Information-driven experimental design in materials science

R. Aggarwal, M. J. Demkowicz, Y. M. Marzouk

Abstract

Optimal experimental design (OED) aims to maximize the value of experiments and the data they produce. OED ensures efficient allocation of limited resources, especially when numerous repeated experiments cannot be performed. This chapter presents a fully Bayesian and decision theoretic approach to OED—accounting for uncertainties in models, model parameters, and experimental outcomes, and allowing optimality to be defined according to a range of possible experimental goals. We demonstrate this approach on two illustrative problems in materials research.

The first example is a parameter inference problem. Its goal is to determine a substrate property from the behavior of a film deposited thereon. We design experiments to yield maximal information about the substrate property using only two measurements. The second example is a model selection problem. We design an experiment that optimally distinguishes between two models for helium trapping at interfaces. In both instances, we provide model-based justifications for why the selected experiments are optimal. Moreover, both examples illustrate the utility of reduced-order or surrogate models in optimal experimental design.

1 Introduction

Experiments are essential prerequisites of all scientific research. They are the basis for developing and refining mathematical models of physical reality. Experimental data are used to infer model parameters, to improve the accuracy of model-based predictions, to discriminate among competing models, to assess model validity, and to improve design and decision-making under uncertainty. Yet experimental observations can be difficult, time-consuming, and expensive to acquire. Maximizing the value of experimental observations—i.e., designing experiments to be *optimal* by some appropriate measure—is therefore a critical task. Experimental design encompasses questions of where and when to measure, which variables to interrogate, and what experimental conditions to employ.

Conventional experimental design methods, such as factorial and composite designs, are largely used as heuristics for exploring the relationship between input factors and response variables. By contrast, *optimal* experimental design uses a concrete hypothesis—expressed as a quantitative model—to guide the choice of experiments for a particular purpose, such as parameter inference, prediction, or model discrimination. Optimal design has seen extensive development for linear models (where the measured quantities depend linearly on the model parameters) endowed with Gaussian

distributions [5]. Extensions to nonlinear models are often based on linearization and Gaussian approximations [15, 36, 21], as analytical results are otherwise impractical or impossible to obtain. With advances in computational power, however, optimal experimental design for nonlinear systems can now be tackled directly using numerical simulation [65, 84, 93, 64, 89, 48, 49, 96].

This chapter will present an overview of model-based optimal experimental design, connecting this approach to illustrative applications in materials science—a field replete with potential applications for optimal experimentation. We will take a fully Bayesian and decision-theoretic approach. In this formulation, one first defines the *utility* of an experiment and then, taking into account uncertainties in both the parameter values and the observations, chooses experiments by maximizing an *expected utility*. We will define these utilities according to information theoretic considerations, reflecting the particular experimental goals at hand.

The evaluation and optimization of information theoretic design criteria, in particular those that invoke complex physics-based models, requires the synthesis of several computational tools. These include: (1) statistical estimators of expected information gain; (2) efficient optimization methods for stochastic or noisy objectives (since expected utilities are typically evaluated with Monte Carlo methods); and (3) reduced-order or surrogate models that can accelerate the estimation of information gain. For a simple film-substrate system, we will present an example of such a reduced-order model, derived from physical scaling principles and an "offline" set of detailed/full model simulations. This is but one example; reduced-order models constructed through a variety of techniques have practical use in a wide range of optimal experimental design applications [48].

The rest of this chapter is organized as follows. Section 2 will present the foundational tools of optimal experimental design, beginning with Bayesian inference and proceeding to discuss several information theoretic design criteria. It will also discuss the computational challenges presented by this formulation. Section 3 will illustrate the information-driven approach with two examples: optimal design for *parameter inference*, in the context of a film-substrate system; and optimal design for *model selection*, in the context of heterophase interfaces in layered metal composites. Section 4 will discuss open questions and topics of ongoing research.

2 The tools of optimal experimental design

We will formulate our experimental design criteria in a Bayesian setting. Bayesian statistics offers a foundation for inference from noisy, indirect, and incomplete data; a mechanism for incorporating multiple heterogeneous sources of information; and a complete assessment of uncertainty in parameters, models, and predictions. The Bayesian approach also provides natural links to decision theory, which we will exploit below.

2.1 Bayesian inference

The essence of the Bayesian paradigm is to describe uncertainty or lack of knowledge *probabilistically*. This idea applies to model parameters, to observations, and even to

competing models. For simplicity, we first describe the case of parameter inference.

Let $\theta \in \Theta \subseteq \mathbb{R}^n$ represent the parameters of a given model. We describe our state of knowledge about these parameters with a prior probability density $p(\theta)$. (For the remainder of this article, we assume that all parameter and data probability distributions have densities with respect to Lebesgue measure.) We would like to update our knowlege about θ by performing an experiment at conditions $\eta \in \mathcal{H} \subseteq \mathbb{R}^d$. η is therefore our vector of experimental design parameters. This experiment will yield observations $y \in \mathcal{Y} \subseteq \mathbb{R}^m$. The relationship between the model parameters, experimental conditions, and observations is captured by the likelihood function $p(y|\theta, \eta)$, i.e., the probability density of the observations given a *particular* choice of θ, η. The likelihood naturally incorporates a physical model of the experiment. For instance, one often has a computational model $G(\theta, \eta)$ that predicts the quantity being measured by a proposed experiment. This prediction may be imperfect, and is almost always corrupted by some observational errors. A simple likelihood then results from the additive model $y = G(\theta, \eta) + \epsilon$, where ϵ is a random variable representing measurement and model errors. If ϵ is Gaussian with mean zero and variance σ^2, and independent of θ and η, then we have the Gaussian likelihood $p(y|\theta, \eta) \sim N\left(G(\theta, \eta), \sigma^2\right)$. More complex likelihoods describe signal-dependent noise, or include more sophisticated representations of model error (e.g., the discrepancy models of [53]).

Putting these ingredients together via Bayes' rule, we obtain the posterior probability density $p(\theta|y, \eta)$ of the parameters:

$$p(\theta|y, \eta) = \frac{p(y|\theta, \eta)p(\theta)}{p(y|\eta)}, \tag{1}$$

where we have assumed (quite reasonably) that the prior knowledge on the parameters is independent of the experimental design. The posterior density describes the state of knowledge about the parameters θ *after* conditioning on the result of the experiment. The design criteria described below will formalize the intuitive idea of choosing values of η to make the posterior distribution of θ as "informed" as possible.

Many problems, whether in materials science or other domains, do not have parameter inference as an end goal. Rather than learning about parameters that appear in a single fixed model of interest, one may wish to collect data that help discriminate among competing models. For instance, different hypothesized physical mechanisms may lead to different models of a phenomenon. In this context, the Bayesian approach involves characterizing a posterior probability distribution over models. Let the model space $\mathcal{M}$ consist of an enumerable number of competing models M_i, $i \in \{1, 2, \ldots\}$. Let each model M_i be endowed with parameters $\theta_i \in \Theta_i \subseteq \mathbb{R}^{n_i}$. Then Bayes' rule writes the posterior probability of a model M_i as:

$$P(M_i|y, \eta) = \frac{p(y|M_i, \eta)P(M_i)}{p(y|\eta)}, \tag{2}$$

where the *marginal likelihood* of each model (i.e., $p(y|M_i, \eta)$ for the ith model) is obtained by averaging the likelihood over the prior distribution on the model's parameters:

$$p(y|M_i, \eta) = \int_{\Theta_i} p(y|\theta_i, \eta, M_i)p(\theta_i|M_i)d\theta_i. \tag{3}$$

Each model has its own parameters θ_i and its own prior $p(\theta_i|M_i)$. The marginal likelihood incorporates an automatic Occam's razor that penalizes unnecessary model complexity [66, 8]. The effective use of the posterior distribution over models $P(M_i|y,\eta)$ can then depend on the goals at hand. For instance, one may wish to know which model is best supported by the data; in this case, one simply selects the model with the highest posterior probability, thus performing Bayesian model selection. Alternatively, if the end goal is to make a prediction that accounts for model uncertainty, one can perform Bayesian model averaging [46] by taking a linear combination of predictions from each model, weighed according to the posterior model probabilities.

2.2 Information theoretic objectives

Following a decision theoretic approach, Lindley [63] suggests that an objective for experimental design should have the following general form:

$$U(\eta) = \int_{\mathcal{Y}} \int_{\Theta} u(\eta, y, \theta)\, p(\theta, y|\eta)\, d\theta\, dy, \tag{4}$$

where $u(\eta, y, \theta)$ is a *utility function* and $U(\eta)$ is the *expected* utility. The utility function u should be chosen to reflect the usefulness of an experiment at conditions η, given a particular value of the parameters θ and a particular outcome y. Since we do not know the precise value of θ and we cannot know the outcome of the experiment before it is performed, we obtain U by taking the expectation of u over the joint distribution of θ and y; hence the name 'expected' utility.

The choice of utility function u reflects the purpose of the experiment. To accommodate nonlinear models and avoid restrictive distributional assumptions on the parameters or model predictions, we advocate the use of utility functions that reflect the gain in *Shannon information* in quantities of interest [42]. For instance, if the object of the experiment is *parameter inference*, then a useful utility function is the relative entropy or Kullback-Leibler (KL) divergence from the posterior to the prior:

$$\begin{aligned} u(\eta, y, \theta) = u(\eta, y) &= D_{\mathrm{KL}}\left(p(\theta|y,\eta) \,\|\, p(\theta)\right) \\ &= \int_{\theta} p(\theta|y,\eta) \log \frac{p(\theta|y,\eta)}{p(\theta)}\, d\theta. \end{aligned} \tag{5}$$

Taking the expectation of this quantity over the prior predictive of the data, as in (4), yields a U equal to the *expected information gain* in θ. This quantity is equivalent to the mutual information [26] between the data and the parameters, $I(y;\theta)$.

Inferring parameters may not be the true object of an experiment, however. For many experiments, the goal is to improve *predictions* of some quantity Q. This quantity may depend strongly on some model parameters and weakly on others. Moreover, some model parameters might simply be "knobs" without a strict physical interpretation or meaning. In this setting, we can put $u(\eta, y, \theta) = u(\eta, y)$ equal to the Kullback-Leibler divergence evaluated from the posterior predictive distribution, $p(Q|y,\eta) = \int p(Q|\theta)p(\theta|y,\eta)d\theta$, to the prior predictive distribution, $p(Q) = \int p(Q|\theta)p(\theta)d\theta$. Taking the expectation of this utility function over the data yields $U(\eta) = I(Q;y|\eta)$, that is, the conditional mutual information between data and predictions. This quantity

implicitly incorporates an information theoretic "forward" sensitivity analysis, as the experiments that are most informative about Q will automatically constrain the directions in the parameter space that strongly influence Q.

