Chemical Science*, 11(2):577–586, 2020.

58. Ke Wang and Alexander W Dowling. Bayesian optimization for chemical products and functional materials. *Current Opinion in Chemical Engineering*, 36: 100728, 2022.

59. Peter I Frazier. Bayesian optimization. In *Recent Advances in Optimization and Modeling of Contemporary Problems*, pages 255–278. INFORMS, 2018.

60. Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13: 455–492, 1998.

61. Stewart Greenhill, Santu Rana, Sunil Gupta, Pratibha Vellanki, and Svetha Venkatesh. Bayesian optimization for adaptive experimental design: A review. *IEEE Access*, 8:13937–13948, 2020.

62. Benjamin Letham, Brian Karrer, Guilherme Ottoni, and Eytan Bakshy. Constrained bayesian optimization with noisy experiments. 2019.

63. Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.

64. Alexander W Dowling, Gerardo Ruiz-Mercado, and Victor M Zavala. A framework for multi-stakeholder decision-making and conflict resolution. *Computers & Chemical Engineering*, 90:136–150, 2016.

65. Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272, 2020.

66. Kaisa Miettinen. *Nonlinear Multiobjective Optimization*, volume 12. Springer Science & Business Media, 1999.

67. Niket Sharma and YA Liu. A hybrid science-guided machine learning approach for modeling chemical processes: A review. *AIChE Journal*, 68(5):e17609, 2022.

CHAPTER 5

WHEN PHYSICS-INFORMED DATA ANALYTICS OUTPERFORMS BLACK-BOX MACHINE LEARNING: A CASE STUDY IN THICKNESS CONTROL FOR ADDITIVE MANUFACTURING

This chapter is based on previously published work:

Ke Wang, Minxiang Zeng, Jialu Wang, Wenjie Shang, Yanliang Zhang, Tengfei Luo, and Alexander W Dowling. When physics-informed data analytics outperforms black-box machine learning: A case study in thickness control for additive manufacturing. *Digital Chemical Engineering*, 6: 100076, 2023.

5.1 Introduction

Recent advances in printing technologies have enabled the direct transformation of nanoparticle inks into functional devices [1]. In particular, aerosol jet printing (AJP) has emerged as a promising additive manufacturing method due to its high printing resolution and broad material options. In the past decade, a wide range of materials have been printed into various complex device architectures, including temperature sensors [2], photodetectors [3], piezoelectric sensors [4], and thermoelectric devices [5, 6]. During the AJP process, the amount of materials deposited, often referred to as deposition thickness, is influenced by several decision variables, such as carrier gas flowrate, sheath flowrate, printing speed, and atomizing voltage [1]. The interplay and nonlinear effect of these printing decision variables make it difficult to accurately predict and precisely control the deposition thickness [7, 8]. Moreover, for certain material systems and additive manufacturing methods, controlling film thickness is

essential to maximize material performance [9]. Thus, there is a great need for reliable methods to precisely control printing thickness to further expand the application domain of AJP.

Currently, there are three main categories of AJP manufacturing optimization studies: experimental empirical investigations, physics-informed modeling, and data-driven black-box machine learning (ML) optimization [10]. In the first category, [7] studied the influence of manufacturing conditions (e.g., carrier gas flowrate) on the line width and thickness of the printed film based on one-dimensional sensitivity analysis. This experimental empirical investigation focused on one-dimensional sensitivity analyses. The method can study the influence of one experiment condition on the measurement by fixing other experimental conditions but ignores the interaction between experimental conditions and can be highly biased by the intuition of the experimental expert.

In the second category, physics-informed modeling, [11] proposed a 3D computational fluid dynamics (CFD) model to investigate the influence of the tamper angle of the nozzle and the jet printing height on the droplet of the printed line. [12] developed a simplified first-principle model to replace Stokes and Saffman force-based numerical calculation for the prediction of the width of the printed line. The physics modeling allows researchers to build up a complete physical model to describe the experiment phenomenon, but the assumptions in the model may be impractical. Additionally, in some scenarios, e.g., computational fluid dynamics, the computation time can be a problem.

In the third and final category, ML optimization, [13] utilized linear regression to study the relationship between manufacturing experimental conditions and printed line features. [14] developed a hybrid machine learning model with k-means clustering and support vector machine (SVM) to identify the optimal operation window of the printing process. [15] proposed a multi-objective optimization framework with

Gaussian process regression (GPR) to increase the overall printed line quality. These black-box machine learning models can give accurate predictions and are widely applied across engineering disciplines, but often requires a large to massive dataset and lack physical interpretability. While physics-informed ML is seen by many as a way to reduce these data requirements [16–18], we are unaware of any prior work using these emerging methods for AJP systems.

Complementary to these prior studies, we consider a data science framework built upon model-based design of experiments (MBDoE) principles to postulate, train, and validate predictive physics-informed models for the AJP process using limited data. We emphasize our approach prioritizes developing physically interpretable models that are well suited for optimization and inverse control, similar to [19] and [20], and can ultimately support molecular-to-systems engineering [21]. The MBDoE framework is a generalization of classical design of experiment (DoE) methods, which provides a systematic framework to recommend new experiments based on previous observations. These so-called black-box DoE methods use response surface models (e.g., polynomial models) to map the factors and response outputs. Black-box DoE methods are suitable for applications with limited knowledge of underlying processes and are usually combined with the analysis of variance (ANOVA) to identify the influential features. For example, [22] studied the influence of operating decision variables (e.g., working pressure, filling ratio), on the performance of the thermoelectric generator (TEG) system by the three-level Box-Behnken response surface method. Likewise, [23] explored the relationship between the thermoelectric properties of the ball-milled materials and main operation parameters (e.g., duration of milling, speed) by full factorial design. MBDoE, on the other hand, builds upon the scientific model with physically meaningful parameters. MBDoE recommends the most informative experiments based on candidate models to accelerate the identification of the optimal model and corresponding parameters [24]. A case study shown by [25] demonstrates

that the MBDoE usually requires fewer experiments and thus can reduce total labor, time, and resource costs.

In this study, we adapted a general data science and MBDoE framework from the literature [24–28] to AJP additive manufacturing processes, which is summarized in Figure 5.1. Starting with data already available experiments from equipment commissioning, we develop and validate physics-informed data-driven mathematical models to describe the film printing process. These models enable precise inverse control of thickness for each printed film with quantified uncertainty. We find the proposed parametric model outperforms a machine learning alternative, e.g., nonparametric GPR, calibrated using the same data. This finding underscores the data efficiency and efficacy of physics-based models and is consistent with the recent popularity of physics-informed ML methods.

This chapter is organized as follows. Section 5.2 introduces the experimental methods and mathematical modeling framework. Section 5.3 reports the results of sensitivity analysis, MBDoE model selection, and inverse design, and shows the performance comparison between the physics-informed model and GPR. Finally, Section 5.4 presents conclusions and future research opportunities.

5.2 Methods

5.2.1 Experimental apparatus and manufacturing process

The AJP setup consists of an ultrasonic atomizer, printhead, temperature control stage, and motion control system that controls the real-time position and velocity during the targeted move and jogging operations. For AJP printing, the inks are atomized in aerosols which are transferred to the printhead by a nitrogen carrier gas flowrate. The printhead applies a sheath flowrate to narrow the aerosols into a fine jet which is deposited onto a pre-cleaned substrate during the temperature

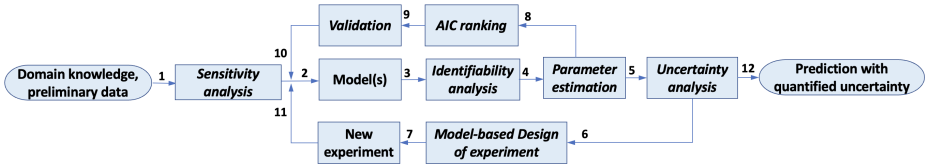

Figure 5.1. Data science tools are integrated into an iterative workflow to postulate, select and calibrate the candidate models, resulting in the final model with quantified prediction uncertainty. The workflow is adapted from the literature [24–28] and starts with domain knowledge and preliminary data on the film printing manufacturing process. This knowledge is used to conduct the sensitivity analysis (arrow 1), exploring the potential relation between decision variables and measurement. These findings help postulate the physics-informed candidate model library (arrow 2). Given the model library, identifiability analysis is conducted (arrow 3) to determine if the parameters in each model can be uniquely inferred from the given data. Next, parameter estimation is conducted with identifiable models (arrow 4). Given the estimated parameters for each candidate model (arrow 8), Akaike Information Criteria (AIC) ranking recommends the best models that balance accuracy and the number of parameters (arrow 9). Using the Fisher information matrix (and its inverse, the parameter covariance matrix) (arrows 5 and 6), MBDoE analysis recommends the most informative new experiments (arrow 7). The down-selected model(s) (arrow 10) and new experiment (arrow 11) provided new information to repeat the entire procedure (arrows 3 to 5). The workflow terminates when the experimental budget has been exhausted or the prediction uncertainty for the selected model is adequately small (arrow 12).

control stage. A heating process accelerates the evaporation of ink solvents to avoid undesired drying effects. The experimental apparatuses are shown in Figure 5.2(a). The morphology of printed film is uniform shown in Figure 5.2(b) and can be modeled as a cubic shown in Figure 5.2(c). Given the fixed length l and width w, the height of the cubic, defined as thickness in our study, is the measurement of our study.

The ultrasonic atomizer generates aerosol droplets from nanoparticle (NPs) dispersion; we use graphene nanoplates as the NPs in this study. The applied voltage U determines the concentration of NPs in the aerosol c. The printhead projects aerosol jet to printed film; carrier gas flowrate Q_a and sheath flowrate Q_s control the amount

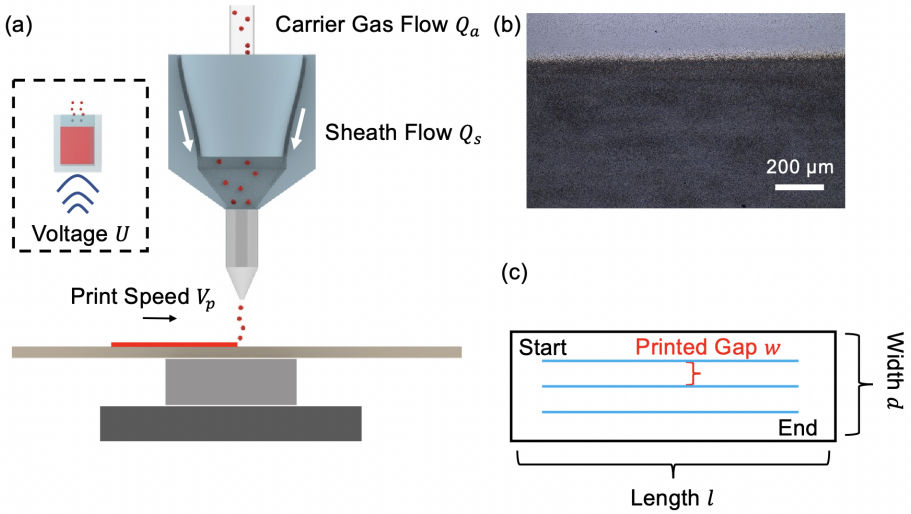

Figure 5.2. (a) The schematic illustration of experiment apparatus. The voltage U applied in the ultrasonic atomizer activates NPs ink to become an aerosol jet, and the aerosol is mixed with the carrier gas flowrate Q_a; then, the sheath flowrate Q_s in printhead projects the mixed particle out; with the motion control system, the nozzle moves back and forth at certain printing speed V_p to print multiple lines to form the film. (b) The printed film. (c) The abstraction of the printed film with thickness h, width d, and length l. The gaps between each line are defined as printed gap w.

of aerosol projected. The motion control system controls the printing speed v_p that determines the printing time t of each sample. We ultimately seek a control strategy to manipulate the experimental degrees of freedom Q_a, Q_s, v_p, and U to print a thin film with desired thickness h. In the remaining of this section, we develop a mathematical modeling framework to help achieve this control objective.

The concentration of NPs, c, is challenging to directly measure. Instead, we assume the voltage U has a nonlinear influence on c, denoted by $g_1(U)$:

$$c = g_1(U). \tag{5.1}$$

Volumetric flow V is the amount of aerosol projected by the nozzle in a unit

of time, which is regulated by aerosol flow Q_a and sheath flowrate Q_s. The exact relationship between these variables is unknown and denoted by $g_2(Q_a, Q_s)$:

$$V = g_2(Q_a, Q_s). \tag{5.2}$$

The momentum flux τ is the weight of projected aerosol in a unit of time and can be derived from the volumetric flow:

$$\tau = c \cdot V. \tag{5.3}$$

The weight m of projected aerosol is controlled by both printing time t and momentum flux τ:

$$m = \tau \cdot t. \tag{5.4}$$

The printed film, shown in Figure 5.2(b), can be modeled as a cuboid, shown in Figure 5.2(c), with length l, width d, and thickness h. Fabrication of the film requires the nozzle to move back and forth to print lines. The interspace exists between lines that the width of it is called printed gap w, as shown in Figure 5.2(c). The porosity ϕ, due to interspace, is defined as a fraction of the volume of voids over the total volume of the printed film; here, we assume the porosity ϕ is a constant. Considering all the factors mentioned above, the weight of the printed film is given by:

$$m = (1 - \phi) \cdot l \cdot d \cdot h \cdot \rho. \tag{5.5}$$

With fixed width d, length l and printed gap w, the printing time t is determined by the printing speed v_p, using the nonlinear function $g_3(v_p)$:

$$t = \frac{l \cdot d}{w} \cdot g_3(v_p). \tag{5.6}$$

We assume the AJP process operates at steady-state, so the weight of film m equals the weight of the nozzle projected m. From Eqs. (5.3) to (5.6), the balanced equation can be derived:

$$(1-\phi)\cdot l\cdot d\cdot h\cdot \rho = \tau\cdot t = c\cdot V\cdot t = \frac{c\cdot V\cdot l\cdot d}{w}\cdot g_3(v_p). \tag{5.7}$$

Simplify the Eq. (5.7) and substituting Eqs. (5.1) and (5.2) gives:

$$h = \frac{c\cdot V}{(1-\phi)\cdot\rho\cdot w\cdot v_p} = \frac{g_1(U)\cdot g_2(Q_a,Q_s)\cdot g_3(v_p)}{(1-\phi)\cdot\rho\cdot w\cdot v_p}. \tag{5.8}$$

In this chapter, we use tools from data science — nonlinear regression, Fisher information matrix analysis (FIM), model-based design of experiment (MBDoE), and Akaike information criteria (AIC) — to systematically determine mathematical expressions for $g_1(U)$, $g_2(Q_a,Q_s)$, and $g_3(v_p)$. Table 5.1 gives numerical values, ranges, and operating windows as bounds for all physical properties used in the above equations.

5.2.2 Decision variables, measurement, and 1D sensitivity analysis

The manufacturing process contains multiple controllable decision variables, including voltage U_i, printing speed v_{pi}, carrier gas flowrate Q_{ai} and sheath flowrate Q_{si}, which influence the measured output variable thickness h_i, where i denotes run i. Let $\mathbf{x}_i = (U_i, v_{pi}, Q_{ai}, Q_{si})$ denote these decision variables. We define the dataset as $\mathbb{D} = \{(\mathbf{x}_i, h_i) \mid \mathbf{x}_i \in \mathbb{R}^4, h_i \in \mathbb{R}, i \in 1,2,\ldots,n\}$. For simplicity, we denote $\mathbb{D} = (\mathbf{X}, \mathbf{h})$, using matrix $\mathbf{X} = (\mathbf{x}_1,\ldots,\mathbf{x}_n)^T$ and vector $\mathbf{h} = (h_1,\ldots,h_n)^T$.

While developing and commissioning the laboratory-scale AJP process, 26 experiments were conducted, which are reported in Table 5.2, and constitute dataset $\mathbb{D}$. Runs 1 through 22 were systematically chosen by an experimental expert to test and develop an understanding of the AJP process. Many of these runs correspond

TABLE 5.1

THIS TABLE DEFINES THE PHYSICAL PROPERTIES IN THE MANUFACTURING PROCESS AND PROVIDES THEIR CORRESPONDING UNIT AND NUMERICAL VALUE OR RANGE.

name	symbol	unit	range or value
Voltage	U	V	$15 - 45$
NPs concentration	c	$\mathrm{g \cdot cm^{-3}}$	$1 - 5 \times 10^{-3}$
NPs density	ρ	$\mathrm{g \cdot cm^{-3}}$	2.267
carrier gas flowrate	Q_a	$\mathrm{cm^3 \cdot min^{-1}}$	$10 - 25$
Sheath flowrate	Q_s	$\mathrm{cm^3 \cdot min^{-1}}$	$40 - 85$
Printing speed	v_p	$\mathrm{cm \cdot s^{-1}}$	$6 - 24$
Volumetric flow	V	$\mathrm{cm^3 \cdot min^{-1}}$	$30 - 100$
Momentum flux	τ	$\mathrm{g \cdot s^{-1}}$	$60 - 250$
Width	d	cm	0.35
Length	l	cm	0.15
Thickness	h	μm	$0.1 - 2.0$
Printed gap	w	cm	0.002
Porosity	ϕ	dimensionless	0.2
Printing time	t	s	$30 - 180$

to one-dimensional (1D) sensitivity analysis between the decision variables $\mathbf{x_i}$ and thickness h_i. These data are typically gathered in the commissioning of a new piece of laboratory equipment with the primary goal of developing expert intuition. In this chapter, we use runs 1 to 22 to postulate a library of mathematical models. Runs 23 to 25 are validation experiments recommended by MBDoE (see Section 5.3.2) and expert opinions. Run 26 is a fourth validation experiment, selected solely by expert

intuition.

The decision variables $\mathbf{x}_i$ and measurement h_i each have different scales (e.g., units) and are accordingly normalized by Eq. (5.9).

$$x_{i,j} \leftarrow \frac{x_{i,j}}{\gamma_j}, \; h_i \leftarrow \frac{h_i}{\gamma_5}. \tag{5.9}$$

Here the index j corresponds to the decision variable, and γ_j denotes the normalizing constants which are given in Table C.1 of the Supporting Information (SI). Unless otherwise noted, the normalized variables are used in the remaining equations in the chapter.

TABLE 5.2

THIS TABLE SUMMARIZES ALL THE RUNS OF EXPERIMENTS DECISION VARIABLES **X** AND MEASUREMENT **h**.

Run	Voltage (V)	Speed ($\text{mm} \cdot \text{s}^{-1}$)	Carrier Gas Flowrate ($\text{cm}^3 \cdot \text{min}^{-1}$)	Sheath Flowrate ($\text{cm}^3 \cdot \text{min}^{-1}$)	Thickness (μm)
1	30	1	14	50	1.23
2	30	2	14	50	0.558
3	30	3	14	50	0.326
4	30	4	14	50	0.27
5	35	2	14	50	0.553
6	40	2	14	50	0.555
7	45	2	14	50	0.623
8	25	2	14	50	0.305
9	15	2	14	50	0

Continued on next page

TABLE 5.2 (CONTINUED)

Run	Voltage (V)	Speed ($\text{mm} \cdot \text{s}^{-1}$)	Carrier Gas Flowrate ($\text{cm}^3 \cdot \text{min}^{-1}$)	Sheath Flowrate ($\text{cm}^3 \cdot \text{min}^{-1}$)	Thickness (μm)
10	10	2	14	50	0
11	35	2	11	50	0.203
12	35	2	12	50	0.268
13	35	2	13	50	0.313
14	35	2	14	80	0.251
15	35	2	14	70	0.357
16	35	2	14	40	0.438
17	35	2	14	30	0.324
18	35	2	14	20	0.162
19	35	2	16	60	0.794
20	35	2	14	60	0.453
21	35	2	12	60	0.193
22	35	2	10	60	0.0158
23	45	1	25	85	0.93
24	35	1	25	60	2.09
25	40	1	25	60	1.8
26	35	1	20	60	1.56

Note: Runs 1 to 22 are from the commissioning of the experimental equipment. Runs 23 to 26 are from validation conducted after modeling.

5.2.3 Parameter estimation

Given the dataset $\mathbb{D}$, model parameters $\boldsymbol{\beta}$ can be estimated with the following procedure. We define the residual sum of square (RSS) function as:

$$\Phi(\boldsymbol{\beta}, \mathbf{X}) \coloneqq \frac{1}{2}(\mathbf{h} - \mathbf{f}(\boldsymbol{\beta}, \mathbf{X}))^T(\mathbf{h} - \mathbf{f}(\boldsymbol{\beta}, \mathbf{X})), \tag{5.10}$$

where $\mathbf{f} : \mathbb{R}^{n \times p} \to \mathbb{R}^n$, and p is the number of parameters. To keep the notation uncluttered, we write $\Phi(\boldsymbol{\beta})$ in the rest of the chapter. The parameter estimation problem is equivalent to minimizing the RSS function, which gives:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \Phi(\boldsymbol{\beta}). \tag{5.11}$$

The optimization problem is solved by MINPACK using the trust region reflective method available in Scipy [29] and central difference to approximate the Jacobian matrix $\nabla_{\boldsymbol{\beta}}\Phi(\hat{\boldsymbol{\beta}})$. A multi-start strategy helps avoid local minima by sampling initial values from a uniform distribution between 0.0 to 1.0 ten times.

5.2.4 Uncertainty propagation

After computing the estimate $\hat{\boldsymbol{\beta}}$, we can obtain the prediction uncertainty via first-order uncertainty propagation [30]. Assume the measured thickness $\mathbf{h}$ is corrupted by random measurement error $\boldsymbol{\varepsilon}$, which follows homoskedastic standard Gaussian distribution, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Assuming the nonlinear regression model $\mathbf{f}(\hat{\boldsymbol{\beta}})$ is true, the observations are:

$$\mathbf{h} = \mathbf{f}(\hat{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon}. \tag{5.12}$$

We wish to estimate how the measurement uncertainty $\boldsymbol{\varepsilon}$ propagates into uncertainty about the estimated regression parameters $\hat{\boldsymbol{\beta}}$. Taking the first-order Taylor expansion on nonlinear regression model $\mathbf{f}(\hat{\boldsymbol{\beta}})$, the unbiased estimator of the σ^2 can be approximated and calculated by [31]:

$$\hat{\sigma}^2 = \frac{2\Phi(\hat{\boldsymbol{\beta}})}{n-p}. \tag{5.13}$$

The corresponding covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ is estimated by [32]:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \hat{\sigma}^2(\nabla_{\boldsymbol{\beta}}\Phi(\hat{\boldsymbol{\beta}})^T\nabla_{\boldsymbol{\beta}}\Phi(\hat{\boldsymbol{\beta}}))^{-1} \tag{5.14}$$

For an arbitrary experiment $\mathbf{x_i}$, the prediction variance $\text{var}(h_i|\hat{\boldsymbol{\beta}}, \mathbf{x_i})$ is given by:

$$\text{var}(h_i|\hat{\boldsymbol{\beta}}, \mathbf{x_i}) = \nabla_{\boldsymbol{\beta}}\mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{x_i})\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}\nabla_{\boldsymbol{\beta}}\mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{x_i})^T + \hat{\sigma}^2, \tag{5.15}$$

where the Jacobian vector $\nabla_{\boldsymbol{\beta}}\mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{x_i})$ is also estimated by central difference for convenience. Finally, we can compute a 95% prediction interval (PI) for the thickness of experiment i as follows:

$$\{\ \ \mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{x_i}) - 1.96\sqrt{\text{var}(h_i|\hat{\boldsymbol{\beta}}, \mathbf{x_i})}, \quad \mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{x_i}) + 1.96\sqrt{\text{var}(h_i|\hat{\boldsymbol{\beta}}, \mathbf{x_i})}\ \ \}. \tag{5.16}$$

5.2.5 Model selection criteria

With the estimated optimal parameter $\hat{\boldsymbol{\beta}}$ for the corresponding model, we can measure the model performance, as well as down-select the model using Akaike Information Criterion (AIC) criteria [33]. The performance of the given model is measured by three criteria: R^2 score, AIC, and mean squared error (MSE). The first two criteria evaluate the model with all data, while the last one evaluates the model with leave-one-out cross-validation, detailed in Section 5.2.7.

The R^2 score shows the goodness of fit by measuring how well the model can

explain the observed data [31] and is calculated by:

$$R^2 = 1 - \frac{2\Phi(\hat{\boldsymbol{\beta}}, \mathbf{X})}{\sum_{i=1}^{n}(h_i - \frac{1}{n}\sum_{i=1}^{n} h_i)}. \tag{5.17}$$

The AIC measures how well the model will generalize to the new data. It is the trade-off between log-likelihood and the model complexity, i.e., the number of parameters, and can be directly calculated by RSS in the least-square regression problem using all data, given by [33]:

$$\mathrm{AIC}(\hat{\boldsymbol{\beta}}, p) = n \cdot \ln\left(\frac{2\Phi(\hat{\boldsymbol{\beta}})}{n}\right) + 2p. \tag{5.18}$$

The preferred model is the one with the smaller AIC value.

The MSE measures the out-of-sample mean squared error and is usually applied with leave-one-out cross-validation to reduce the bias of estimated error. It can be calculated by:

$$\mathrm{MSE} = \frac{2\Phi(\hat{\boldsymbol{\beta}}, \mathbf{X})}{n}. \tag{5.19}$$

5.2.6 Model-based design of experiment (MBDoE)

Design of experiments (DoE) methods aim to identify experiments that maximize the information gain, either to discriminated between models or improve parameter precisions. Model-based design of experiments (MBDoE) methods leverage science-based models to measure such information gain with the Fisher information matrix (FIM) [24]. FIM is defined as the variance of the partial derivative of the natural logarithm of the likelihood function $\mathbf{f}$ with respect to the unknown parameter $\boldsymbol{\beta}$. The FIM of previous experiments, defined as $\mathbf{M}(\boldsymbol{\beta}, \mathbf{X}, \hat{\sigma})$, can be calculated by:

$$\mathbf{M}(\boldsymbol{\beta}, \mathbf{X}, \hat{\sigma}) = \hat{\sigma}^{-2} \sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}} \mathbf{f}(\boldsymbol{\beta}, \mathbf{x_i})^T \nabla_{\boldsymbol{\beta}} \mathbf{f}(\boldsymbol{\beta}, \mathbf{x_i}), \tag{5.20}$$

where i is the index and n is the number of experiments. When adding a new experiment, the total information content, $\mathbf{M}(\boldsymbol{\beta}, \mathbf{X}, \mathbf{x}_{n+1}, \hat{\sigma})$, is measured by the accumulation of all FIM from previous experiments $\mathbf{X}$, and the new experiment $\mathbf{x_{n+1}}$:

$$\mathbf{M}(\boldsymbol{\beta}, \mathbf{X}, \mathbf{x}_{n+1}, \hat{\sigma}) = \mathbf{M}(\boldsymbol{\beta}, \mathbf{X}, \hat{\sigma}) + \hat{\sigma}^{-2} \nabla_{\boldsymbol{\beta}} \mathbf{f}(\boldsymbol{\beta}, \mathbf{x_{n+1}})^T \nabla_{\boldsymbol{\beta}} \mathbf{f}(\boldsymbol{\beta}, \mathbf{x_{n+1}}). \tag{5.21}$$

Large FIM indicates that significant information is contained in the model; A-, D-, and E-optimality metrics are the three most used design criteria [34]. These metrics can be computed using the eigendecomposition of $\mathbf{M}$ [35], where λ_j for $j = \{1, 2, \ldots, p\}$ are the eigenvalues:

$$\mathrm{A-optimality} : \mathrm{trace}(\mathbf{M}) = \sum_{j=1}^{p} \lambda_j, \tag{5.22}$$

$$\mathrm{D-optimality} : \det(\mathbf{M}) = \prod_{j=1}^{j=p} \lambda_j, \tag{5.23}$$

$$\mathrm{E-optimality} : \min_{\lambda_j} \{\lambda_1, \lambda_2, \ldots, \lambda_p\}. \tag{5.24}$$

FIM suffers from the unidentifiable issue when it is (near) singular [36]; in practice, $\mathbf{M}$, is near-singular when the condition number, C_M, is larger than 10^{12} [37]:

$$C_M = \frac{\max_{\lambda_j} \{\lambda_1, \lambda_2, \ldots, \lambda_p\}}{\min_{\lambda_j} \{\lambda_1, \lambda_2, \ldots, \lambda_p\}} \geq 10^{12}. \tag{5.25}$$

Given n previous experiments, and obtaining the next most informative experiment, $\mathbf{x}_{n+1}$, D-optimality is the most popular MBDoE objective function to maximize:

$$\max_{\mathbf{x_{n+1}}} \log(\det \mathbf{M}(\boldsymbol{\beta}, \mathbf{X}, \mathbf{x}_{n+1})) = \sum_{j=1}^{p} \log(\lambda_j). \tag{5.26}$$

where λ_j are the eigenvalues of $\mathbf{M}(\boldsymbol{\beta}, \mathbf{X}, \mathbf{x_{n+1}})$. The optimization problem is solved using Pyomo.DoE [25].