As mentioned above, another common experimental goal is *model discrimination*. From the Bayesian perspective, we wish to maximize the relative entropy between the posterior and prior distributions over models:

$$u(\eta, y) = \sum_i P(M_i|y,\eta) \log \frac{P(M_i|y,\eta)}{P(M_i)}. \tag{6}$$

Moving from this utility to an expected utility requires integrating over the prior predictive distribution of the data, as specified in (4). Since the utility function u here does not depend on the parameters θ, we simply have $U(\eta) = \int_{\mathcal{Y}} u(\eta, y) p(y|\eta) dy$. Because we are now considering multiple competing models, however, the prior predictive distribution is itself a mixture of the prior predictive distribution of each model:

$$p(y|\eta) = \sum_i P(M_i) p(y|M_i, \eta) = \sum_i P(M_i) \int_{\Theta_i} p(y|\theta_i, \eta, M_i) p(\theta_i|M_i) d\theta_i. \tag{7}$$

The resulting expected information gain in model space favors designs that are expected to focus the posterior distribution onto fewer models [75]. In more intuitive terms, we will be driven to test where we know the least and where we also expect to learn the most.

2.3 Computational considerations

Evaluating expected information gain. Except in special cases (e.g., linear-Gaussian models), the expected utilities described above cannot be evaluated in closed form. Instead, the integrals in these expressions must be approximated numerically. Note that, even in the simplest case of parameter inference—with utility given by (5)—evaluating the posterior density of the parameters requires calculating the posterior normalizing constant, which (like the posterior distribution itself) is a function of the data y and the design parameters η. In this situation, it is convenient to rewrite the expected information gain in the parameters θ as follows:

$$\begin{aligned} U(\eta) &= \int_{\mathcal{Y}} \int_{\Theta} p(\theta|y,\eta) \log \left(\frac{p(\theta|y,\eta)}{p(\theta)} \right) d\theta \, p(y|\eta) \, dy \\ &= \int_{\mathcal{Y}} \int_{\Theta} \log \left(\frac{p(y|\theta,\eta)}{p(y|\eta)} \right) p(y|\theta,\eta) \, p(\theta) \, d\theta \, dy \\ &= \int_{\mathcal{Y}} \int_{\Theta} \{\log p(y|\theta,\eta) - \log p(y|\eta)\} \, p(y|\theta,\eta) \, p(\theta) \, d\theta \, dy, \end{aligned} \tag{8}$$

where the second equality is due to the application of Bayes' rule to the quantities both inside and outside the logarithm. Introducing Monte Carlo approximations of the evidence $p(y|\eta)$ and the outer integrals, we obtain the nested Monte Carlo estimator proposed by Ryan [84]:

$$U(\eta) \approx \hat{U}_{N,M}(\eta) := \frac{1}{N} \sum_{i=1}^{N} \left\{ \log \left(p(y^{(i)}|\theta^{(i)}, \eta) \right) - \log \left(\frac{1}{M} \sum_{j=1}^{M} p(y^{(i)}|\tilde{\theta}^{(i,j)}, \eta) \right) \right\}. \tag{9}$$

Here $\{\theta^{(i)}\}$ and $\{\tilde{\theta}^{(i,j)}\}$, $i = 1 \dots N$, $j = 1 \dots M$, are independent samples from the prior $p(\theta)$, and each $y^{(i)}$ is an independent sample from the likelihood $p(y|\theta^{(i)}, \eta)$, for $i = 1 \dots N$. The variance of this estimator is approximately $A(\eta)/N + B(\eta)/(NM)$, and its bias is (to leading order) $C(\eta)/M$ [84], where A, B, and C are terms that depend only on the distributions at hand. The estimator $\hat{U}_{N,M}$ is thus biased for finite M, but asymptotically unbiased.

Analogous, though more complex, Monte Carlo estimators can be derived for the expected information gain in some predictions Q, or for the expected information gain in the model indicator M_i.

Optimization approaches. Regardless of the particular utility function u used to define U, selecting an optimal experimental design requires solving an optimization problem of the form:

$$\max_{\eta \in \mathcal{H}} U(\eta). \tag{10}$$

Using the Monte Carlo approaches described above, only noisy estimates (e.g., $\hat{U}_{N,M}$) of the objective function U are available. Hence, the optimal design problem becomes a stochastic optimization problem, typically over a continuous design space $\mathcal{H}$. Many algorithms have been devised to solve continuous optimization problems with stochastic objectives. While some do not require the direct evaluation of gradients (e.g., Nelder-Mead [76], Kiefer-Wolfowitz [54], and simultaneous perturbation stochastic approximation [90]), other algorithms can use gradient evaluations to great advantage. Broadly, these algorithms involve either stochastic approximation (SA) [56] or sample average approximation (SAA) [87], where the latter approach must also invoke a gradient-based deterministic optimization algorithm. SA requires an unbiased estimator of the gradient of the objective, computed anew at each optimization iteration. SAA approaches, on the other hand, "freeze" the randomness in the objective and solve the resulting deterministic optimization problem, the solution of which yields an *estimate* of the solution of (10) [6]. Hybrids of the two approaches are possible as well. [49] presents a systematic comparison of SA and SAA approaches in the context of optimal experimental design, where SAA is coupled with a BFGS quasi-Newton method for deterministic optimization.

An alternative approach to the optimization problem (10) involves constructing and optimizing Gaussian process models of $U(\eta)$, again from noisy evaluations. As presented in [96], this approach generalizes the EGO (efficient global optimization) algorithm of [51] by choosing successive evaluation points η according to an expected quantile improvement criterion [80].

Surrogate models. An efficient optimization approach is only one part of the computational toolbox for optimal experimental design. Evaluating estimators such as $\hat{U}_{N,M}(\eta)$ (9) for even a single value of η can be computationally taxing when the likelihood $p(y|\theta, \eta)$ contains a computationally intensive model $G(\theta, \eta)$—a situation that occurs very often in physical systems, including in materials science. As a result, considerable effort has gone into the development of reduced-order or "surrogate" models, designed to serve as computationally inexpensive replacements for G.

Useful surrogate models can take many different forms. [34] categorizes surrogates into three different classes: data-fit models, reduced-order models, and hierarchical models. Data-fit models are typically generated using interpolation or regression of of the input-output relationship induced by the high-fidelity model $G(\theta, \eta)$, based on evaluations of G at selected input values $(\theta^{(i)}, \eta^{(i)})$. This class includes polynomial chaos expansions that are constructed non-intrusively [41, 57, 100] and, more broadly, interpolation or pseudospectral approximation with standard basis functions on (adaptive) sparse grids [24, 40, 101]. Gaussian process emulators [53, 99], widely used in the statistics community, fall into this category as well. Indeed, the systematic and efficient construction of data-fit surrogates, particularly for high-dimensional input spaces, has been the focus of a vast body of work in computational mathematics and statistics over the past decade. While many of these methods are used in forward uncertainty propagation (e.g., the solution of PDEs with random input data), recent work [48] has employed sparse grid polynomial surrogates specifically for the case of optimal Bayesian experimental design.

Reduced-order models are commonly derived using a projection framework; that is, the governing equations of the forward model are projected onto a subspace of reduced dimension. This reduced subspace is defined via a set of basis vectors, which, for general nonlinear problems, can be calculated via the proper orthogonal decomposition (POD) [47, 88, 81] or with reduced basis methods [77, 43]. For both approaches, the empirical basis is pre-constructed using full forward problem simulations or "snapshots." Systematic projection-based model reduction schemes for parameter-dependent models have also seen extensive development in recent years [17, 22]. To our knowledge, such reduction schemes have not yet been used for optimal experimental design, but in principle they are directly applicable.

Hierarchical surrogate models span a range of physics-based models of lower accuracy and reduced computational cost. Hierarchical surrogates are derived from higher-fidelity models using approaches such as simplifying physics assumptions, coarser grids, alternative basis expansions, and looser residual tolerances. These approaches may not be particularly systematic, in that their success and applicability are strongly problem-dependent, but they can be quite powerful in certain cases. One of the examples in the next section will use a reduced order model derived from a combination of simplifying physics assumptions and fits to simulation data from a higher-fidelity model.

3 Examples of optimal experimental design

In this section, we present two examples of Bayesian experimental design in materials-related applications. The first illustrates experimental design for parameter estimation in a simple substrate-film model. This example also demonstrates the usefulness of reduced-order models in accelerating the design process. The second example is concerned with experimental design for model selection. It will illustrate this process using competing models of impurity precipitation at heterophase interfaces.

3.1 Film-substrate systems: design for parameter inference

A classical application of Bayesian methods to physical modeling involves inferring the properties of the interior of an object from observations of its surface, e.g., of the mantle or core of the Earth from observations at the Earth's crust [16, 44]. In the context of materials science, similar problems arise when observing the surface of a material and trying to infer the subsurface properties. One example of such a problem involves observing a thin film deposited on a heterogeneous substrate. The heterogeneity of the substrate—e.g., in temperature [58], local chemistry [3], or topography [14]—induces some corresponding heterogeneity in the film—e.g., melting [58], condensation [3], or buckling [14]. The goal is to deduce information about the substrate from the behavior of the film.

We have recently developed a convenient model for studying the inference of substrate properties from film behavior [2]. Figure 1 shows a film deposited on a substrate. Though the substrate is not directly observable, we would like to infer its properties from the behavior of the film deposited above. In the present example, we will use this simple model to demonstrate aspects of Bayesian experimental design. Our objective will be to choose experiments that provide maximal information about a parameter of interest for a fixed number of allowed experiments.

3.1.1 Physical background

In our model problem, the substrate is described by a non-uniform scalar field $\mathcal{T}(x, y)$ on a two-dimensional spatial domain, $(x, y) \in \Omega := [0, L_D] \times [0, L_D]$. In other words, $\mathcal{T}(x, y)$ describes the variation of the substrate property $\mathcal{T}$ over a square domain. Realizations of the substrate are random, and hence we model $\mathcal{T}(x, y)$ as a zero-mean Gaussian random field with a squared exponential covariance kernel [82]. One of the key parameters of this covariance kernel is the characteristic length scale ℓ_s, which describes the scale over which spatial variations in the random field occur. When ℓ_s is large, realizations of the substrate field have a relatively coarse structure, while smaller values of ℓ_s produce realizations with more fine-scale variation. The film deposited on the substrate is a two-component mixture represented by an order parameter field $c(x, y, t)$. The order parameter takes values in the range $[-1, 1]$, where $c = -1$ and $c = 1$ represent single-component phases and $c = 0$ represents a uniformly mixed phase.