5.2.7 Model validation

Given the set of candidate models, we can further investigate the correctness of the model and choose the optimal model(s). The MBDoE workflow facilitates this by recommending the most informative experiments. Here, two frameworks are adopted for model validation: train-test split and leave-one-out validation. Both of them split dataset $\mathbb{D}$ into the training dataset $\mathbb{D}_1$ and testing dataset $\mathbb{D}_2$. In the train-test split validation framework, dataset $\mathbb{D}$ is split only once to demonstrate the actual prediction performance of the given model using all previously collected data. Mathematically, $\mathbb{D}_1 = (\mathbf{X}_{1:q}, \mathbf{h}_{1:q})$ and $\mathbb{D}_2 = (\mathbf{X}_{k:n}, \mathbf{h}_{k:n})$, where q and k denote the run q and k, respectively, and ':' is an abbreviation for in-between samples.

The leave-one-out validation framework consistently splits dataset $\mathbb{D}$ iteratively for n times to compute leave-one-out mean squared error (LOO-MSE), the out-of-sample error. Mathematically, $\mathbb{D}_1 = (\mathbf{X}_{-i}, \mathbf{h}_{-i})$ and $\mathbb{D}_2 = (\mathbf{X}_i, \mathbf{h}_i)$, where $-i$ denotes all data except sample i. The optimal parameters $\hat{\boldsymbol{\beta}}$ are estimated via Eq. (5.11), using training data $\mathbb{D}_1$. The prediction mean is computed by:

$$\overline{\mathbf{h}} = \mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{X}), \tag{5.27}$$

where $\mathbf{f}$ is the given model. The prediction uncertainty is estimated by using the uncertainty propagation via Eq. (5.16).

5.2.8 Gaussian process regression

In contrast to the parametric model $\mathbf{f}(\hat{\boldsymbol{\beta}})$ described above, Gaussian process regression (GPR) is a nonparametric model, which learns from data directly [38, 39]. The GPR model, $f(\mathbf{x})$, also assumes a random measurement error, similarly to Eq. (5.12), and is defined as [39]:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad \mathbf{x}, \mathbf{x}' \in \mathbb{R}^p, \tag{5.28}$$

where the mean function $m(\mathbf{x})$ is assumed to be zero and $k(\mathbf{x}, \mathbf{x}')$ is the radial basis function (RBF):

$$k(\mathbf{x}, \mathbf{x}') = e^{-\frac{1}{2}\sum_{j=1}^{p}\left(\frac{x_j - x'_j}{\ell_j}\right)^2}. \tag{5.29}$$

The $\boldsymbol{l}$ are the hyperparameters to measure the closeness between different samples; the closer the samples are in the kernel space, the more correlated these two samples [40]. The hyperparameters $\boldsymbol{l}$ is estimated by maximizing log-marginal likelihood function [41]:

$$\boldsymbol{l}^* = \arg\max_{\boldsymbol{l}} \ \log p(\mathbf{h} \mid \mathbf{X}, \boldsymbol{l}) = -\frac{1}{2}\mathbf{h}^T\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{h} - \frac{1}{2}\ln|\mathbf{K}(\mathbf{X}, \mathbf{X})| - \frac{n}{2}\ln 2\pi. \tag{5.30}$$

The hyperparameters $\boldsymbol{l}^*$ are estimated using the L-BFGS-B algorithm implemented in Scipy [29] with ten multi-starts.

Given the dataset $\mathbb{D}$ and hyperparameters $\boldsymbol{l}^*$, we can write the multivariate joint Gaussian distribution in the matrix form:

$$\begin{bmatrix} \mathbf{h} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathbf{X}) \\ \mathbf{m}(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I} & \mathbf{K}(\mathbf{X},\mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*,\mathbf{X}) & \mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) \end{bmatrix} \right), \tag{5.31}$$

where $\mathbf{X}_*$, is new inputs values and $\mathbf{f}_*$ is the corresponding prediction. $\mathbf{K}(\cdot,\cdot)$ is kernel function $k(\cdot,\cdot)$ evaluated element-wise. Through the conjugacy properties in exponential family [42], we can analytically obtain the prediction mean $\mu_*(\mathbf{X}_*)$ and variance $\mathbf{\Sigma}_*(\mathbf{X}_*)$:

$$\mu_*(\mathbf{X}_*) = \mathbf{m}(\mathbf{X}_*) + \mathbf{K}(\mathbf{X}_*,\mathbf{X})\left[\mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I}\right]^{-1}(\mathbf{h} - \mathbf{m}(\mathbf{X})) \tag{5.32a}$$

$$\mathbf{\Sigma}_*(\mathbf{X}_*) = \mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*,\mathbf{X})\left[\mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I}\right]^{-1}\mathbf{K}(\mathbf{X},\mathbf{X}_*) \tag{5.32b}$$

The GPR model is implemented in Scikit-learn [43].

5.3 Results

We apply the data science workflow, described in Figure 5.1 and Section 5.2, with the ultimate goal of controlling AJP sample thickness.

5.3.1 1D sensitivity analysis builds a library of models

5.3.1.1 1D analysis provides data-informed nonlinear equations

The domain knowledge and 1D sensitivity analysis provide possible nonlinear relationships $g_1(U), g_2(Q_a, Q_s)$, and $g_3(v_p)$ between the decision variables and the measured thickness. This step corresponds to arrows 1 and 2 in Figure 5.1. Figure 5.3 visualizes these experiments and shows the corresponding data-informed fitting curve. The parameter estimates (arrows 3 and 4 in Figure 5.1) are calculated via

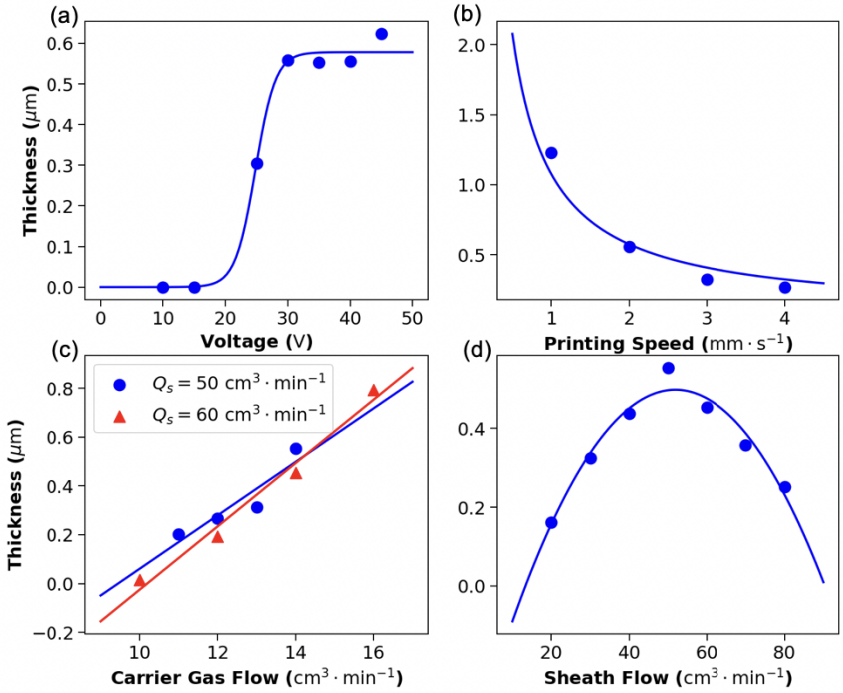

Figure 5.3. One-dimensional sensitivity analysis shows the effect of decision variables on the measured thickness h. (a) Voltage U acts like a threshold, and follows a sigmoid function with thickness. Runs 2, 5, 6, 7, 8, 9, and 10 are used from Table 5.2. (b) Printing speed v_p is inverse proportional to thickness considering the fixed film length and width. Runs 1, 2, 3, and 4 are used from Table 5.2. (c) The carrier gas flowrate Q_a is linear with thickness based on the one-dimensional sensitivity analysis. Runs 5, 11, 12, 13, 19, 20, 21, and 22 are used from Table 5.2. (d) The sheath flowrate Q_s has quadratic relation with thickness based on the one-dimensional sensitivity analysis. Runs 5, 14, 15, 16, 17, 18, and 20 are used from Table 5.2. In all subfigures, the lines show the submodels evaluated with optimized parameters summarized in Table C.2 in the SI.

Eq. (5.11).

In the ultrasound atomizer, the voltage acts as a threshold for generating the NPs. Under a certain threshold, the NPs cannot be activated to become aerosol. Above a certain threshold, NPs will become steady aerosol with a (near) constant NP

concentration. Combining this prior knowledge with the data shown in Figure 5.3(a), we postulate the following sigmoid relationship between voltage U and thickness h:

$$g_1(U) = \frac{1}{1 + \exp\{-\alpha_1 U + \alpha_0\}}, \tag{5.33}$$

where α_0 and α_1 are two fitted parameters. Because such threshold will not vary with the change of other decision variables, we use $\alpha_0^* = 15.38$ and $\alpha_1^* = 30.99$ estimated from the one-dimension analysis in all models. For this nonlinear regression problems, we replace $\mathbf{f}(\boldsymbol{\beta}, \mathbf{X})$ with assumed sigmoid function $g_1(U)$ in Eq. (5.10), and estimate the parameters α_0^* and α_1^* via Eq. (5.11). We chose a sigmoid function as a smooth and differentiable approximation to this switching behavior.

Likewise, from Figure 5.3(b), we propose an inverse proportional relation between printing speed v_p and thickness h:

$$g_3(v_p) = \frac{1}{v_p}. \tag{5.34}$$

Such relation follows the intuition that with fixed width d, length l, printed gap w, and momentum flux τ, thickness is proportional to printing time t, and the printing time t should be inverse proportion to printing speed v_p. Even without the fitting parameters, we observe the data agrees with the model in Figure 5.3(b).

Figure 5.3(c) shows the linear relation between carrier gas flowrate Q_a and thickness h that two lines are plotted with different sheath flowrate Q_s. With fixed sheath flowrate Q_s^*, the relation is given by:

$$g_2(Q_a, Q_s^*) = aQ_a + b, \tag{5.35}$$

where a and b are the regressed parameters.

Finally, Figure 5.3(d) shows the quadratic relation between sheath flowrate Q_s and thickness h with fixed carrier gas flowrate Q_a^*. The relation is given by:

$$g_2(Q_a^*, Q_s) = aQ_s^2 + Q_s b + c, \tag{5.36}$$

where a, b and c are the regressed parameters. Assume the carrier gas flowrate Q_a has linear relation with $g_2(Q_a^*, Q_s)$, we postulate the general formula for $g_2(Q_a, Q_s)$, given by:

$$g_2(Q_a, Q_s) = \beta_{00} + \beta_{10}Q_s + \beta_{20}Q_s^2 + \beta_{01}Q_a + \beta_{11}Q_aQ_s + \beta_{21}Q_aQ_s^2, \tag{5.37}$$

where β_{00}, β_{10}, β_{20}, β_{01}, β_{11}, and β_{21} are parameters, denoted by $\boldsymbol{\beta}$.

5.3.1.2 Model library and down-selection

Combining the above one-dimensional sensitivity analysis results Eqs. (5.33) - (5.37) into Eq. (5.8), gives the most general model, named model 0:

$$h = \frac{1}{1 + \exp(-\alpha_1 U + \alpha_0)} \frac{\beta_{00} + \beta_{10}Q_s + \beta_{20}Q_s^2 + \beta_{01}Q_a + \beta_{11}Q_aQ_s + \beta_{21}Q_aQ_s^2}{\phi\rho w v_p}. \tag{5.38}$$

Model 0 contains all possible combinations of interactions between Q_a and Q_s. We then systematically remove quadratic and (or) cubic terms in the expressions for Q_a and Q_s, giving models 1 to 7. The postulated model library is summarized in Table 5.3.

Table 5.3 shows all models have a moderated high R^2 value, from 0.93 to 0.96. Model 7 is the exception with an R^2 value of 0.85. Instead, we use AIC, computed via Eq. (5.18), to consider the trade-offs between the quality of fit and the number of parameters, as shown in arrows 8 and 9 of Figure 5.1. We down-select and retain only models 0, 1, 3, and 5, which have the highest four AIC values reported in Table

TABLE 5.3

THE POSTULATED MODEL LIBRARY WITH A COMPARISON OF THE GOODNESS OF FIT USING R^2 SCORE AND AIC CRITERIA.