The behavior of the film is modeled by the Cahn-Hilliard equation [18]:

$$\frac{\partial c}{\partial t} = \Delta \left(\frac{\partial g}{\partial c} - \epsilon^2 \Delta c \right), \tag{11}$$

where

$$g\left(c, \mathcal{T}(x, y)\right) = \frac{c^4}{4} + \mathcal{T}(x, y)\frac{c^2}{2} \tag{12}$$

is a substrate-dependent energy potential function. The two components of the film separate in regions of the substrate where $\mathcal{T}(x, y) < 0$ and mix in regions where $\mathcal{T}(x, y) > 0$. Hence, the substrate field can be thought of as a difference from some critical temperature, where temperatures above the critical value promote phase mixing while those below the critical value promote phase separation. The parameter ϵ

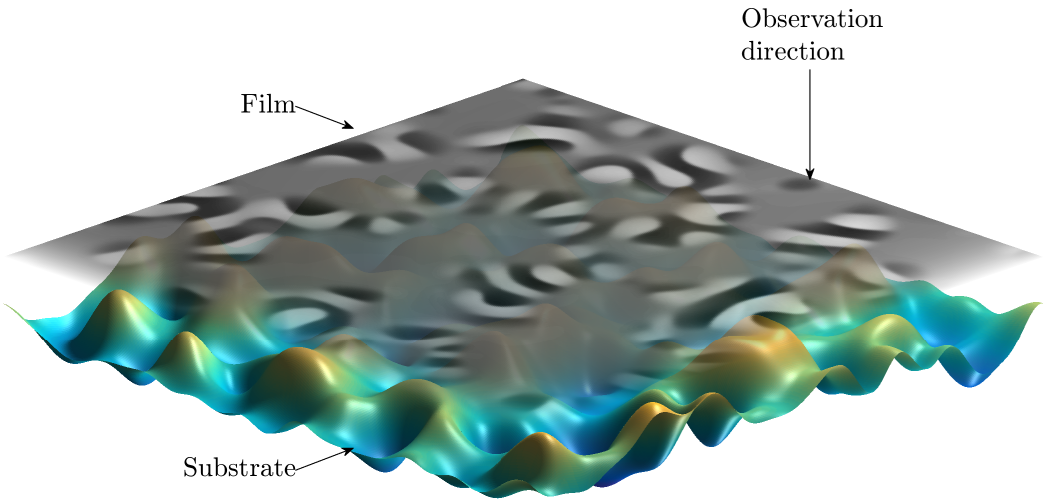


Figure 1: A film deposited on top of a substrate. The substrate is not directly observable, but some of its properties may be inferred from the behavior of the film.

in (11) governs the thickness of the interface between separated phases. Films with larger values of ϵ have thicker interfaces between their phase-separated regions than films with lower values of ϵ.

We model the time evolution of an initially uniform film $c(x, y, t = 0) = 0$ deposited on a substrate by solving the Cahn-Hilliard equation using Eyre's method for time discretization [35]. We find that the order parameter field c converges to a static configuration in the long time limit for any combination of ℓ_s and ϵ. A detailed description of the model implementation and analysis of the time-dependence of c is given in [2]. For the purpose of the example presented here, it suffices to know that the converged order parameter field has a characteristic length scale of its own, which we call Λ_∞.

Figure 2 illustrates converged order parameter fields of films with two different values of ϵ ($\epsilon = 0.02$ and $\epsilon = 0.04$) deposited on substrates with two different values of ℓ_s ($\ell_s = 0.77$ and $\epsilon = 0.13$). For both substrates, we observe that increasing the value of ϵ increases the value of Λ_∞. Yet the behavior of the film on the two substrates is qualitatively different. For the substrate with $\ell_s = 0.77$, the thickness of interfaces between phase-separated parts of the film is sufficiently small for fluctuations in c to be correlated with fluctuations in $\mathcal{T}$. By contrast, no direct correlation of this sort exists

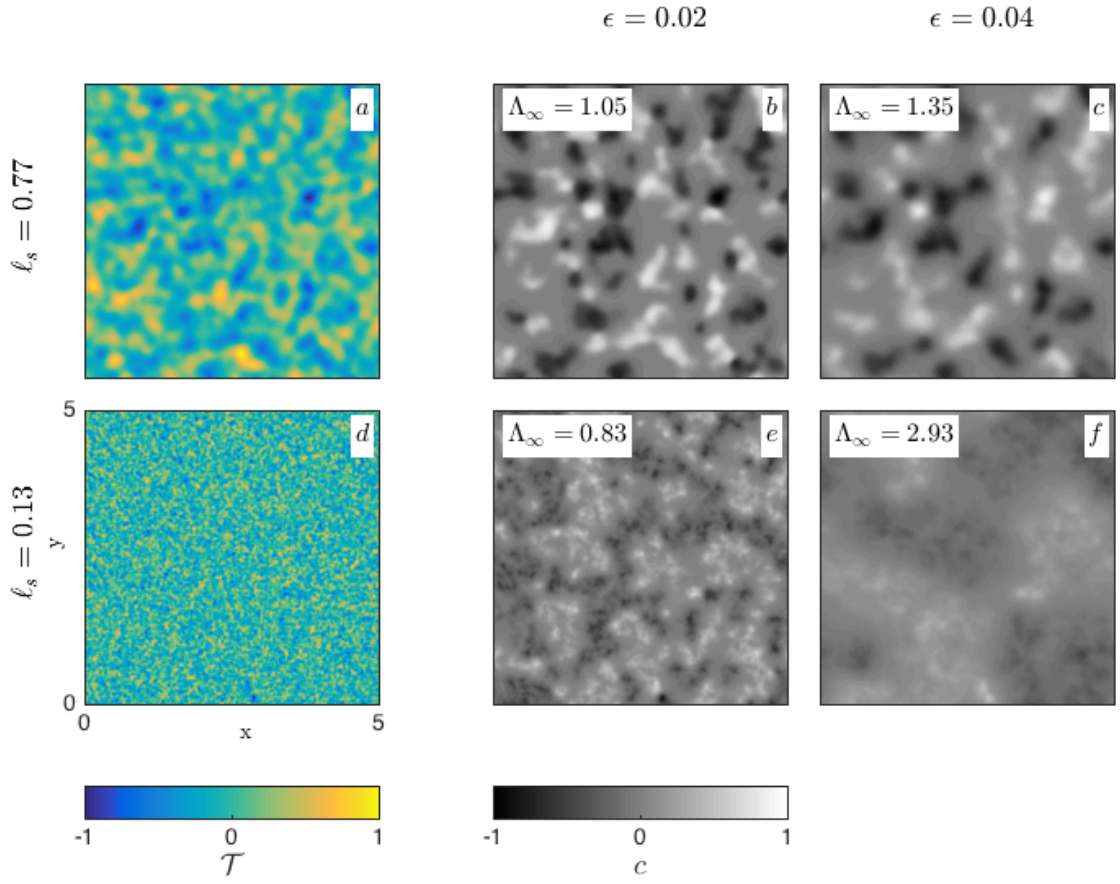


Figure 2: Substrate fields with (a) $\ell_s = 0.77$ and (d) $\ell_s = 0.13$. Plots (b) and (c) show converged order parameter distributions for films deposited on the substrate in (a) with $\epsilon = 0.02$ and $\epsilon = 0.04$ respectively. Similarly, plots (e) and (f) show converged order parameter distributions for films deposited on the substrate in (d) with $\epsilon = 0.02$ and $\epsilon = 0.04$ respectively. The converged length scale Λ_∞ is indicated for each film.

for the substrate with $\ell_s = 0.13$, because its characteristic length is smaller than the thickness of interfaces between phase-separated parts of the films in Figure 2(e) and (f). Instead, the fluctuations in c for these films reflects a local spatial average of T over a length scale that depends on ϵ.

The value of ϵ determines how Λ_∞ changes with ℓ_s. For example, in films with $\epsilon = 0.02$, reducing the value of ℓ_s from 0.77 to 0.13 reduces Λ_∞ from 1.05 to 0.83. However, the opposite effect is observed for $\epsilon = 0.04$, where reducing ℓ_s from 0.77 to 0.13 increases Λ_∞ from 1.35 to 2.93. These observations show that Λ_∞, ℓ_s, and ϵ are

related, albeit in a non-trivial way.

Our goal is to infer the substrate length scale ℓ_s from the value of Λ_∞ of a film of known ϵ, deposited on the substrate. In this context, Λ_∞ is the data obtained from an experiment, ℓ_s is the value to be inferred, and ϵ is a parameter of the experiment that we control (e.g., by manipulating the chemical composition of the film). In previous work, we showed how to perform this inference and how to improve it by performing multiple measurements of Λ_∞ using films with different ϵ values [2]. In the experimental design problem described here, we would like to choose optimal values of ϵ that lead to the most efficient inference of ℓ_s.

For any given ℓ_s and ϵ, Λ_∞ may be obtained by solving the Cahn-Hilliard equation for the time evolution of the film on the substrate. This calculation does not call for extraordinary computational resources; indeed, it can be performed in roughly 100 seconds on a modern workstation. In Bayesian experimental design, however, this calculation would have to be carried out many millions of times. The potential computational effort of this approach is compounded by the stochasticity of $\mathcal{T}(x, y)$; to evaluate the likelihood function for any given value of ℓ_s, we must account for many possible substrate field realizations. Therefore, to make optimal experimental design tractable, we construct a "reduced order model" (ROM) relating Λ_∞, ℓ_s, and ϵ. We use a relation of the form

$$\Lambda_\infty = \overbrace{\underbrace{f(\epsilon, \ell_s)}_{\text{deterministic term}} + \underbrace{\gamma(\epsilon, \ell_s)}_{\text{random term}}}^{\text{reduced order model}} . \tag{13}$$

The deterministic term captures the average response of the film/substrate system, and the random term captures the inherent stochasticity of the film/substrate system and any systematic error in the deterministic term. The stochasticity of the film/substrate system is due to the random nature of the substrate field and the initial condition of the Cahn-Hilliard equation, among other factors [2].