	R^2	AIC	Parameters
Model 0	0.96	-104.0	all regressed
Model 1	0.95	-101.4	$\beta_{21} = 0$
Model 2	0.95	-100.4	$\beta_{20} = 0$
Model 3	0.95	-101.2	$\beta_{11} = 0$
Model 4	0.93	-92.1	$\beta_{21} = 0, \beta_{20} = 0$
Model 5	0.95	-101.9	$\beta_{21} = 0, \beta_{11} = 0$
Model 6	0.94	-98.7	$\beta_{20} = 0, \beta_{11} = 0$
Model 7	0.85	-75.9	$\beta_{21} = 0, \beta_{20} = 0, \beta_{11} = 0$

5.3 (all above -101). These four retained models have high R^2 score values of 0.95 to 0.96, suggesting the variations in the decision (input) variables explain a majority of the variations in thickness (output variable). Table C.3 reports the estimated parameter values for models 0, 1, 3, and 5. Among four selected models, model 0 is unidentifiable using the initial runs 1 to 22 in Table 5.2, but becomes identifiable after adding the optimal MBDoE experiment, run 23, which is explained in Section 5.3.2; models 1, 3, and 5 are identifiable with the initial dataset.

Figure 5.4 is the parity plot for retained models 0, 1, 3, and 5 with optimal trained parameters $\hat{\boldsymbol{\beta}}$. The horizontal axis and vertical axis are experimental and predicted thickness, respectively. The blue dots are all the experiments, runs 1 to 26 in Table 5.2 and the error bars show 95% PI computed via Eq. (5.16). From Figure 5.4, we observe that all of the prediction uncertainty bars capture the dashed parity line,

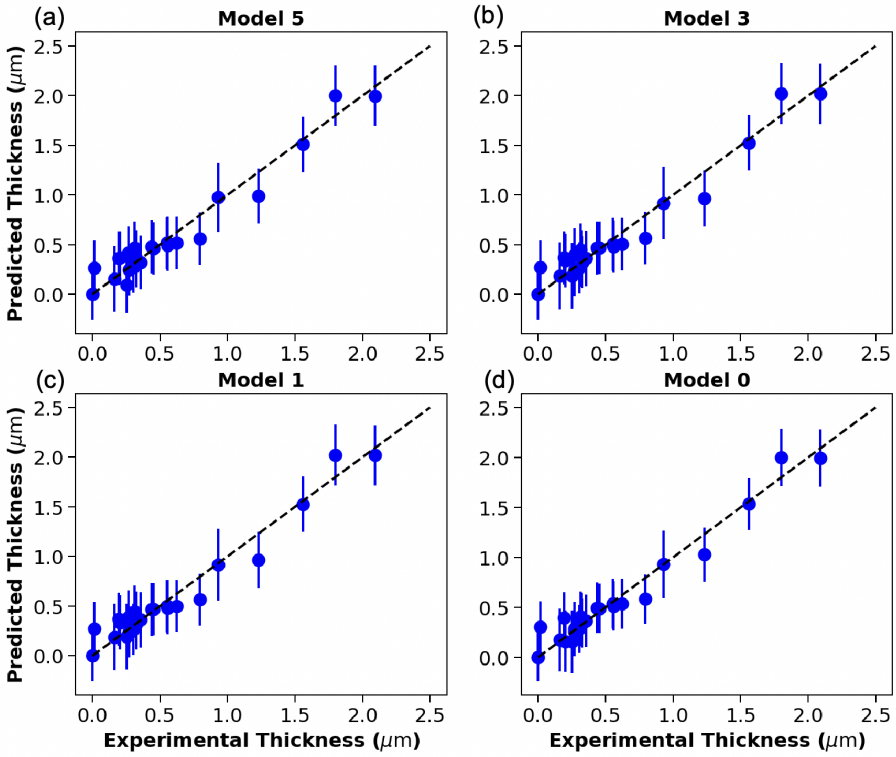

Figure 5.4. Parity plots showing prediction and experimental thickness for (a) model 0, (b) model 3, (c) model 1, and (d) model 5, which were the four best models selected by the AIC ranking. The optimal parameters $\hat{\boldsymbol{\beta}}$ are estimated via Eq. (5.11).

suggesting the four models can successfully explain the relation between decision variables and thickness. This is consistent with the high R^2 scores reported in Table 5.3.

5.3.2 MBDoE informs three validation experiments

The model library in Table 5.3 reports eight models with physically meaningful parameters. We then apply MBDoE to recommend the new experiment(s) to validate these models.

TABLE 5.4

COMPARISON OF THE A-, D-, AND E-OPTIMALITY METRICS FOR MODELS 0 TO 7 WHEN ADDING DIFFERENT SETS OF NEW EXPERIMENTS.

Model	0	1	2	3	4	5	6	7
$\log_{10}$trace($\mathbf{M}$), A-optimality								
Runs 1 to 22	9.237	4.989	4.694	4.961	3.929	4.976	5.201	3.696
Runs 1 to 22: Run 23	9.581	5.160	5.131	5.131	4.236	5.148	5.398	3.869
Runs 1 to 22: Run 24	9.422	5.109	4.932	5.081	4.112	5.095	5.343	3.868
Runs 1 to 22: Run 25	9.422	5.109	4.933	5.082	4.113	5.096	5.344	3.869
Runs 1 to 22: Run 26	9.378	5.092	4.864	5.064	4.069	5.079	5.313	3.823
$\log_{10}$det($\mathbf{M}$), D-optimality								
Runs 1 to 22	26.227	9.523	7.620	9.406	4.704	11.860	12.472	4.072
Runs 1 to 22: Run 23	38.676	12.226	10.334	12.315	7.390	13.179	14.253	5.363
Runs 1 to 22: Run 24	27.746	10.901	8.999	10.784	6.081	13.116	13.717	5.315
Runs 1 to 22: Run 25	27.141	10.903	9.001	10.786	6.083	13.117	13.718	5.317
Runs 1 to 22: Run 26	26.984	10.446	8.543	10.329	5.625	12.686	13.288	4.885

Continued on next page

TABLE 5.4 (CONTINUED)

Model	0	1	2	3	4	5	6	7
$\log_{10}$trace($\mathbf{M}$), A-optimality								
$\log_{10}\min(\text{eig}(\mathbf{M}))$, E-optimality								
Runs 1 to 22	-7.031	-2.370	-2.759	-2.368	-2.128	1.515	1.375	-0.632
Runs 1 to 22: Run 23	2.488	-0.988	-1.525	-0.781	-1.025	1.516	1.644	-0.575
Runs 1 to 22: Run 24	-6.892	-2.247	-2.638	-2.245	-2.034	1.564	1.396	-0.622
Runs 1 to 22: Run 25	-7.498	-2.247	-2.638	-2.245	-2.034	1.564	1.396	-0.622
Runs 1 to 22: Run 26	-7.198	-2.273	-2.665	-2.272	-2.054	1.563	1.398	-0.615

5.3.2.1 Recommendation of new experiments

Before MBDoE analysis, we perform the identifiable analysis of all the models using previous runs 1 to 22. Table C.4 in the SI reports the eigendecomposition of the Fisher information matrix for all of the considered models. The results show that model 0 is unidentifiable using the prior 22 experiments indicated by a near-zero eigenvalue. Models 1 to 7 are identifiable with condition numbers ranging from 10^4 to 10^7. Next, we consider a batch of four more experiments, runs 23 to 26, to quantify the effectiveness of MBDoE and to validate the models. Run 23 in Table 5.2 is the experiment that maximizes D-optimality criteria computed by Eq. (5.26), which corresponds to arrows 6, 7, and 11 in Figure 5.1. Interestingly, for all eight models, MBDoE recommends the same optimal experiment, run 23, with

$U = 45$ V, $v_p = 1$ mm $\cdot$ s^{-1}, $Q_a = 25$ cm$^3 \cdot$ min^{-1}, $Q_s = 85$ cm$^3 \cdot$ min^{-1}. Figure C.1 - C.7 in the SI shows that the different structures of the eight models do not have a large effect on MBDoE recommendation. Recall from Table 5.1, we observe that the experiment designs mentioned above are located in the boundary of the experiment device operation range. This result indicates that, though different models have drastically different mathematical formulas, the MBDoE approach consistently finds that new experiments at the bound contain the most information.

In addition to the one optimal experiment identified by the MBDoE approach, we also propose another two experiments, runs 24 and 25 in Table 5.2, by selecting two arbitrary conditions with high D-optimality values from the corresponding heatmap of model 5, shown in row 1 of Figure 5.5. The last experiment, run 26 in Table 5.2, was selected by only expert intuition without considering the MBDoE results. Figure 5.5 shows these proposed experiments and their influence on the MBDoE metrics.

5.3.2.2 A-, D-, and E-optimality comparison

To quantify the benefits of the newly proposed experiments, Table 5.4 compares the A-, D-, and E-optimality metrics across the model library. Each row is generated by considering calculating the previous experiments, runs 1 to 22, and adding a different proposed experiment, chosen from runs 23 to 26; each column corresponds to one of the eight models. The A-, D-, and E-optimality metrics were $\log_{10}$ transformed. Table C.3 reports the parameters used to compute the A-, D-, and E-optimality metrics.

Comparison between rows in Table C.3, we observe that the A-, D-, and E-optimality metrics increase, which is expected based on the FIM formula, Eq. (5.21). Each new experiment adds to the prior FIM. We also notice that the MBDoE optimal choice, run 23, caused a consistently larger increase in the MBDoE metrics compared to the other three experiments, run 24 to 26; considering model 5 as an example,

compared with the original FIM, the MBDoE optimal choice improved A-optimality by 49%, D-optimality by 1990%, and E-optimality by 0.23%. In contrast, the experiment chosen based on expert intuition (run 26) improved A-optimality by 26%, D-optimality by 570%, and E-optimality by 5.44%. This result indicates the MBDoE optimal choice indeed recommend experiment with higher information.

Comparison between columns in Table C.3, though the dataset is the same, A-, D-, and E-optimality metrics vary across models. Such variation can be again explained by Eq. (5.21) that different models have different estimated parameters $\hat{\boldsymbol{\beta}}$ and variance $\hat{\sigma}^2$, leading to different Jacobian matrix $\nabla_{\boldsymbol{\beta}}\mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{x_{n+1}})$; the estimated parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ for all models are reported in Table C.5 of the SI.

5.3.2.3 Comparison of sensitivity variation with new experiment

Apart from numerical comparison, Figure 5.5 shows the sensitivity analysis of A-, D-, and E-optimality metrics for the same original data and a different new experiment for model 5. Row 1 shows the heatmap with only original data, runs 1 to 22 in Table 5.2, while rows 2 to 5 show the heatmap with original data and a sample from runs 23 to 26 in Table 5.2, accordingly. Figures C.1 - C.7 in the SI provide a similar analysis for the other models.

Next, we perform sensitivity analysis to confirm the A-, D-, and E-optimality metrics, which are computed using local sensitivity estimates, Eq. (5.21), are robust to variations in the model parameters. Following Algorithm 2 described in the SI [44], we draw 1000 parameter samples $\tilde{\boldsymbol{\beta}}$ from a multivariate normal distribution defined by the parameter estimates and their covariance matrix. For each sample, we recompute the A-, D-, and E-optimality metrics and record the distance to the nominal value for these metrics. The histograms of these differences, Figure C.8 in the SI, shows all of the MBDoE metrics change less than 1.5% difference across all Monte Carlo samples. To further verify this conclusion, Figure C.9 in the SI visualizes the heatmaps from

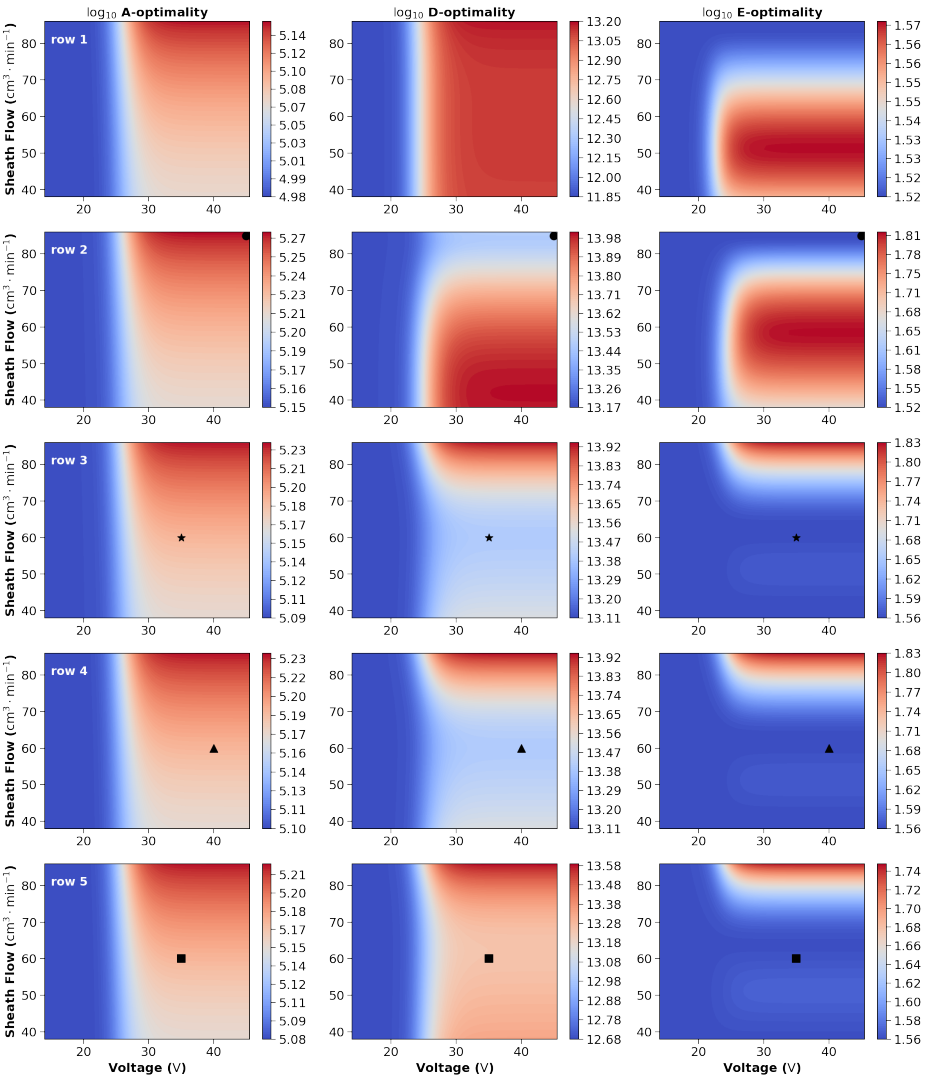

Figure 5.5. This figure shows the sensitivity of A-, D-, and E-optimality metrics for model 5 after a new experiment is added with $Q_a = 25\ \text{cm}^3 \cdot \text{min}^{-1}$ and $V_p = 1\ \text{mm} \cdot \text{s}^{-1}$ fixed. The color indicates the numerical value of corresponding $\log_{10}$ optimality metrics. Row 1 shows figures only considering original data, runs 1 to 22. Rows 2 to 5 consider these prior data (runs 1 to 22) and one new experiment, and are labeled in black dot (add run 23), star (add run 24), triangle (add run 25), and square (add run 26), respectively.