The proposed ROM can be simplified using the Buckingham Pi theorem [102]. Since ϵ, ℓ_s, and Λ_∞ all have dimensions of length, we can form two Pi groups: $(\Lambda_\infty/\epsilon)$ and (ℓ_s/ϵ). The ROM may then be simplified to

$$\frac{\Lambda_\infty}{\epsilon} = F\left(\frac{\ell_s}{\epsilon}\right) + \Gamma\left(\frac{\ell_s}{\epsilon}\right). \tag{14}$$

To obtain the form of $F(\ell_s/\epsilon)$ and $\Gamma(\ell_s/\epsilon)$, we carried out multiple runs of the Cahn-Hilliard model, with values of ℓ_s sampled over $[0.1, 1]$ and values of ϵ sampled over $[0.01, 0.1]$. Figure 3(a) plots Λ_∞/ϵ as a function of ℓ_s/ϵ, confirming that these quantities lie on a single curve, on average. However, there is a spread about this curve as well. This is caused by the stochastic nature of the relation between Λ_∞/ϵ and ℓ_s/ϵ, and justifies the random term in the ROM. The exact forms of $F(\ell_s/\epsilon)$ and $\Gamma(\ell_s/\epsilon)$ are then:

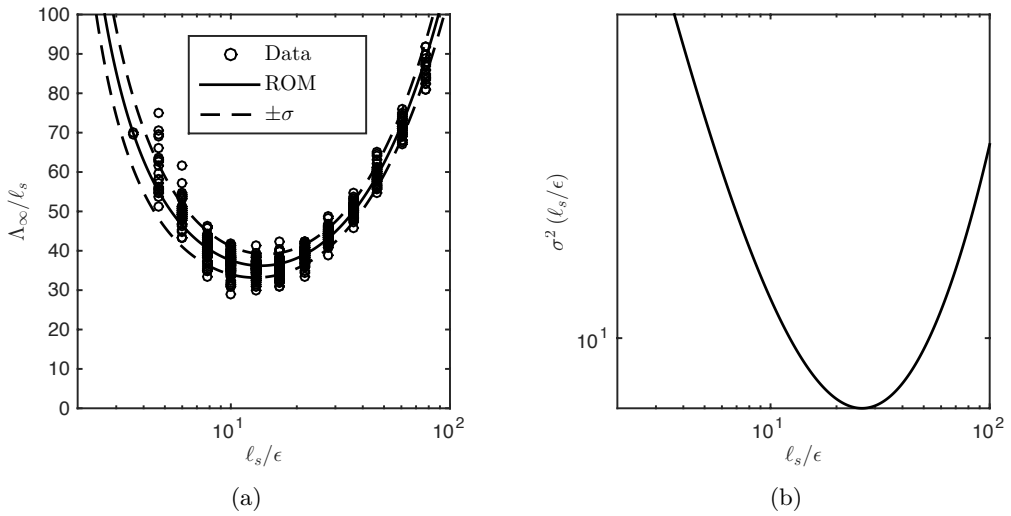
$$F\left(\frac{\ell_s}{\epsilon}\right) = \frac{\ell_s}{\epsilon}\left(a + \frac{b}{(\ell_s/\epsilon - 1)^c}\right) \tag{15}$$

$$\Gamma\left(\frac{\ell_s}{\epsilon}\right) \sim N\left(0, \sigma^2\left(\frac{\ell_s}{\epsilon}\right)\right) \tag{16}$$

with parameters of the mean term F obtained by least squares fitting:

$$a = 1.05 \quad b = 79.51 \quad c = 1.54.$$

The dependence of σ^2 on (ℓ_s/ϵ) is captured nonparametrically using Gaussian process regression [82], as shown in Figure 3(b). Details of the derivation of the ROM can be found in [2].



(a) (b)

Figure 3: (a) A plot of Λ_∞/ϵ against ℓ_s/ϵ. (b) A plot of the non-stationary variance of the random term $\Gamma(\ell_s/\epsilon)$.

To perform inference, we use the Cahn-Hilliard model as a proxy for a physical experiment. We generate multiple realizations of substrates with the same value of ℓ_s. Then, using each substrate as an input, we run the Cahn-Hilliard model, which also requires ϵ as a parameter. Given one or more choices for ϵ and the values of Λ_∞ thus obtained, we infer the value of ℓ_s using the ROM. Inference may be conducted using one or multiple $(\Lambda_\infty, \epsilon)$ pairs.

To infer ℓ_s in a Bayesian setting, we need to calculate the likelihood $p(\Lambda_\infty, \epsilon)$. This can be done using the ROM as follows:

$$p\left(\Lambda_\infty|\ell_s,\epsilon\right) = \frac{1}{\sqrt{2\pi}\sigma(\ell_s/\epsilon)} \exp\left(-\frac{\left(\Lambda_\infty/\ell_s - F(\ell_s/\epsilon)\right)^2}{2\sigma^2(\ell_s/\epsilon)}\right). \tag{17}$$

Since runs of the Cahn-Hilliard equation are conditionally independent given ℓ_s and ϵ, the likelihood for multiple $(\Lambda_\infty, \epsilon)$ pairs can be found using the product rule

$$p\left(\Lambda_{\infty,1:n}|\ell_s,\epsilon_{1:n}\right) = \prod_i p\left(\Lambda_{\infty,i}|\ell_s,\epsilon_i\right). \tag{18}$$

Finally, the posterior density is calculated using Bayes' rule

$$p(\ell_s|\Lambda_{\infty,1:n},\epsilon_{1:n}) = \frac{p(\Lambda_{\infty,1:n}|\ell_s,\epsilon_{1:n})p(\ell_s)}{\int p(\Lambda_{\infty,1:n}|\ell_s,\epsilon_{1:n})p(\ell_s)d\ell_s}. \tag{19}$$

We use a truncated Jeffreys prior [50] for ℓ_s

$$p(\ell_s) \propto \ln(1/\ell_s),\ \ell_s \in [0.1,1]. \tag{20}$$

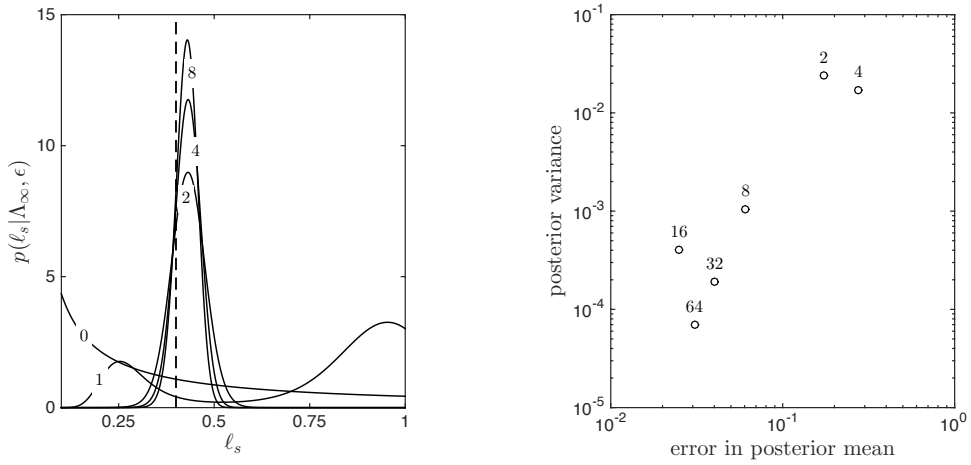
The prior density is set to zero outside the range $[0.1,1]$. This restriction is imposed for reasons of computational convenience and may easily be relaxed.

The results of an iterative inference process that incorporates successive $(\Lambda_\infty,\epsilon)$ pairs are shown in Figure 4(a). Here, the true value of the substrate length scale (i.e., the value used to generate the data) is $\ell_s = 0.4$. Values of ϵ are selected by sampling uniformly in log-space over the interval $[0.01,0.1]$. The probability density marked '0' (i.e., with zero data points) is the prior. The posterior probability density with one data point (marked '1') is bimodal, but the bimodality of the posterior vanishes with two or more data points. As additional $(\Lambda_\infty,\epsilon)$ pairs are introduced, the peak in the posterior moves towards the true value of $\ell_s = 0.4$. Any number of point estimates of ℓ_s may be calculated from the posterior, such as the mean, median, or mode, but the posterior probability density itself gives a full characterization of the uncertainty in ℓ_s. As an example, we have plotted in Figure 4(b) both the posterior variance (a measure of uncertainty) and the absolute difference between the posterior mean and the true value of ℓ_s (a measure of error) for different numbers of data points. As more data are used in the inference problem, both the posterior variance and the error in the posterior mean decrease. Note that the ultimate convergence of the posterior mean towards the true value of ℓ_s, as the number of data points approaches infinity, is a more subtle issue; it is related to the frequentist properties of this Bayesian estimator, here in the presence of model error. For a fuller discussion of this topic, see [2].

3.1.2 Bayesian experimental design

Thus far, we have described a model problem wherein the characteristic length scale ℓ_s of a substrate is inferred from the behavior of films with known values of ϵ, deposited on the substrate. In the preceding calculations, we chose ϵ randomly from a distribution. Since ϵ is in fact an experimental parameter that we can control, this choice is equivalent to performing experiments at random. Now we would like to consider a more focused experimental campaign, choosing values of ϵ to maximize the information gained with each experiment. In the language of Section 2, we will take our utility function u to be the Kullback-Leibler (KL) divergence from the posterior to the prior (5). The expected utility (4) will represent expected information gain in the parameters ℓ_s. To connect the present problem to the general formulation of Section 2, note that Λ_∞ is the experimental data y, ℓ_s is the parameter θ to be inferred, and ϵ is the experimental parameter η over which we will optimize the expected utility.

The expected KL divergence from posterior to prior is estimated via the Monte Carlo estimator in (9). To perform the calculation, we need to be able to sample $\ell_s^{(i)}$ and $\tilde{\ell}_s^{(i,j)}$ from the prior $p(\ell_s)$, and $\Lambda_\infty^{(i)}$ from the likelihood $p(\Lambda_\infty|\ell_s^{(i)},\epsilon)$. The length



(a) (b)

Figure 4: (a) Posterior probability densities for different numbers of $(\Lambda_\infty, \epsilon)$ pairs. With the inclusion of ever more data, uncertainty in the posterior on ℓ_s decreases steadily. (b) Posterior variance and error in posterior mean for different numbers of $(\Lambda_\infty, \epsilon)$ pairs. Both error and variance decrease with increasing numbers of data points.

scales ℓ_s can be sampled from the truncated Jeffreys prior using a standard inverse CDF transformation [83]. The observation Λ_∞ is Gaussian given ϵ and ℓ_s, and can be sampled by evaluating (14) with distributional information given in (15)–(16).

We will use this formulation to design an optimal experiment consisting of two measurements. In other words, two films with independently controlled values of ϵ will be deposited on substrates with the same value of ℓ_s, and the two values of Λ_∞ generated will be used for inference. The values of ϵ will be restricted to the design range $[0.01, 0.095]$. As before, this restriction is not essential and is easily relaxed. Figure 10 shows the resulting Monte Carlo estimates of expected information gain $U(\epsilon_1, \epsilon_2)$.

Because the ordering of the experiments is immaterial, the map of the expected information gain is symmetric about the $\epsilon_1 = \epsilon_2$ line, aside from Monte Carlo estimation error. We draw attention to three points marked by squares in Figure 10. The first is at $(\epsilon_1, \epsilon_2) = (0.025, 0.025)$, where $U(\epsilon_1, \epsilon_2) = 0.49$; it is near the minimum of the expected utility function. This point corresponds to the least useful pair of experiments. The second is at $(\epsilon_1, \epsilon_2) = (0.01, 0.095)$, with $U(\epsilon_1, \epsilon_2) = 2.9$; it is the maximum of the expected utility map and is expected to yield the most informative experiments. The point $(\epsilon_1, \epsilon_2) = (0.08, 0.08)$, where $U(\epsilon_1, \epsilon_2) = 2.0$, lies midway between these extremes: it is expected to be more informative than the first design but less informative than the second.

To illustrate how the three (ϵ_1, ϵ_2) pairs highlighted above yield different expected utilities, we carry out the corresponding inferences of ℓ_s following the procedure de-

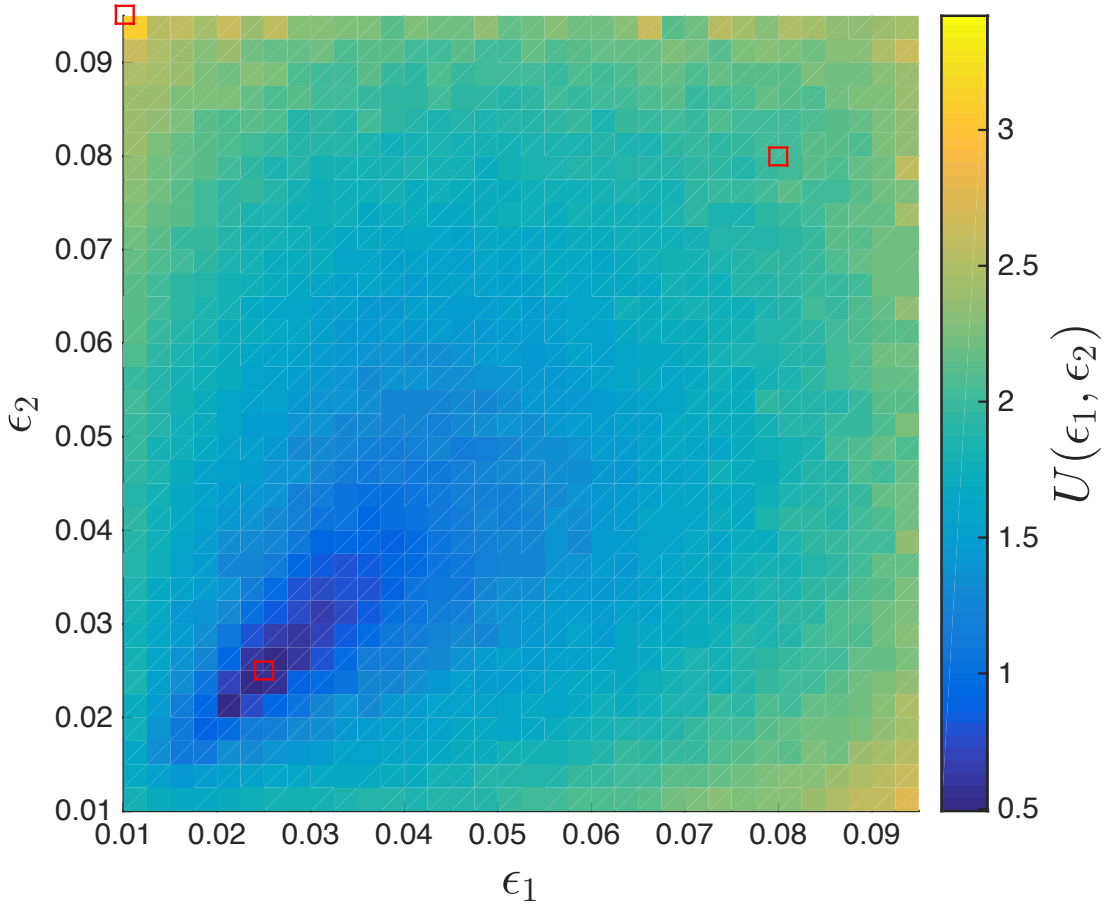


Figure 5: Map of expected information gain $U(\epsilon_1, \epsilon_2)$ in the substrate length scale parameter ℓ_s, as a function of experimental design parameters ϵ_1 and ϵ_2. The three experiments discussed in the text are marked with red squares.

scribed in Section 3.1.1. To simulate each experiment, we fix ℓ_s and the desired value of ϵ, then generate a converged order parameter length scale Λ_∞ by generating a realization of the substrate and simulating the Cahn-Hilliard equation. Given the data $\Lambda_{\infty,1}$ and $\Lambda_{\infty,2}$ corresponding to (ϵ_1, ϵ_2), we evaluate the corresponding

posterior density and calculate the actual KL divergence from posterior to prior, $D_{KL}\left(p(\ell_s|\Lambda_{\infty,1:2},\epsilon_{1:2})\|p(\ell_s)\right)$. The results of these three experiments are summarized in Figure 6(a). As expected, the second experiment, performed at $(\epsilon_1,\epsilon_2) = (0.01, 0.095)$ is the most informative, and has a large information gain of $D_{KL} = 2.10$ nats.[1] The first experiment is the least informative, with a small information gain of $D_{KL} = 0.99$ nats. The third experiment, with $D_{KL} = 1.44$ nats, lies in between. The actual values of D_{KL} are different from their expected values because the expected information gains are calculated by averaging over all possible prior values of ℓ_s and all possible experimental outcomes, whereas the actual values are calculated only for particular ℓ_s and Λ_∞ values, given ϵ. However, the values of D_{KL} follow the same trend as their expectations.

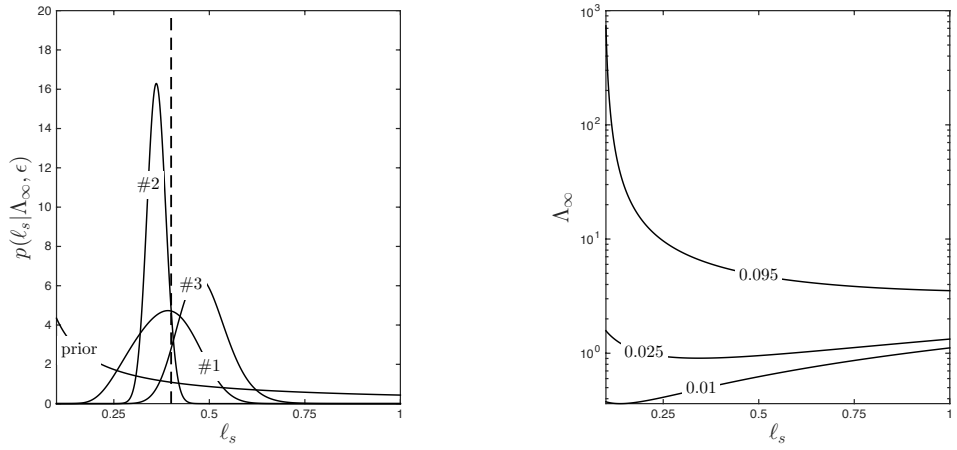
To better understand why these experiments produce different values of the information gain, Figure 6(b) plots $\Lambda_\infty = \epsilon F(\ell_s/\epsilon)$ as a function of ℓ_s for $\epsilon = 0.01$, 0.025, and 0.095. We observe that Λ_∞ is not very sensitive to variations in ℓ_s for $\epsilon = 0.025$. This explains why an experiment with $(\epsilon_1,\epsilon_2) = (0.025, 0.025)$ is not particularly informative. On the other hand, Λ_∞ is sensitive to variations in ℓ_s for $\epsilon = 0.095$ and $\epsilon = 0.01$. Additionally, Λ_∞ is a decreasing function of ℓ_s for $\epsilon = 0.095$, and an increasing function for $\epsilon = 0.01$. The complementarity of these trends makes the experiment $(\epsilon_1,\epsilon_2) = (0.01, 0.095)$ especially useful.

We can also compare the optimal experiment to the random experiments shown in Figure 4. The information gained in the optimal experiment ($D_{KL} = 2.10$ nats), with two values of ϵ, is comparable to the information gained from the experiment with *eight* randomly selected values of ϵ ($D_{KL} = 2.29$ nats). Hence by using optimal Bayesian experimental design in this example, we are able to reduce the experimental effort over a random strategy by roughly a factor of four! This reduction is especially valuable when experiments are difficult or expensive to conduct.

3.2 Heterophase interfaces: design for model discrimination

As noted in Section 2.1, experiments often yield data that may be explained by multiple models. Additional measurements may then be required to determine which of many possible models is best supported by the data. In such situations, it is desirable to determine which further experiments are likely to distinguish between alternative models most efficiently. Naturally, this guidance is needed before the additional work is actually carried out. Determining which experiments are most informative for distinguishing between alternative models is the goal of Bayesian experimental design for model selection [63, 75]. This capability is especially useful when the experiments are very resource-intensive and brute force data acquisition over a wide parameter range is not feasible. This section will illustrate Bayesian experimental design for model selection on an example taken from investigations of heterophase interfaces in layered metal composites.

[1] A nat is a unit of information, analogous to a bit, but with a natural logarithm rather than a base two logarithm in (5).



(a) (b)

Figure 6: (a) Experiments corresponding to the three (ϵ_1, ϵ_2) pairs indicated in Figure 5. The posterior densities from the three experiments are marked #1, #2, and #3. (b) $\Lambda_\infty = \epsilon F(\ell_s/\epsilon)$ versus ℓ_s for $\epsilon = 0.01, 0.025, 0.095$.