Figure 5.5 using the single sample from the Monte Carlo analysis with the largest difference in D-optimality. Figures C.9 and 5.5 are practically indistinguishable, further demonstrating the results are robust to parameter uncertainty.

Comparison between rows in Figure 5.5 yields three important observations. First, the sensitivities of A-, D-, and E-optimality metrics, given the same new experiment, are different. Adding a new experiment has a relatively small influence on A-optimality because it is the bounding box of a given FIM ellipsoid, and adding a new measurement has minimal impact on these bounds. In contrast, D-optimality and E-optimality metrics significantly vary with the addition of a new experiment. D-optimality is the volume of the FIM ellipsoid and E-optimality is the smallest eigenvalue, which corresponds to the smallest minor axis of the FIM ellipsoid. The new experiment enlarges this smallest minor axis and increases the ellipsoid volume. Secondly, after adding an experiment not recommended by the MBDoE approach, e.g., any experiment from runs 24 to 26, which are rows 3 to 5 in Figure 5.5, the next recommendation of experiment self-corrects. The next optimal MBDoE recommendation is run 23, as shown in the red region of rows 3 to 5 in Figure 5.5. This indicates that most information can be obtained only by considering the MBDoE optimal recommendation experiment (or region). Finally, after adding the MBDoE optimal experiment, run 23, the next MBDoE optimal one is located in the lower right region, see row 2 of Figure 5.5, which verifies the claim above that boundary usually contains the most information. Such observations are common across all models as shown by Figure C.1 - C.7 in the SI.

5.3.3 Validated models enable precise thickness control

With the new experiments recommended by MBDoE, we select the best model and use it for predictive inverse control of thickness.

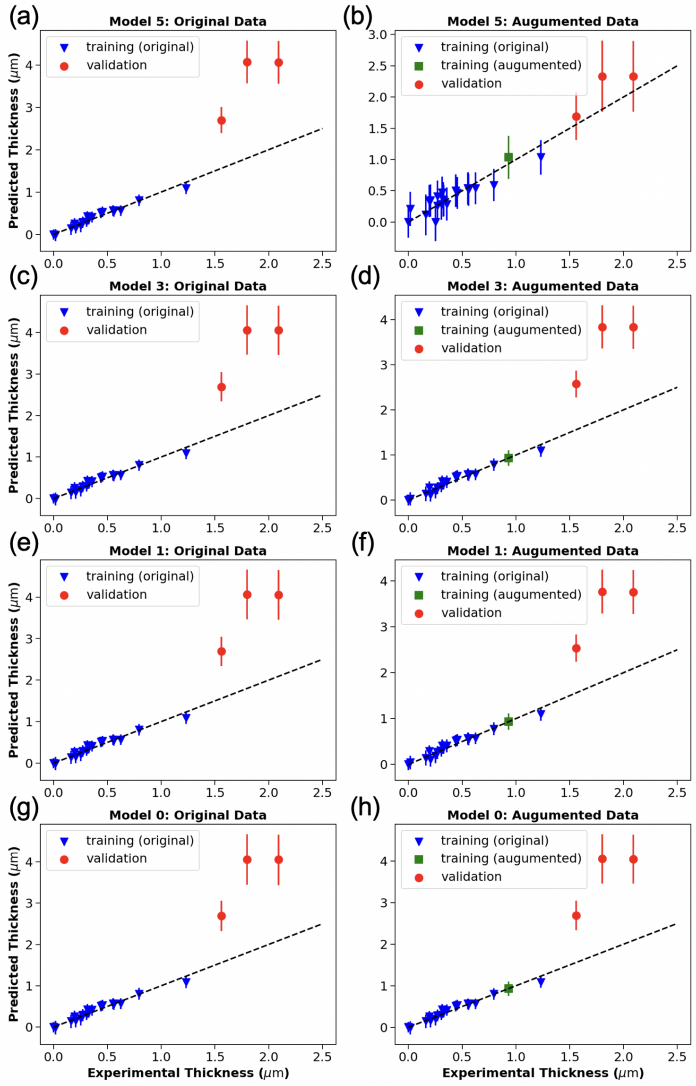

Figure 5.6. This parity plot helps validate the models 0, 1, 3, and 5. The models in the first column are trained with original data (blue triangle), runs 1 to 22; models in the second column are trained with original data (blue triangles) and augmented data (green square), run 23. The red dots show the validation experiments, runs 24 to 26. The error bars are 95% PI which are calculated via Eq. (5.16).

5.3.3.1 Optimal model identification

TABLE 5.5

COMPARISON OF PARAMETER CHANGE IN MODEL 5 BETWEEN TRAINED WITH ORIGINAL DATA (RUNS 1 TO 22) AND WITH ADDING AUGMENTED DATA (RUNS 1 TO 23).

Parameters	β_{00} ($cm^3 \cdot min^{-1}$)	β_{10} dimensionless	β_{20} ($min \cdot cm^{-3}$)	β_{01} dimensionless
Original (runs 1 to 22)	-0.0180	0.0290	-0.0255	0.0248
Augmented (runs 1 to 23)	-0.0116	0.0336	-0.0314	0.0116

The model validation process identifies the optimal model (arrow 10 in Figure 5.1). We selected the optimal model by using the train-test split and leave-one-out validation frameworks. Figure 5.6 shows the train-test split validation results for models 0, 1, 3, and 5. The left column uses original data, runs 1 to 22 in Table 5.2, as train data $\mathbb{D}_1$ to obtain optimal parameters $\hat{\boldsymbol{\beta}}$, and make predictions for the final three experiments, runs 24 to 26 in Table 5.2 as testing data $\mathbb{D}_2$. The right column follows a similar procedure but with run 23 in Table 5.2 included in the training data $\mathbb{D}_1$. A side-by-side comparison between two columns for different models shows the augmented data assists model 5 in successfully predicting the validation experiment, while models 0, 1, and 3 fail the test-split validation. This result reinforces our

conclusions that model 5 is the best model to describe the relationship between decision variables and thickness. The leave-one-out prediction results in Figure C.10 and Table C.6 in the SI further verifies this finding that model 0, 1, and 3 suffers from overfitting issue; in other words, model 5 has the best-generalized performance [42]. The physical interpretation of optimal model 5 is that there is no interaction between carrier gas flowrate Q_a and sheath flowrate Q_s. Given the most general model 0, Eq. (5.38), we have six parameters to estimate, including β_{00}, β_{10}, β_{20}, β_{01}, β_{11}, and β_{21}; the model 5 directly set $\beta_{11} = 0$ and $\beta_{21} = 0$, which means the interaction terms, $Q_a^2 Q_s$ and $Q_a Q_s$, can be ignored.

Table 5.5 summarizes the optimal estimated parameters $\hat{\boldsymbol{\beta}}$ comparison between original data and original data with augmented data for model 5. Two parameters, β_{00} and β_{10} have increased 35.6% and 15.9%, while another two parameters, β_{20} and β_{01} have decreased 23.1% and 53.2%, respectively. The highest variation of parameter, β_{01}, is the linear term of carrier gas flowrate Q_a. Meanwhile, Table C.7 in the SI shows the uncertainty of estimated parameters $\hat{\boldsymbol{\beta}}$ increases when run 23 is also used to regress model 5. This is because the FIM is calculated by using estimated variance, $\hat{\sigma^2}$, and local sensitivity, via Eq. (5.21); the augmented data increase estimated variance $\hat{\sigma^2}$ from 3.66×10^{-3} to 1.58×10^{-2}, which causes the increase of the estimated uncertainty of the parameters. This is expected because run 23 is dramatically different from the experimental conditions in runs 1 to 22 and thus increases the residuals. A similar comparison for other models is summarized in Table C.5 in the SI.

Figure C.11 in the SI visualizes the correlation between parameters in model 5. We observe high correlation, greater than 0.90, between parameters β_{00}, β_{10}, and β_{20}. The variables of the three parameters correspond to the Q_s with order 0, 1, 2, respectively, that can provide highly correlated relation between parameters [45]. The parameter β_{01} corresponds to variable Q_a that shows no correlation, less than

0.2, between the other three parameters.

5.3.3.2 Predictive inverse control of thickness

The selection and validation of model 5 enable the predictive inverse control of thickness based on sensitivity analysis (arrow 12 in Figure 5.1). Specifically, Figure 5.7 predicts how film thickness depends on the two most important decision variables — printing speed v_p and carrier gas flowrate Q_a — with voltage $U = 35$ V. Each row considers sheath flowrate rates Q_s, values of 50, 65, and 80 $\text{cm}^3 \cdot \text{min}^{-1}$. Similar plots for the voltage $U = 45$ V and $U = 20$ V are reported in Figure C.12 and C.13 of the SI. The comparison of these three figures indicates the voltage has no influence on the thickness once it reaches to proper operation condition (e.g., $U \geq 30$ V).

Figure 5.7(a), (c), and (e) show the prediction thickness as the function of printing speed v_p and carrier gas flowrate Q_a with other decision variables fixed. The thickness contours are linear because thickness h is inverse proportional to printing speed v_p and proportional to carrier gas flowrate Q_a, as shown in Eq. (5.38). Comparison between Figure 5.7(a), (c), and (e), shows thickness h decreases with the increase of sheath flowrate Q_s. From Figure 5.7(b), (d), and (f), we observe that the uncertainty increase with the larger sheath flowrate Q_s, which indicates the limitations of model predictive performance in the high sheath flowrate Q_s and low carrier gas flowrate Q_a region. Specifically, we notice that the lower carrier gas flowrate Q_a region in Figure 5.7(e) predicts a negative thickness which is physically infeasible.

One can fabricate the desired thickness film by selecting the corresponding point(s) from Figure 5.7. For example, for a target film thickness of $h = 0.50$ μm, we can choose any point near the contour line in Figure 5.7(c) that equals 0.50 μm. Then along this contour we, choose the conditions that minimize the prediction uncertainty using Figure 5.7(d), which is voltage $U = 35$ V, sheath flowrate $Q_s = 65$ $\text{cm}^3 \cdot \text{min}^{-1}$, carrier gas flowrate $Q_a = 25$ $\text{cm}^3 \cdot \text{min}^{-1}$, and printing speed $v_p = 3.75$ $\text{mm} \cdot \text{s}^{-1}$ for

this example.

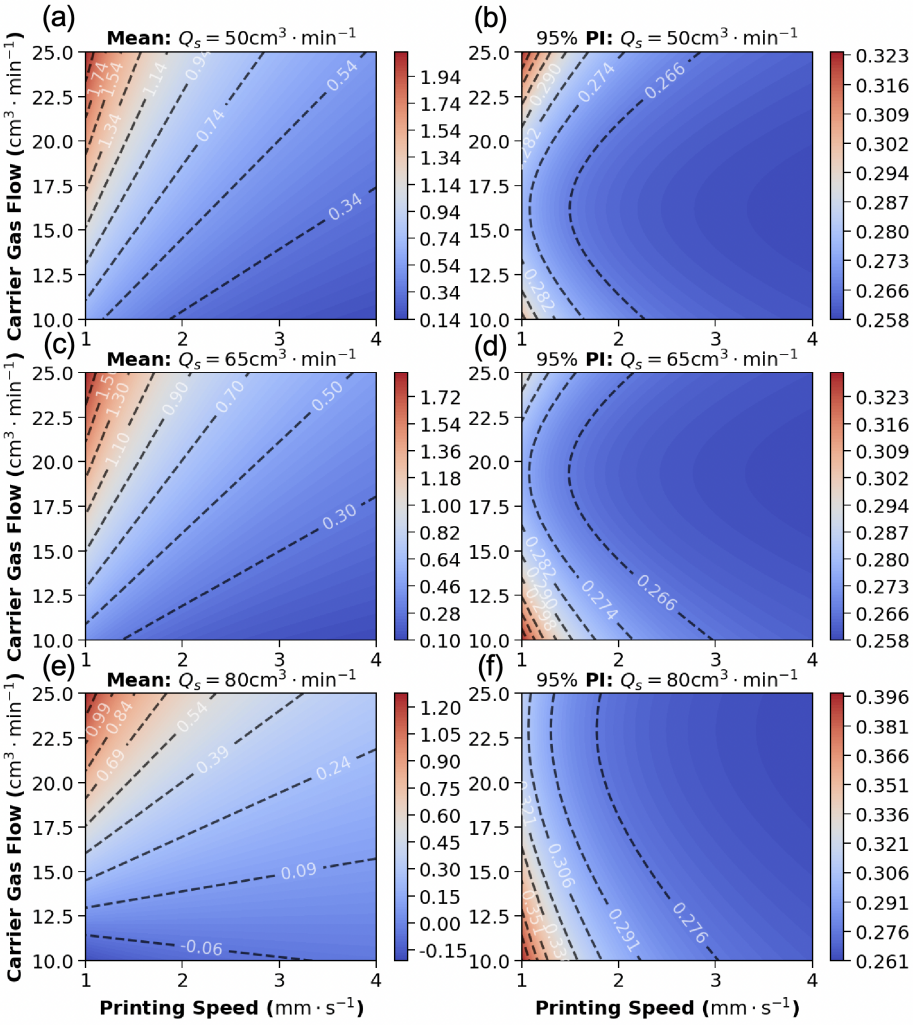

Figure 5.7. This figure shows the sensitivity analysis of inverse design with fixed $U = 35$ V using model 5. The first column is the mean prediction and the second column is the corresponding 95% PI.