3.2.1 Physical background

Figure 7 shows a multilayer composite of two transition metals—copper (Cu) and niobium (Nb)—created by physical vapor deposition (PVD) [74]. In this synthesis technique, atoms impinge upon a flat substrate, adhere to it, and aggregate into crystalline layers. By alternating the elements being deposited—e.g., first Cu, then Nb, then Cu again, and so on—layered composites such as the one in Figure 7 may be made. The thickness of each layer may be controlled by changing the total deposition time for each element. Many multilayer compositions besides Cu/Nb have been synthesized this way, including Cu/V (V: vanadium) [37, 105], Cu/Mo (Mo: molybdenum) [62], Ag/V (Ag: Silver) [98, 97], Al/Nb (Al: Aluminum) [37, 61, 62], Fe/W (Fe: Iron, W: Tungsten) [60], and others [11, 12, 86].

Layered composites are ideal for studying the properties of heterophase interfaces. In Figure 7, each pair of adjacent Cu and Nb layers forms one Cu-Nb interface. The total amount of interface area per unit volume of the material may be changed by adjusting the thickness of the layers. For composites where all the individual layers have identical thickness l, the volume of material corresponding to interface area A is $V = A \times l$. Thus, the interface area per unit volume is $A/V = 1/l$: as the layers are made thinner, A/V rises and the influence of interfaces on the physical properties of the composite as a whole increases. For l in the nanometer range, i.e., $l \lesssim 10$ nm, interfaces dominate the behavior of the multilayer composite, leading to enhanced strength [73], resistance to radiation [72], and increased fatigue life [95]. In multilayer composites, all of these desirable properties are due to the influence of interfaces. Thus, considerable

effort continues to be invested into elucidating the structure and properties of individual interfaces [71, 19, 10].

The present example will consider the relationship between the structure of metal-metal heterophase interfaces and trapping of helium (He) impurities. Implanted He is a major concern for the performance of materials in nuclear energy applications [92, 106]. Trapping and stable storage of these impurities at interfaces is one way of mitigating the deleterious effects of implanted He [78, 32]. The influence of interfaces on He may be clearly seen in multilayer composites. Experiments carried out on Cu-Nb [29], Cu-Mo [62], and Cu-V [38] multilayers synthesized by PVD show that implanted He is preferentially trapped at the interfaces. Moreover, not all interfaces are equally effective at trapping He: the maximum concentration c of interfacial He—expressed as the number of He atoms per unit interface area—that may be trapped at an interface before detectable He precipitates form differs from interface to interface [32].

Figure 8 plots c for Cu-Nb, Cu-Mo, and Cu-V interfaces as a function of one parameter: the interface "misfit" m. Cu, Nb, Mo, and V all have cubic crystal structures: Cu is face-centered cubic (fcc) while Nb, Mo, and V are body-centered cubic (bcc). Thus, all three composites used in Figure 8 are made up of alternating fcc (Cu) and bcc (Nb, Mo, or V) layers. The edge length of a single cubic unit cell in fcc or bcc crystals is the lattice parameter, a_{fcc} or a_{bcc}. The misfit m is defined as $m = a_{\text{bcc}}/a_{\text{fcc}}$. Intuitively, m measures the mismatch in inter-atomic spacing in the adjacent crystals that make up an interface.

According to Figure 8, the ability of interfaces to trap He, as measured by c, depends on the misfit: $c = c(m)$. A simple model that may be proposed based on this data is that the relationship between c and m is linear: $c = \alpha_0 + \alpha_1 m$. Indeed, a linear fit represents the available data reasonably well. Its most apparent drawback is that it predicts negative c values for $m \lesssim 0.83$. Thus, a better model might be $c = \alpha|m - m_0|$. This model predicts that c drops to zero as m decreases to m_0 and begins to rise again as m is further reduced below m_0. Physically, this model may be rationalized by stating that at some special value of misfit, m_0, the atomic matching between adjacent layers is especially good, leading to few sites at the interface where He impurities may be trapped. As m departs from m_0 in either direction, the atomic matching becomes worse, giving rise to more He trapping sites and therefore higher c.

The structure of fcc/bcc interfaces—including the degree of atomic matching—may be investigated in detail by constructing atomic-level models [33, 30]. Figure 9 shows such a model of the terminal atomic planes of Cu and Nb found at PVD Cu-Nb interfaces. The pattern of overlapping atoms from the two planes contains sites where a Cu atom is nearly on top of a Nb atom. Such sites are thought to be preferential locations for trapping of He impurities [52]. They arise from the geometrical interference of the overlapping atomic arrangements in the adjacent crystalline layers and, in that sense, may be viewed as analogous to features in a Moiré pattern.

The density and distribution of He trapping sites of the type shown in Figure 9 may be computed directly for any given fcc/bcc interface as a function of the geometry of the interface using the well-known O-lattice theory [13]. In PVD Cu/Nb, Cu/Mo, and Cu/V composites, the relative orientation of the adjacent crystals is identical. Thus, differences in the geometry Cu-Nb, Cu-Mo, and Cu-V interfaces arise solely from differences in the lattice parameters of the adjacent crystals, as described by the

misfit parameter, m. The areal density of He trapping sites for these interfaces is therefore only a function of m and may be written as

$$f(m) = \frac{\left|(4m-3)(\sqrt{3}m-\sqrt{2})\right|}{3a_{\mathrm{Cu}}^2 m^2}. \tag{21}$$

Using this expression, we propose a second model for the dependence of c on m, namely: $c = \beta f(m)$. Here, the proportionality constant β determines the number of He atoms that may be trapped at a single site of the type illustrated in Figure 9. The best fit for this second model is plotted in Figure 8. Both this model and the previously described linear model fit the available experimental data reasonably well. Moreover, both predict c values of zero for $m \approx 0.82$–0.83. However, unlike the linear model, $c = \beta f(m)$ predicts that c is also zero at $m \approx 0.75$.

We wish to determine what additional experimental data will help distinguish between the two models described above. However, since measuring even a single value of c requires considerable resources, our goal is to limit the additional data to just one (c, m) pair. In an experiment, we may select m by choosing to synthesize a fcc/bcc multilayer composite of specific composition. In other words, we control m. However, we do not know c in advance. In this context, our goal is to determine what one value of m is most likely to distinguish between the two models, regardless of the c value actually found in the subsequent experiment. In the following section, we will apply Bayesian experimental design to address this challenge. In addition to the two models described above, we will also consider a third model encapsulating the hypothesis that c does not depend on m at all: $c = \gamma = \text{constant}$.

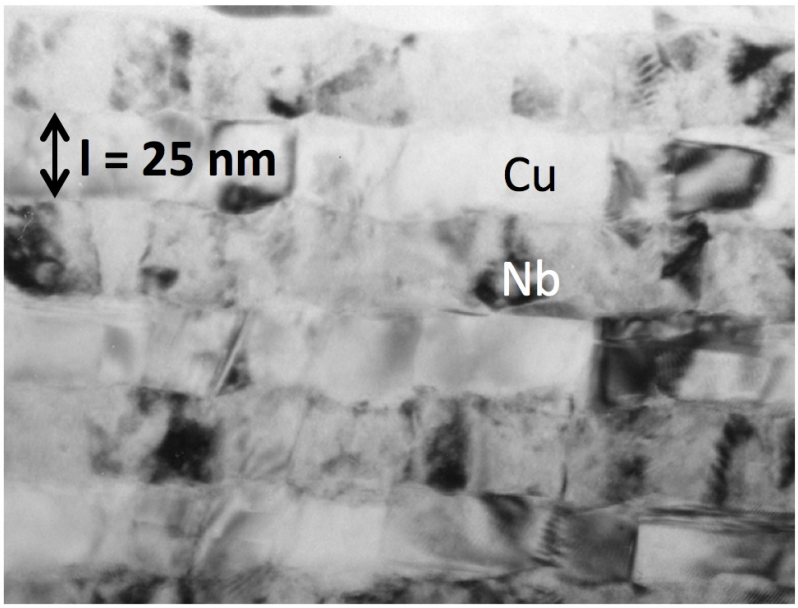


Figure 7: A Cu-Nb multilayer composite synthesized by PVD [31].

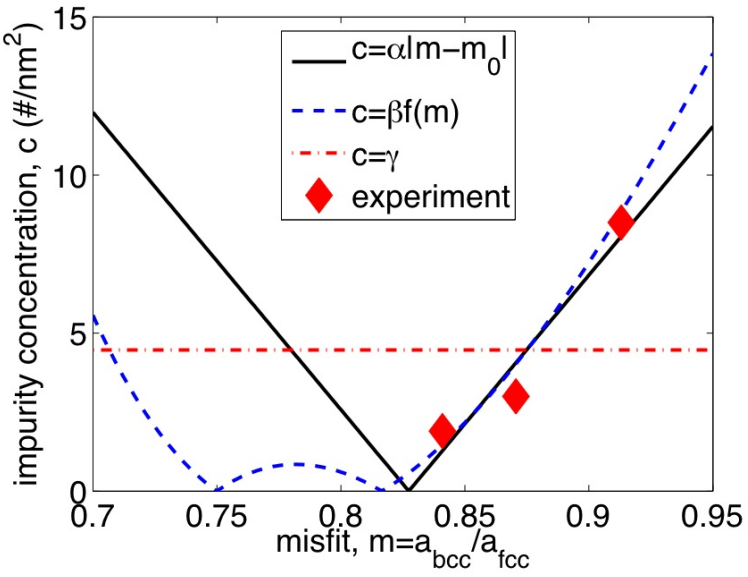


Figure 8: Maximum interfacial He impurity concentration, c, plotted as a function of misfit, m.

3.2.2 Bayesian experimental design for model selection

As described in Section 2, the goal of experimental design is to maximize the expectation of some measure of information. In the present example, we will maximize the expected KL divergence, as applied to model discrimination, described in (6) and (7). In this context, m is the experimental parameter η that we control; c is the observed data y; M_1, M_2, and M_3 are the competing models; (α, m_0) are the parameters θ_1 of model M_1; β is the parameter θ_2 of model M_2; and γ is the parameter θ_3 of model M_3.

The expected KL divergence can be computed by combining (6) and (7); this requires knowing both the prior $p(M_i)$ and the posterior $p(M_i|c,m)$ for each model. We we use a flat or "indifference" prior over models, $p(M_i) = 1/3$. The posterior model probabilities are calculated from Bayes rule as given in (2). Evaluating Bayes' rule in this setting requires that we calculate the marginal likelihood for each model and proposed experiment, $p(c|M_i, m)$, as shown in (3). We now detail this procedure.

The previous section identified three models connecting c and m. They are:

$$M_1: \quad c = \alpha|m - m_0| + \epsilon_1 \tag{22}$$

$$M_2: \quad c = \beta f(m) + \epsilon_2 \tag{23}$$

$$M_3: \quad c = \gamma + \epsilon_3 \tag{24}$$

where $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. In addition to specifying the functional form of each model, each expression above also contains an additive noise term ϵ_i. This term is a random variable that describes uncertainty in the measured c, i.e., due to the observational

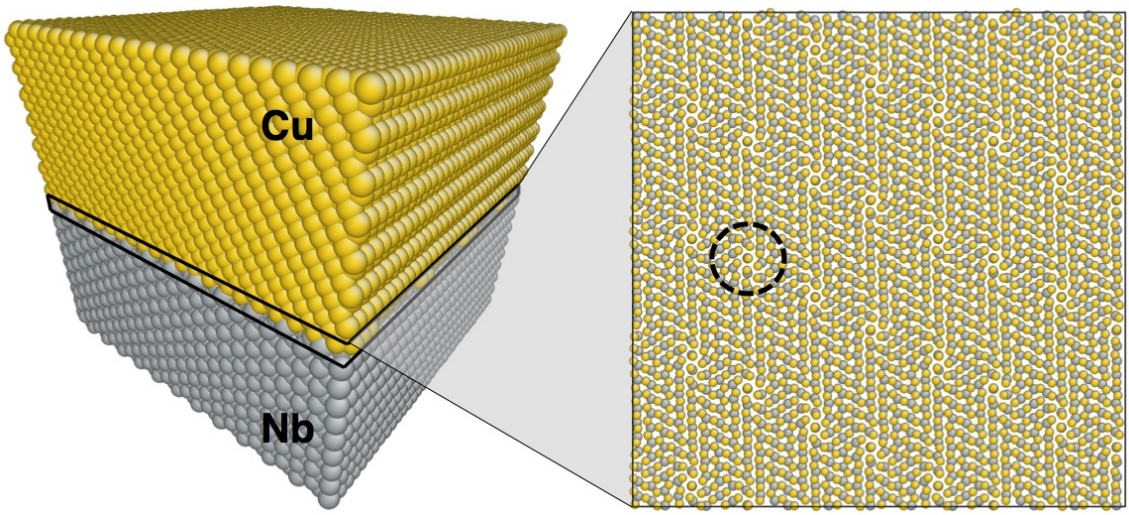


Figure 9: Left: a Cu-Nb bilayer. Right: the terminal Cu and Nb planes that meet at the Cu-Nb interface. He trapping occurs at sites where a Cu atom is nearly on top of a Nb atom. One such site is shown with the dashed circle.

process itself. For simplicity, we assume the observational error variance σ_ϵ^2 to be known. The model parameters α, m_0, β, and γ are endowed with priors that reflect our state of knowledge after performing the three experiments shown in Figure 8, before beginning the current experimental design problem. These priors are taken to be Gaussian. In other words, we suppose that they are the result of Bayesian linear regression with Gaussian priors or improper uniform priors; the posterior following the three previous experiments becomes the prior for the current experimental design problem. We denote the current prior means by $\bar{\alpha}$, $\bar{m}_0$, $\bar{\beta}$, and $\bar{\gamma}$, and the current prior standard deviations as σ_α, σ_{m_0}, σ_β, and σ_γ.

Given these assumptions, we can express the probability density of the observation c for each parameterized model as:

$$p(c|m,\alpha,m_0,M_1) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon}\exp\left(-\frac{(c-\alpha|m-m_0|)^2}{2\sigma_\epsilon^2}\right) \tag{25}$$

$$p(c|m,\beta,M_2) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon}\exp\left(-\frac{(c-\beta f(m))^2}{2\sigma_\epsilon^2}\right) \tag{26}$$

$$p(c|m,\gamma,M_3) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon}\exp\left(-\frac{(c-\gamma)^2}{2\sigma_\epsilon^2}\right). \tag{27}$$

Each of these densities is normal with mean given by the model and variance σ_ϵ^2. For fixed m and c, these densities can be viewed as the *likelihood functions* for the corresponding model parameters, i.e., α and m_0 for model 1, β for model 2, and γ for model 3. To obtain the *marginal likelihoods* $p(c|m,M_i)$, we marginalize out these

parametric dependencies as follows:

$$p(c|m, M_1) = \int_{-\infty}^{\infty} p(c|m, \alpha, m_0, M_1)p(\alpha)p(m_0)\, d\alpha\, dm_0 \tag{28}$$

$$p(c|m, M_2) = \int_{-\infty}^{\infty} p(c|m, \beta, M_2)p(\beta)\, d\beta \tag{29}$$

$$p(c|m, M_3) = \int_{-\infty}^{\infty} p(c|m, \gamma, M_1)p(\gamma)\, d\gamma \tag{30}$$

Here, $p(\alpha)$, $p(m_0)$, $p(\beta)$, and $p(\gamma)$ denote the Gaussian prior probability densities described above, e.g.,

$$p(\alpha) = \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left(-\frac{(\alpha - \bar{\alpha})^2}{2\sigma_\alpha^2}\right), \text{ etc.} \tag{31}$$

In the expressions for $p(c|m, M_i)$, integration over α, β, and γ can be performed analytically, e.g.,

$$p(c|m, M_2) = \frac{1}{\sigma_\beta \sigma_\epsilon \sqrt{\frac{2\pi}{\sigma_\beta^2} + \frac{2\pi f(m)^2}{\sigma_\epsilon^2}}} \exp\left(-\frac{(c - \bar{\beta} f(m))^2}{2(f(m)^2\sigma_\beta^2 + \sigma_\epsilon^2)}\right). \tag{32}$$

The integral over m_0 in the expression for $p(c|m, M_1)$ must be found numerically, however. In the present example, this integral is easily computed using standard numerical quadrature. If the integral had been too high dimensional, however, then a Monte Carlo scheme might be used instead [83]. We carry out these calculations using prior parameters listed in Table 1. The experimental uncertainty was set to $\sigma_\epsilon = 2.5/\text{nm}^2$, following [29].

To calculate the expected information gain U in the model indicator, as a function of the m value for a single additional experiment, we first substitute the prior and posterior model probabilities calculated above into (6). Then we take the expectation of this utility over the prior predictive distribution, as in (7), by integrating over the data c. More explicitly, we calculate:

$$U(m) = \int u(m, c)p(c|m)\, dc, \tag{33}$$

where the utility is

$$u(m, c) = \sum_{i=1}^{3} P(M_i|c, m) \log \frac{P(M_i|c, m)}{P(M_i)},$$

and the design-dependent prior predictive probability density is

$$p(c|m) = \sum_{i=1}^{3} P(M_i)p(c|m, M_i).$$

The integral in (33) formally is taken over $(-\infty, \infty)$, since this is the range of the prior predictive. Negative values of c are not physical, of course, but they are exceedingly rare: the mean predictions of models 1 and 2 are necessarily positive, and the Gaussian prior on γ in model 3 is almost entirely supported above zero. The Gaussian measurement noise ϵ can also lead to negative c values, but it too has a relatively small variance.

Model	Parameter	Standard Deviation
M_1	$\bar{\alpha} \approx 94/\text{nm}^2$	$\sigma_\alpha \approx 0.49/\text{nm}^2$
	$\bar{m_0} \approx 0.83$	$\sigma_{m_0} \approx 0.62$
M_2	$\bar{\beta} \approx 26/\text{nm}^2$	$\sigma_\beta \approx 4.2/\text{nm}^2$
M_3	$\bar{\gamma} \approx 4.5/\text{nm}^2$	$\sigma_\gamma = 2.0/\text{nm}^2$

Table 1: Prior model parameters.

Figure 10 plots $U(m)$ computed using all three models. For comparison, the figure also shows $U(m)$ found using only models 1 and 2, i.e., excluding the constant model $c = \gamma$. Values of m that maximize $U(m)$ are the best choices for an experiment to distinguish between models. When all three models are considered, $U(m)$ is greatest for high misfit, i.e., $m \approx 0.95$. By contrast, when only models 1 and 2 are considered, $U(m)$ is least in the high m limit. The reason for this difference is clear from comparing Figure 10 with Figure 8: models 1 and 2 predict comparable c at high m while model 3 predicts a markedly lower c. Thus, when all three models are considered, the value of $U(m)$ is high for $m \approx 0.95$ because a measurement at that m value makes it possible to distinguish models 1 and 2 from model 3. By contrast, when only models 1 and 2 are considered, $U(m)$ is least at high m because a measurement in that m range has limited value for distinguishing between models 1 and 2.

Putting aside model 3, $U(m)$ predicts greatest utility for an experiment carried out in the range $0.74 < m < 0.84$, i.e., in the vicinity of the minima of function $f(m)$. To understand the reason for the significance of this m range, it is important to realize that the plots in Figure 8 only show a single realization of models 1 and 2, namely those corresponding to $\alpha = \bar{\alpha}$, $m_0 = \bar{m_0}$, and $\beta = \bar{\beta}$ (the prior means on the parameters). Since we assume that α, m_0, and β are normally distributed, many other realizations of these models are possible. Figure 11 shows 100 different realizations of models 1 and 2 obtained by drawing α, m_0, and β at random from their prior distributions.

Figure 11 makes clear an important distinction between models 1 and 2 that is not apparent in Figure 8: model 1 exhibits an extreme sensitivity to its fitting parameters within the range of uncertainty of those parameters. In particular, the minimum in c predicted by model 1 may occur at many different m values. By contrast, model 2 is relatively less sensitive to its fitting parameters, especially for $0.74 < m \lesssim 0.84$. Unlike model 1, the locations of its minima are fixed. Thus, measuring a low c value for $0.74 < m \lesssim 0.84$ has the potential to exclude a large number of realizations of model 1, while measuring a high c value in that range essentially excludes model 2. Bayesian methods naturally capture this subtle aspect of experimental design without any special prior analysis of the competing models.