5.3.4 Comparison between data-driven physical model and Gaussian process regression

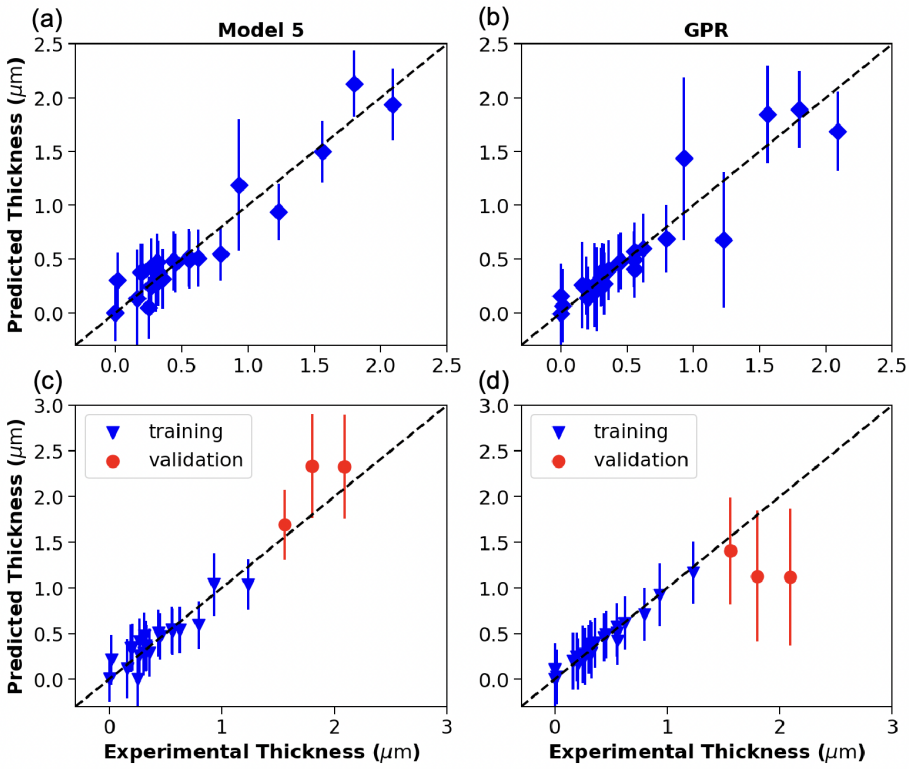

Figure 5.8. This figure shows the comparison between optimal model 5 and GPR performance in the validation process, the same setup as in Section 5.3.3; (a) and (b) are the leave-one-out cross-validation, while (c) and (d) are the train-test split validation. The hyperparameters $\boldsymbol{l}^*$ are estimated via Eq. (5.30).

So far in this chapter, we employed nonlinear regression to select and validate a parametric predictive model, which requires multiple modeling assumptions (e.g.,

linear, quadratic, sigmoid) about the parametric relationship between the decision (input) variables and measurement (output). An alternate data-driven approach is to use a nonparametric model, such as GPR, which replaces the parametric assumptions with alternate assumptions about how a kernel interpolates between the data [31, 38, 39]. We now make a side-by-side comparison between the optimal model 5 and GPR. The decision variables and measurement of the GPR are the same as the model 5, $\mathbf{x}_i = (U_i, v_{pi}, Q_{ai}, Q_{si})$ and h_i, respectively. The hyperparameters of the GPR model are trained by maximizing log-marginal likelihood shown in Eq. (5.30).

5.3.4.1 Comparison of model performance between model 5 and GPR with all experiments

Figure 5.8(a) and (b) show the leave-one-out prediction comparison between optimal model 5 and GPR. We can observe that model 5 makes more precise predictions in higher experimental thickness region (e.g., $h \geq 0.9$ μm), while GPR performs better for lower experimental thickness (e.g., $h < 0.9$ μm). Specifically, in lower experimental thickness, $h < 0.9$ μm, the LOO-MSE metric of GPR is 67.1% lower than model 5, while in the higher experimental thickness, $h \geq 0.9$ μm, the LOO-MSE metric of model 5 is 30.7% lower than GPR. Table C.8 reports the results of LOO-MSE for higher and lower thicknesses. To explain this, we need to introduce the key difference between the parametric and nonparametric models. The parametric model has a fixed equation structure and is often re-calibrated after each new experiment. However, the nonparametric model directly learns from data and each data sample is used to make predictions, as shown in Eq. (5.32a). The GPR prediction at a new experimental condition can be interpreted as summing the prior measurements using weights from the kernel function Eq. (5.29). When the samples are close in kernel distance, their prediction will behave similarly. In dataset $\mathbb{D}$, we have abundant data in the lower thickness region, which makes GPR perform better than model 5; while,

in the higher thickness region, the lack of data makes model 5 performs better. Table C.6 shows the comparison of LOO-MSE for model 5 and GPR. The result indicates the GPR performs slightly better, 16.7% in the LOO-MSE metric, than model 5, due to the better performance of GPR in the abundant data region, $h < 0.9$ μm.

Figure 5.8(c) and (d) show the train-test split prediction comparison between optimal model 5 and GPR. The setup is the same in Section 5.3.3. Comparison between the training data prediction shows the predicted thickness and experimental thickness of the GPR method are the same, which is the property of the nonparametric method that it has no bias [38]. In GPR, the predicted thickness of validation samples is consistently smaller than the experimental thickness, which can be explained via Eq. (5.32a), as the weighted sum can never be higher than the maximum thickness of training samples.

Figure 5.9(a) and (c) show the mean prediction comparison between model 5 and GPR. In the regions that have abundant samples (black star), we observe similar trends in equipotential surface across model 5 and GPR. However, in scarce data region (e.g., $v_p \geq 2.5$ mm $\cdot$ s^{-1}), the equipotential surface has flattened. With sufficient data, the GPR will converge to parametric model 5 [46], and even perform better, because the assumptions of the parametric model may become invalid in regions near bounds. Figure 5.9(b) and (d) compare the 95% PI between model 5 and GPR. The uncertainty in the data abundant region, black stars, is small, and the data-scarce region, upper right, is larger. This can be explained by Eq. (5.32b) that the first term can be regarded as prior uncertainty from new inputs $\mathbf{X}^*$, and the second term can be regarded as the uncertainty induced by training data $\mathbf{X}$; the subtraction shows the uncertainty of new input $\mathbf{\Sigma}_*(\mathbf{X}_*)$ is reduced by providing training data $\mathbf{X}$, and the more training data, the less the uncertainty of new input $\mathbf{\Sigma}_*(\mathbf{X}_*)$. The comparison suggests the parametric model is more data efficient because it provides physically realistic estimates across the entire experimental domain using a small dataset.

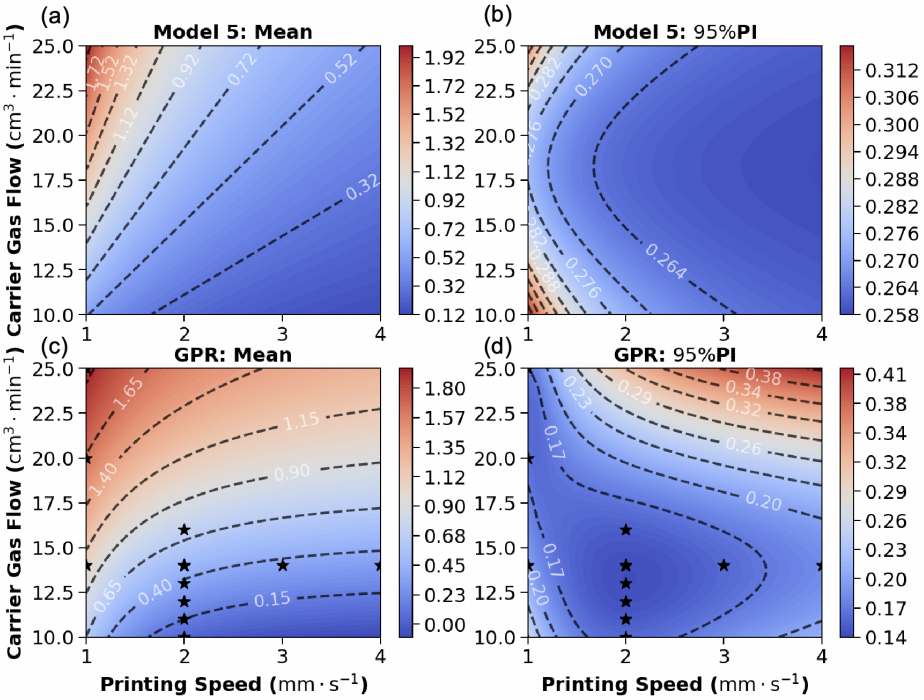

Figure 5.9. This figure shows the heatmap comparison of model 5 and GPR with fixed $U = 35$ V and $Q_s = 60$ cm$^3 \cdot$ min^{-1}. The first column is the mean prediction and the second column is the corresponding 95% PI calculated via Eq. (5.16). The black stars, in (c) and (d), are the runs 1 to 26, from Table 5.2, projected into 2D space.

5.3.4.2 Comparison of data efficiency between model 5 and GPR with experiments selected by D-optimality

To compare the data efficiency between parametric and nonparametric models, we down-select five experiments with maximum $\log_{10}$ D-optimality metric. Specifically, Algorithm 1 iterates over all possible combinations of experiments to form datasets containing 1 to 26 samples. This analysis is conducted using the parameter values reported in Table 5.2. From Figure 5.10, we observe the D-optimality metric is near zero for fewer than 4 experiments. This makes sense as a 4-parameter model requires at least 4 experiments to be identifiable (otherwise $\nabla_{\boldsymbol{\beta}}\mathbf{f}(\boldsymbol{\beta}, \mathbf{x_i})^T \nabla_{\boldsymbol{\beta}}\mathbf{f}(\boldsymbol{\beta}, \mathbf{x_i})$ cannot

be full rank). As expected, the $\log_{10}$ D-optimality demonstrates diminishing returns. Recall each point on the blue line corresponds to the best combination of experiments for a given dataset size and not all experiments in Table 5.2 are equally informative. The most informative experiments are the first selected by Algorithm 1. Based on Figure 5.10, we select the five experiments (runs 1, 18, 23, 24, and 25) as training data, and use the rest as validation data, to further make a comparison between model 5 and GPR.

Algorithm 1 Calculate the maximum $\log_{10}$ D-optimality as a function of the number of experiments

1: **Given**: Total number of experiments n, number of desired experiments k, estimated parameters $\hat{\boldsymbol{\beta}}$, estimated uncertainty $\hat{\sigma}$, dataset $\mathbb{D}$
2: Calculate the FIM for each experiment, $\mathbf{M}_i = \mathbf{M}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i, \hat{\sigma})$, $\forall i \in \{1, \cdots, n\}$ with $\boldsymbol{\beta}$ and $\hat{\sigma}$ fixed
3: **for** subset π_j in all combinations of n choose k experiments **do**
4: Calculate $\mathbf{M}^j = \sum_{i \in \pi_j} \mathbf{M}_i$
5: Calculate $D^j = \log_{10}\det(\mathbf{M}^j)$
6: **end for**
7: **Return**: $D = \max D^j$, $\pi_j^* = \arg\max_j D^j$

Figure 5.11 shows the train-test split prediction comparison between optimal model 5 and GPR using only five experiments as training data and the other 21 experiments as testing data described above. Comparing Figure 5.11 and Figure 5.8(c) and (d), we notice that small data training, five experiments, leads to larger error bars in both model 5 and GPR. In Figure 5.11, the test data MSE reported by model 5 is 0.035, while the same metric reported by GPR is 0.092. Thus, the GPR has a 62% larger MSE than model 5. Side-by-side comparison between Figure 5.11(a) and (b) shows the training data of GPR are exactly located in the parity line as expected; while, the training data of model 5 is around the parity line, in-

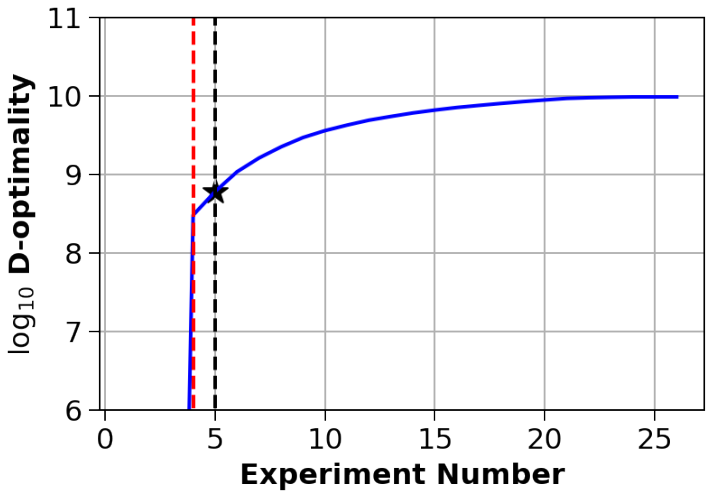

Figure 5.10. This figure shows the largest possible $\log_{10}$ D-optimality metric as a function of the number of experiments for model 5. The red vertical dashed line at $n = 4$ marks the number of parameters in model 5. The black vertical dashed line marks the smallest number of experiment suitables to estimate the model. The black star at $n = 5$ marks the number of experiments chosen to compare model 5 and GPR.

dicating the trained model can well-explained the variation of training data. The validation data of model 5 has uniform uncertainty, while the same data in GPR has a larger variation in prediction uncertainty. This observation can also be explained by that GPR provides relatively larger uncertainty in regions without data and the parametric model can provide uniform uncertainty across regions.

Figure 5.12 shows the mean and uncertainty prediction comparison between model 5 and GPR with limited training data (five experiments). The settings are the same as Figure 5.9, which uses all of the available data (26 experiments). The comparison between 5.12(a)(b) and (c)(d), we observe that the mean and uncertainty prediction of the parametric model can provide a consistent contour shape. In contrast, the prediction and uncertainty contours of the nonparametric model are largely determined by the distribution of samples. With only five experiments, the parametric

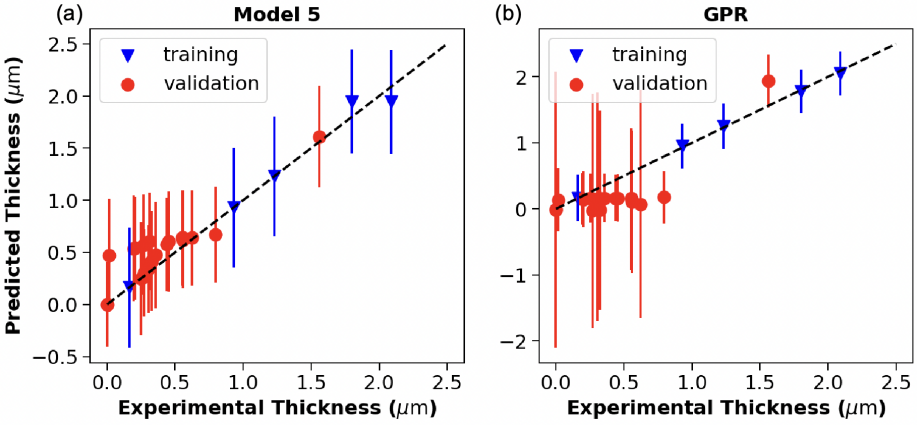

Figure 5.11. This figure shows the comparison of train-test split prediction between model 5 and GPR. The training data, blue triangles, are run 1, 18, 23, 24, and 25 from Table 5.2; the validation data, red dots, are the rest 21 experiments. The hyperparameters $\boldsymbol{l}^*$ are estimated via Eq. (5.30).

model can provide a reasonable prediction with relatively small uncertainty in broad regions.

Next, we focus on the comparison between using five experiments and all experiments as training data. First, in comparison between Figure 5.9(c) and Figure 5.12(c), we observe that the prediction mean of the latter one, trained with five experiments, performed poorly in regions with no data. Second, by comparing Figure 5.9(b)(d) and Figure 5.12(b)(d), we observe the consistently higher uncertainty with only five experiments. The uncertainty is relatively small in regions that have samples and relatively large in regions without data. These can both be explained by that the nonparametric model directly learns from data and cannot provide confident prediction in regions without data nearby.

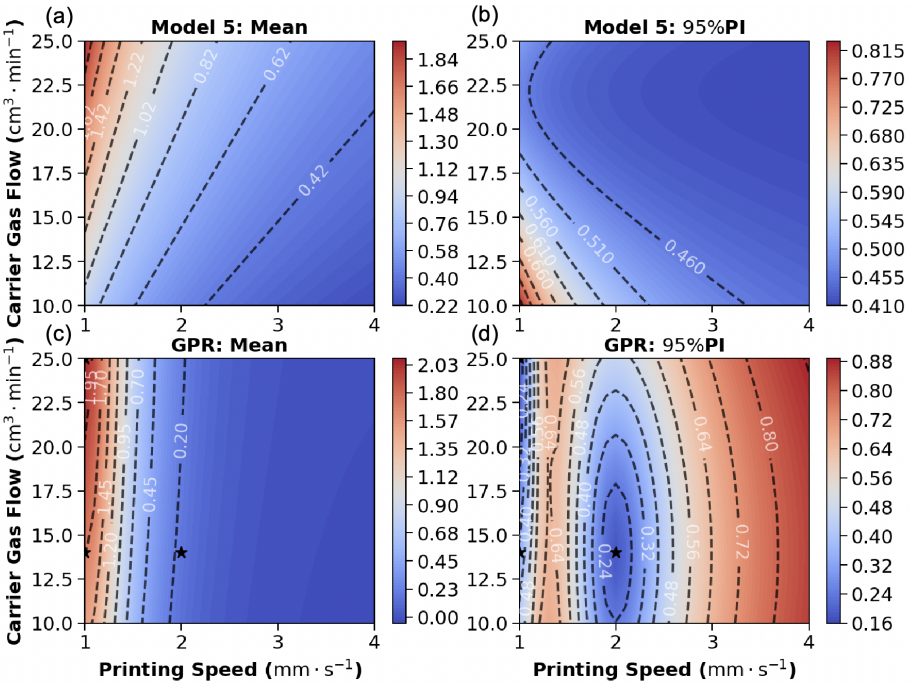

Figure 5.12. This figure shows the heatmap comparison of model 5 and GPR using five experiments, run 1, 18, 23, 24, and 25 from Table 5.2, with fixed $U = 35$ V and $Q_s = 60$ cm$^3 \cdot$ min^{-1}. The first column is the mean prediction and the second column is the corresponding 95% PI calculated via Eq. (5.16). The black stars, in (c) and (d), are run 1, 18, 23, 24, and 25, projected into 2D space.

5.4 Conclusions

In this work, we apply powerful data science tools such as model-based design of experiments (MBDoE) to facilitate precise control of printed film thickness in AJP processes for the first time. We start with an otherwise disregarded dataset containing 22 experiments from AJP apparatus commissioning and other projects. Using these preliminary data, a library of eight candidate physics-informed models is proposed and down-selected to four models using AIC. MBDoE identifies a single optimal experiment to improve the prediction accuracy of these models. Four additional

experiments are conducted, and as predicted the experimental conditions informed by MBDoE are more informative than experiments selected based solely on expert intuition. Finally, the single best model identified via cross-validation is used to identify strategies to manipulate voltage, carrier gas flowrate, sheath flowrate, and printing speed to achieve the desired film thickness with minimal model prediction uncertainty. Moreover, retrospective analysis using the determinant of the Fisher information matrix determines that only five optimized experiments are needed to train the identified physics-informed nonlinear model. This suggests that only a handful of experiments are needed to translate the proposed workflow to AJP with other material systems.

To benchmark the effectiveness of the approach, we train black-box nonparametric models using the same data. As expected, we find the GPR model performs best in data-rich regions (e.g., thin films for this dataset). In contrast, the physics-informed nonlinear parametric model outperforms GPR in regions with limited data. This trend is especially highlighted when considering the minimal dataset of only five optimal experiments. These results are expected and consistent with trends in recent physics-informed ML literature. Whenever possible, it is best to incorporate scientific and phenomenological information into mathematical models to reduce the data required for training. This chapter provides an additional case study to reinforce this observation.

As future work, we expect the identified relationship between decision variables including voltage, carrier gas flowrate, sheath flowrate and printing speed, and thickness can be applied to other AJP material systems such as AgSe [47] or BiSbTe [48] for thermoelectric materials and devices. We optimistically predict adaptations to new material systems will require as few as five to ten optimally designed experiments for initial modeling, but with two important changes: (1) the bounds of decision variables should updated to consider changes in the ink viscosity; and (2) the regression

coefficient will likely be different for each system.

Another direction for future work is increasing the model accuracy near operational bounds. For example, given the best model identified, we observe the assumed linear and quadratic relationships may not hold near the operational bounds. We hypothesize this subtle model-form error explains the superior performance of black-box GPR with extensive data. In the future, hybrid models [49, 50] can be used to correct this small systematic bias in thickness predictions. Moreover, the identified physics-informed models and proposed data science framework can be extended to optimize more steps and decisions in additive manufacturing processes such as ink composition for AJP.

5.5 Acknowledgments

This work was supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Advanced Manufacturing Office Award Number DE-EE0009103. We also thankfully acknowledge additional support from the U.S. National Science Foundation under awards CBET-1941596 and CMMI-1747685.

5.6 References

1. Minxiang Zeng and Yanliang Zhang. Colloidal nanoparticle inks for printing functional devices: Emerging trends and future prospects. *Journal of Materials Chemistry A*, 7(41):23301–23336, 2019.

2. Mortaza Saeidi-Javash, Yipu Du, Minxiang Zeng, Brian C Wyatt, Bowen Zhang, Nicholas Kempf, Babak Anasori, and Yanliang Zhang. All-printed mxene–graphene nanosheet-based bimodal sensors for simultaneous strain and temperature sensing. *ACS Applied Electronic Materials*, 3(5):2341–2348, 2021.

3. Minxiang Zeng, Wenzheng Kuang, Irfan Khan, Dali Huang, Yipu Du, Mortaza Saeidi-Javash, Lecheng Zhang, Zhengdong Cheng, Anthony J Hoffman, and Yanliang Zhang. Colloidal nanosurfactants for 3D conformal printing of 2D van der Waals materials. *Advanced Materials*, 32(39):2003081, 2020.

4. Yipu Du, Ruoxing Wang, Minxiang Zeng, Shujia Xu, Mortaza Saeidi-Javash, Wenzhuo Wu, and Yanliang Zhang. Hybrid printing of wearable piezoelectric sensors. *Nano Energy*, 90:106522, 2021.

5. Minxiang Zeng, Hongyao Xie, Mortaza Saeidi-Javash, ANM Tanvir, Yipu Du, Jiahao Chen, Mercouri G Kanatzidis, and Yanliang Zhang. Scalable nanomanufacturing of chalcogenide inks: a case study on thermoelectric V–VI nanoplates. *Journal of Materials Chemistry A*, 9(39):22555–22562, 2021.

6. Minxiang Zeng, Duncan Zavanelli, Jiahao Chen, Mortaza Saeidi-Javash, Yipu Du, Saniya LeBlanc, G Jeffrey Snyder, and Yanliang Zhang. Printing thermoelectric inks toward next-generation energy and thermal devices. *Chemical Society Reviews*, 51:485–512, 2021.

7. Ankit Mahajan, C Daniel Frisbie, and Lorraine F Francis. Optimization of aerosol jet printing for high-resolution, high-aspect ratio silver lines. *ACS Applied Materials & Interfaces*, 5(11):4856–4864, 2013.

8. Rebecca R Tafoya, Adam W Cook, Bryan Kaehr, Julia R Downing, Mark C Hersam, and Ethan B Secor. Real-time optical process monitoring for structure and property control of aerosol jet printed functional materials. *Advanced Materials Technologies*, 5(12):2000781, 2020.

9. Mortaza Saeidi-Javash, Ke Wang, Minxiang Zeng, Tengfei Luo, Alexander Dowling, and Yanliang Zhang. Machine learning-assisted ultrafast flash sintering of high-performance and flexible silver-selenide thermoelectric devices. *Energy & Environmental Science*, 15:5093–5104, 2022.

10. Haining Zhang and Seung Ki Moon. Reviews on machine learning approaches for process optimization in noncontact direct ink writing. *ACS Applied Materials & Interfaces*, 13(45):53323–53345, 2021.

11. J. Q. Feng. A computational study of particle deposition patterns from a circular laminar jet. *Journal of Applied Fluid Mechanics*, 10(4):1001–1012, 2017.

12. Sebastian Binder, Markus Glatthaar, and Edda Rädlein. Analytical investigation of aerosol jet printing. *Aerosol Science and Technology*, 48(9):924–929, 2014.

13. Kan Wang, Yung-Hang Chang, Chuck Zhang, and Ben Wang. Evaluation of quality of printed strain sensors for composite structural health monitoring applications. In *SAMPE Fall Technical Conference, Wichita, KS, Oct*, pages 21–24, 2013.

14. Haining Zhang, Seung Ki Moon, and Teck Hui Ngo. Hybrid machine learning method to determine the optimal operating process window in aerosol jet 3d printing. *ACS Applied Materials & Interfaces*, 11(19):17994–18003, 2019.

15. Haining Zhang, Joon Phil Choi, Seung Ki Moon, and Teck Hui Ngo. A multi-objective optimization framework for aerosol jet customized line width printing via small data set and prediction uncertainty. *Journal of Materials Processing Technology*, 285:116779, 2020.

16. George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

17. Jin-Long Wu, Heng Xiao, and Eric Paterson. Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. *Physical Review Fluids*, 3(7):074602, 2018.

18. K Kashinath, M Mustafa, A Albert, JL Wu, C Jiang, S Esmaeilzadeh, K Azizzadenesheli, R Wang, A Chattopadhyay, A Singh, et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, 379(2194):20200093, 2021.

19. Kanishka Ghosh, Sergio Vernuccio, and Alexander W Dowling. Nonlinear reactor design optimization with embedded microkinetic model information. *Frontiers in Chemical Engineering*, 4:898685, 2022.

20. Jonathan A Ouimet, Xinhong Liu, David J Brown, Elvis A Eugene, Tylar Popps, Zachary W Muetzel, Alexander W Dowling, and William A Phillip. Data: Diafiltration apparatus for high-throughput analysis. *Journal of Membrane Science*, 641:119743, 2022.

21. Elvis A Eugene, William A Phillip, and Alexander W Dowling. Data science-enabled molecular-to-systems engineering for sustainable water treatment. *Current Opinion in Chemical Engineering*, 26:122–130, 2019.

22. Abhishek Kumar, Kuljeet Singh, and Ranjan Das. Response surface based experimental analysis and thermal resistance model of a thermoelectric power generation system. *Applied Thermal Engineering*, 159:113935, 2019.

23. A Kanatzia, CH Papageorgiou, CH Lioutas, and TH Kyratsi. Design of ball-milling experiments on Bi2Te3 thermoelectric material. *Journal of Electronic Materials*, 42(7):1652–1660, 2013.

24. Gaia Franceschini and Sandro Macchietto. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63(19): 4846–4872, 2008.

25. Jialu Wang and Alexander W Dowling. Pyomo.DOE: An open-source package for model-based design of experiments in Python. *AIChE Journal*, page e17813, 2022.

26. Gary E Blau, Joseph F Pekny, Vishal A Varma, and Paul R Bunch. Managing a portfolio of interdependent new product candidates in the pharmaceutical industry. *Journal of Product Innovation Management*, 21(4):227–245, 2004.

27. Bing H Chen and Steven P Asprey. On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations. *Industrial & Engineering Chemistry Research*, 42(7):1379–1390, 2003.

28. SP Asprey and S Macchietto. Statistical tools for optimal dynamic model building. *Computers & Chemical Engineering*, 24(2-7):1261–1267, 2000.

29. Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272, 2020.

30. Harry H Ku et al. Notes on the use of propagation of error formulas. *Journal of Research of the National Bureau of Standards*, 70(4):263–273, 1966.

31. Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*, volume 112. Springer, 2013. ISBN 978-1461471370.

32. Yonathan Bard. *Nonlinear Parameter Estimation*. Academic Press, 1974. ISBN 978-0120782505.

33. Yuanlin Gu, Hua-Liang Wei, and Michael M Balikhin. Nonlinear predictive model selection and model averaging using information criteria. *Systems Science & Control Engineering*, 6(1):319–328, 2018.

34. Ronald Aylmer Fisher. Design of experiments. *British Medical Journal*, 1(3923): 554, 1936.

35. Gilbert Strang, Gilbert Strang, Gilbert Strang, and Gilbert Strang. *Introduction to Linear Algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA, 1993. ISBN 978-0961408855.

36. Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

37. Oana-Teodora Chis, Alejandro F Villaverde, Julio R Banga, and Eva Balsa-Canto. On the relationship between sloppiness and identifiability. *Mathematical Biosciences*, 282:147–161, 2016.

38. Larry Wasserman. *All of Nonparametric Statistics*. Springer Science & Business Media, 2006. ISBN 978-0387251455.

39. Christopher KI Williams and Carl Edward Rasmussen. *Gaussian Processes for Machine Learning*, volume 2. MIT press Cambridge, MA, 2006. ISBN 978-0262182539.

40. Ke Wang and Alexander W Dowling. Bayesian optimization for chemical products and functional materials. *Current Opinion in Chemical Engineering*, 36: 100728, 2022.

41. Ke Wang, Mortaza Saeidi-Javash, Minxiang Zeng, Zeyu Liu, Yanliang Zhang, Tengfei Luo, and Alexander W Dowling. Gaussian process regression machine learning models for photonic sintering. In *Computer Aided Chemical Engineering*, volume 49, pages 1819–1824. Elsevier, 2022.

42. Christopher M Bishop and Nasser M Nasrabadi. *Pattern Recognition and Machine Learning*, volume 4. Springer, 2006. ISBN 978-0387310732.

43. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

44. Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.

45. Paul D Allison. *Multiple Regression: A Primer*. Pine Forge Press, 1999.

46. Aretha L Teckentrup. Convergence of Gaussian process regression with estimated hyper-parameters and applications in Bayesian inverse problems. *SIAM/ASA Journal on Uncertainty Quantification*, 8(4):1310–1337, 2020.

47. Nazli Turan, Mortaza Saeidi-Javash, Jiahao Chen, Minxiang Zeng, Yanliang Zhang, and David B Go. Atmospheric pressure and ambient temperature plasma jet sintering of aerosol jet printed silver nanoparticles. *ACS Applied Materials & Interfaces*, 13(39):47244–47251, 2021.

48. Courtney Hollar, Zhaoyang Lin, Madhusudan Kongara, Tony Varghese, Chinnathambi Karthik, Jesse Schimpf, Josh Eixenberger, Paul H Davis, Yaqiao Wu, Xiangfeng Duan, et al. High-performance flexible bismuth telluride thin film from solution processed colloidal nanoplates. *Advanced Materials Technologies*, 5(11): 2000600, 2020.

49. Elvis A Eugene, Xian Gao, and Alexander W Dowling. Learning and optimization with bayesian hybrid models. In *2020 American Control Conference (ACC)*, pages 3997–4002. IEEE, 2020.

50. Jialu Wang, Elvis A Eugene, and Alexander W Dowling. Scalable stochastic programming with bayesian hybrid models. In *Computer Aided Chemical Engineering*, volume 49, pages 1309–1314. Elsevier, 2022.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

6.1 Conclusions

The integration of machine learning (ML) and operation research (OR) into experimental workflow represents a novel paradigm for advancing thermoelectric materials and additive manufacturing. This thesis demonstrates how strategic collaboration between domain expertise, data-driven algorithms, and active learning frameworks can overcome the inefficiencies of traditional Edisonian search. Here, I synthesize and summarize the insights for best practices of this framework in three aspects.

Surrogate modeling, which uses machine learning to map relationships between process variables and outcomes, is the key to efficient experimental optimization. In Chapter 2, Gaussian Process Regression (GPR) stands out as a non-parametric surrogate model that excels in data-scarce regimes by interpolating sparse datasets and quantifying uncertainty for known and unknown samples. In Chapter 3, a critical lesson from this work is the robustness of GPR hyperparameter training: two distinct optimization methods (log-marginal likelihood and leave-one-out cross-validation) converged to identical hyperparameter values, indicating that the choice of training framework is inconsequential. These hyperparameters also provided interpretable insights, such as feature importance rankings and aligned model predictions with domain knowledge. However, parametric models like non-linear regression remain valuable in the manufacturing domain. In Chapter 5, we showed that physics-informed nonlinear regression, combined with leave-one-out cross-validation and uncertainty

quantification, achieves high predictive accuracy with minimal data, a key advantage for resource-constrained applications. Notably, parametric models outperformed non-parametric models in data efficiency but required careful incorporation of first-principles knowledge to offset their reliance on predefined mathematical structures. This underscores a broader best practice: *Choosing parametric and non-parametric surrogate models should consider data availability and prior mechanistic understanding.* Non-parametric models like GPR are ideal for exploratory, high-uncertainty scenarios, while parametric models succeed when domain knowledge can guide their modeling structure, enabling precise predictions with fewer experiments.

Optimization frameworks leverage pre-trained surrogate models and domain-specific objective functions to iteratively identify high-value experiments, accelerating the convergence to optimal experiment design. Central to this approach is the adaption of prior observations into the optimization loop, which strategically balances exploration (regions with high uncertainty) and exploitation (known high-performance regions). In Chapter 3, Bayesian optimization (BO) demonstrated this strategy, achieving optimal sintering conditions for a real-world flash sintering process within 32 experiments in high-dimensional and noisy environments. However, optimization efficiency can be further improved by incorporating constraints. In Chapter 4, we integrated black-box constraints into BO to narrow the search space. However, a critical lesson learned: Constraints should not be incorporated early if their mechanisms are unclear, as they may overly restrict the search space. Beyond single-objective optimization, manufacturing often demands trade-offs between competing objectives. For such scenarios, multi-objective optimization with Pareto frontiers provides a principled way to visualize and navigate progress. In Chapter 4, Pareto-front identified knee points, regions where marginal gains in one objective (e.g., electrical conductivity) incur disproportionate losses in another (e.g., substrate temperature), enabling data-driven prioritization of candidate conditions. This underscores a broader best

practice: *Optimization frameworks should leverage multi-objective strategies to resolve competing priorities and should avoid over-constraining the search space in the early stage of the experiment.*

While surrogate models and optimization algorithms provide sufficient tools for solving manufacturing optimization challenges, expert intuition remains indispensable for efficiently navigating complex experimental search space. This thesis demonstrates how collaboration with experts at varying experience levels can be tailored to maximize the value of human-machine synergy. In Chapter 3, we collaborated with highly experienced experts to refine the BO workflow: BO constrained the search space to regions of high Expected Improvement (EI), while experts down-selected experiments within these regions based on practical feasibility and domain-specific heuristics. Conversely, in Chapter 4, collaboration with less-experienced experts required a more iterative, ML-guided strategy. Here, optimization frameworks shaped the experimental roadmap from low-dimension to high-dimension search space, incrementally incorporating expert feedback to align outcomes with broader goals (e.g., identifying knee points in Pareto frontiers). This underscores a broader best practice: *The role of expert intuition should adapt to the practitioner's experience level.* For seasoned experts, leverage their deep domain knowledge to bias or validate ML-driven recommendations; for novices, use ML-driven optimization to conduct decision-making, iteratively refining strategies as their understanding grows. This dynamic integration ensures that human insight improves the efficiency of data-driven experimentation.

6.2 Recommendations for future work

Opportunity 1: Broadening to Generalized Manufacturing Systems. In Chapters 3 to 5, we demonstrate the integration of machine learning, operations research, and domain expertise to accelerate the identification of optimal experimental designs

in thermoelectric materials and additive manufacturing. The opportunity lies in expanding this well-developed framework to broader manufacturing systems, such as carbon capture[1], water treatment[2], and biomanufacturing[3]. These fields share common challenges, including multi-parameter interaction, high-dimensional nonlinear process responses, and rigid resource constraints (time, cost, data). The proven capability of the framework, efficient optimization under data scarcity, quantification of multi-objective trade-offs, and human-machine collaborative decision-making can seamlessly transition to these complex manufacturing fields.

Opportunity 2: Large language model (LLM). In Chapters 3 and 4, we showed how the predictive accuracy of GPR can be improved in data-scarce scenarios by incorporating physical knowledge into the machine learning model. However, the process of extracting this physical knowledge from the literature requires advanced expertise in mathematics and physics. This presents an opportunity to leverage pre-trained LLMs (e.g., Llama[4], GPT[5]) by designing an LLM-based agent[6] to automatically acquire the necessary physical knowledge and build the hybrid model. There is also a growing area of research focusing on the development of digital twins powered by LLMs, which exploit few-shot capabilities to generate warm-up datasets with a small amount of initial experimental data.[7] For example, by fine-tuning a pre-trained LLM [8, 9] on a small dataset, generating large synthetic datasets, and using the large dataset to train the subsequent machine learning model, it can achieve more accurate predictive modeling with less real-world experimental data.

Opportunity 3: Constraint Learning. In Chapter 5, we demonstrated that model-based design of experiments (MBDoE) can effectively determine optimal experiment designs for model selection when using physics-informed machine learning models with explicit mathematical formulas. However, optimization becomes more complex when constraints are learned from black-box machine learning models (e.g., random forests or neural networks) and are implicitly defined. In these cases, the

solver must obtain predictions from the machine learning model each time it needs to calculate gradients, making the process computationally intensive. There is an opportunity to develop an MBDoE framework that integrates constraint learning[10, 11]. In this framework, implicit constraints are learned from data using machine learning models. Then, the pre-trained models are embedded into the optimization framework by reformulating it as a mixed-integer problem.[12] This approach offers two key advantages: first, it can better adapt to the mixed-integer solver, making it faster to solve; second, it can extend to dynamic optimization where current decisions are influenced by previous decisions in sequential experiments.[13]

Opportunity 4: Autonomous experimental design. Throughout Chapters 3 to 5, the active learning framework provides optimal experiment designs to experimentalists who manually conduct new experiments. However, the manufacturing process is highly uncertain and time-consuming, with many factors that can influence the accuracy of the final results. This presents an opportunity for autonomous experimentation built upon active learning [14, 15] and multi-modal large language models [16, 17], where a robotic system is developed, and the agent robot/device automatically designs optimal experiments, conducts them, measures results, and iterates until the desired product requirements are reached.

6.3 References

1. Ahmed I Osman, Mahmoud Hefny, MIA Abdel Maksoud, Ahmed M Elgarahy, and David W Rooney. Recent advances in carbon capture storage and utilisation technologies: a review. *Environmental Chemistry Letters*, 19(2):797–849, 2021.
2. Saja M Alardhi, Nisreen S Ali, Noori M Cata Saady, Sohrab Zendehboudi, Issam K Salih, Jamal M Alrubaye, and Talib M Albayati. Separation techniques in different configurations of hybrid systems via synergetic adsorption and membrane processes for water treatment: a review. *Journal of Industrial and Engineering Chemistry*, 130:91–104, 2024.
3. Kamini Pandey, Muskan Pandey, Vinay Kumar, Upasana Aggarwal, and Barkha

Singhal. Bioprocessing 4.0 in biomanufacturing: paving the way for sustainable bioeconomy. *Systems Microbiology and Biomanufacturing*, 4(2):407–424, 2024.

4. Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

5. Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

6. T Guo, X Chen, Y Wang, R Chang, S Pei, NV Chawla, O Wiest, and X Zhang. Large language model based multi-agents: A survey of progress and challenges. In *33rd International Joint Conference on Artificial Intelligence (IJCAI 2024)*. IJCAI; Cornell arxiv, 2024.

7. Yicheng Sun, Qi Zhang, Jinsong Bao, Yuqian Lu, and Shimin Liu. Empowering digital twins with large language models for global temporal feature learning. *Journal of Manufacturing Systems*, 74:83–99, 2024.

8. Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. Chronos: Learning the language of time series. *Machine learning*, 1(2):3.

9. Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. Moment: A family of open time-series foundation models. In *Forty-first International Conference on Machine Learning*.

10. Donato Maragno, Holly Wiberg, Dimitris Bertsimas, Ş İlker Birbil, Dick den Hertog, and Adejuyigbe O Fajemisin. Mixed-integer optimization with constraint learning. *Operations Research*, 2023.

11. Adejuyigbe O Fajemisin, Donato Maragno, and Dick den Hertog. Optimization with constraint learning: A framework and survey. *European Journal of Operational Research*, 314(1):1–14, 2024.

12. Donato Maragno, Tabea Elina Röber, and Ilker Birbil. Counterfactual explanations using optimization eith constraint learning. In *OPT 2022: Optimization for Machine Learning (NeurIPS 2022 Workshop)*.

13. Risheng Liu, Jiaxin Gao, Xuan Liu, and Xin Fan. Learning with constraint learning: New perspective, solution strategy and various applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

14. Zhichu Ren, Zekun Ren, Zhen Zhang, Tonio Buonassisi, and Ju Li. Autonomous experiments using active learning and ai. *Nature Reviews Materials*, 8(9):563–564, 2023.

15. Cliff Lovell, Gareth Jones, Steve R Gunn, and Klaus-Peter Zauner. Autonomous experimentation: Active learning for enzyme response characterisation. In *Active Learning and Experimental Design workshop In conjunction with AISTATS 2010*, pages 141–155. JMLR Workshop and Conference Proceedings, 2011.

16. Can Cui, Yunsheng Ma, Xu Cao, Wenqian Ye, Yang Zhou, Kaizhao Liang, Jintai Chen, Juanwu Lu, Zichong Yang, Kuei-Da Liao, et al. A survey on multimodal large language models for autonomous driving. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 958–979, 2024.

17. Zhiyuan Xu, Kun Wu, Junjie Wen, Jinming Li, Ning Liu, Zhengping Che, and Jian Tang. A survey on robotics with foundation models: toward embodied ai. *arXiv preprint arXiv:2402.02385*, 2024.