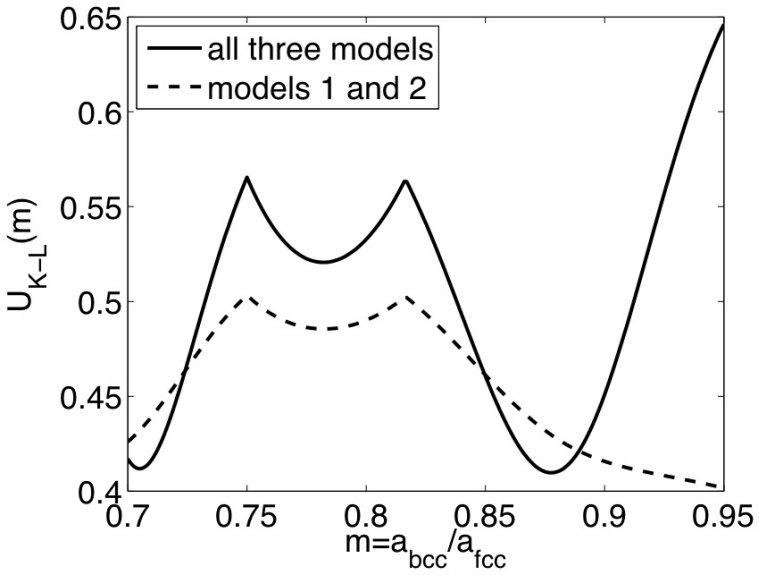


Figure 10: Expected information gain for model discrimination $U(m)$.

4 Outlook

The examples presented here demonstrate how the formalism of optimal Bayesian experimental design, coupled with information theoretic objectives, can be adapted to different questions of optimal data collection in materials science. In one example, we seek the best pair of experiments for inferring the parameters of a given model. In another example, we seek the single experiment that can best distinguish between competing models, where each model has a distinct form and a distinct set of uncertain parameters. Though simple, these examples also demonstrate the use of key computational and analytical tools, including the Monte Carlo estimation of expected information gain (in the nonlinear and non-Gaussian setting of our first example) and the use of reduced-order models (also in the first example).

Reduced order models (ROMs) are increasingly being recognized as crucial to materials science, especially computational materials design [39, 70, 94, 104]. The reason for their utility is clear: strictly speaking, the complete set of degrees of freedom describing a material is the complete set of positions and types of all its constituent atoms. This set defines a design space far too vast to explore. Even if mesoscale entities such as crystal defects (e.g., dislocations [45] or interfaces [91]) or microstructure [1] are used to define the degrees of freedom for design, the resulting design space may nevertheless remain too vast to examine comprehensively.

Therefore, it is crucial to identify only those degrees of freedom that significantly affect properties of interest (e.g., those affecting performance metrics in a design) and create a ROM to connect the two. Yet while formal methods of model order reduction

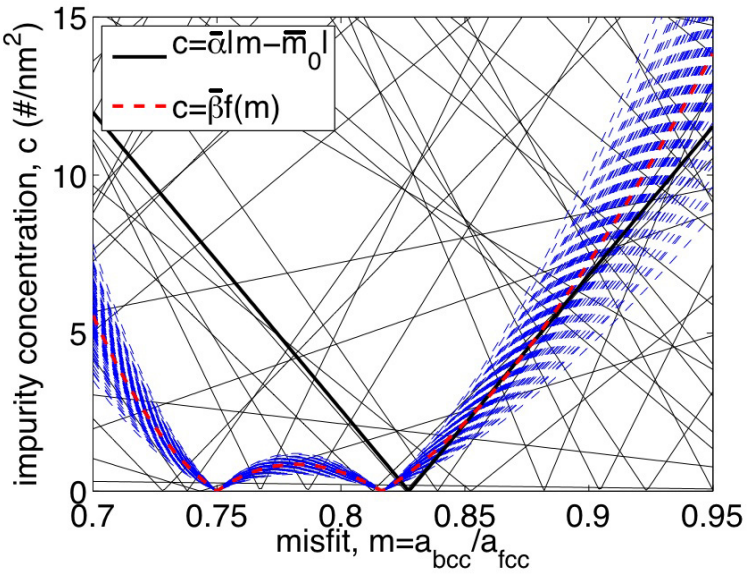


Figure 11: 100 different realizations of models 1 ($c = \alpha|m - m_0|$) and 2 ($c = \beta f(m)$) obtained by drawing parameters α, m_0, and β at random from their prior distributions. The thick lines plot the realizations of the models at the prior mean values of these parameters.

are well established in many fields of science and engineering (as reviewed in Section 2.3), the automated and systematic construction of ROMs in materials contexts is in an early stage of development. In practice, most ROMs in materials science are constructed "manually." The inherently collective and multiscale character of many materials-related phenomena calls for the development and validation of new methods of automatic model order reduction to address materials-specific challenges.

Surrogate or reduced-order models are also essential to making Bayesian inference computationally tractable, particularly inference with computationally intensive physics-based models. Indeed, the past several years have seen a steady stream of developments in model reduction for Bayesian inference, mostly in the applied mathematics, computational science, and statistics communities. These include many types of prior-based ROMs [68, 69, 67], posterior-focused approximations [59] and projection-based reduced-order models [27], hierarchical surrogates [23], and numerous other approaches [25, 85]. The utility of ROMs is increasingly being recognized in materials-related inference problems as well. For instance, the model film-substrate problem described in Section 3.1 relies on a ROM to circumvent computationally expensive forward problem evaluations, thereby making rapid inference of substrate properties tractable [2]. Earlier examples of this approach in materials science problems include [28], which, starting with what is effectively a ROM for the energy-temperature relation, inferred the melting point of Ti_2GaN. Data-fit surrogates constructed based on existing literature may also serve an analogous purpose to a ROM. Using such a

surrogate, [103] modeled the creep rupture life of Ni-base superalloys.

Similar to reduced-order modeling, the usefulness of Bayesian approaches is now becoming better recognized within the materials community. They can be applied to parameter inference and model inference, as demonstrated here, but also to problems involving prediction under uncertainty. For example, [55] used Bayesian inference to assess the uncertainty of cluster expansion methods for computing the internal energies of alloys. These authors point out that cluster expansions are also a kind of surrogate model—i.e., a ROM—and that uncertainty quantification should, among other goals, assess how well the surrogate reproduces the output of a more computationally expensive reference model.

Despite growing interest in Bayesian methods within the materials community, there are fewer examples of their application to experimental design. An early (yet very recent) effort is [4], which applies information-theoretic criteria and Bayesian methods to stress-strain response and texture evolution in polycrystalline solids. Nevertheless, opportunities for expanded application of optimal Bayesian experimental design abound in materials-related work. In particular, detailed and resource-intensive experiments such as those described in Section 3.2 are poised to benefit from it immensely. One potential hurdle to widespread adoption is the up-front investment of effort currently needed to understand and implement the associated mathematical formalism. Thus, expanded availability of user-friendly, well-documented, and multi-functional software [79] is likely to accelerate the adoption and integration of Bayesian experimental design into mainstream materials research.

Finally, we emphasize that optimal experimental design itself—not limited to the materials science context—is the topic of much current research. This research focuses on both formulational issues and on computational methodology. Examples of the latter include developing reduced-order or multi-fidelity models tailored to the needs of stochastic optimization, or devising more efficient estimators of expected information gain, using importance sampling, high-dimensional kernel density estimators, and other approaches. An interesting foundational challenge, on the other hand, involves understanding and accounting for model error or misspecification in optimal design. If the model relating parameters of interest to experimental observables is incomplete or under-resolved, how useful—or close to optimal—are the experiments designed according to this relationship? When a convergent sequence of models of differing fidelity is available (as in the ROM setting), then this question is more tractable. But if all available models are inadequate, many questions remain open. One promising approach to this challenge uses nonparametric statistical models, perhaps formulated in a hierarchical Bayesian manner, to account for interactions and inputs missing from the current model of the experiment. Sequential experimental design is also useful in this context, as successive batches of experiments can help uncover the unmodeled mismatch between a model and physical reality.

Sequential experimental design is useful much more broadly as well. Recall that in all the examples of this chapter, we designed a single batch of experiments all-at-once: even if the batch contained multiple experiments, we chose the design parameters before performing any of the experiments. Sequential design, in contrast, allows information from each experiment to influence the design of the next. The most widely used sequential approaches are greedy, where one designs the next batch of experi-

ments as if it were the final batch—using the current state of knowledge as the prior distribution, with design criteria similar to those used here. But greedy approaches are sub-optimal in general, as they do not account for the information to be gained from *future* experiments. An optimal approach can instead be obtained by formulating sequential experimental design as a problem of dynamic programming [9, 20, 7]. Making this dynamic programming approach computationally tractable, outside of specialized settings, remains a significant challenge.

References

[1] B. L. Adams, S. R. Kalidindi, and D. T. Fullwood, *Microstructure sensitive design for performance optimization*, Butterworth-Heinemann, 2012.

[2] R. Aggarwal, M. Demkowicz, and Y. Marzouk, *Bayesian inference of substrate properties from film behavior*, Modelling and Simulation in Materials Science and Engineering, 23 (2015), p. 015009.

[3] J. Aizenberg, A. Black, and G. Whitesides, *Controlling local disorder in self-assembled monolayers by patterning the topography of their metallic supports*, Nature, 394 (1998), pp. 868–871.

[4] S. Atamturktur, J. Hegenderfer, B. Williams, and C. Unal, *Selection criterion based on an exploration-exploitation approach for optimal design of experiments*, Journal of Engineering Mechanics, 141 (2014).

[5] A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs*, Oxford Statistical Science Series, Oxford University Press, 1992.

[6] G. Bayraksan and D. P. Morton, *Assessing solution quality in stochastic programs via sampling*, INFORMS Tutorials in Operations Research, 5 (2009), pp. 102–122.

[7] I. Ben-Gal and M. Caramanis, *Sequential doe via dynamic programming*, IIE Transactions, 34 (2002), pp. 1087–1100.

[8] J. Berger and L. Pericchi, *Objective Bayesian methods for model selection: introduction and comparison*, Model Selection (P.Lahiri, editor), IMS Lecture Notes – Monograph Series, 38 (2001), pp. 135–207.

[9] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, 3rd ed., 2007.

[10] I. Beyerlein, M. Demkowicz, A. Misra, and B. Uberuaga, *Defect-interface interactions*, Progress in Materials Science, (2015).

[11] D. Bhattacharyya, N. Mara, P. Dickerson, R. Hoagland, and A. Misra, *Transmission electron microscopy study of the deformation behavior of cu/nb and cu/ni nanoscale multilayers during nanoindentation*, Journal of Materials Research, 24 (2009), pp. 1291–1302.

[12] ——, *Compressive flow behavior of al–tin multilayers at nanometer scale layer thickness*, Acta Materialia, 59 (2011), pp. 3804–3816.

[13] W. Bollmann, *Crystal defects and crystalline interfaces*, Springer, 1970.

[14] N. Bowden, S. Brittain, A. Evans, J. Hutchinson, and G. Whitesides, *Spontaneous formation of ordered structures in thin films of metals supported on an elastomeric polymer*, Nature, 393 (1998), pp. 146–149.

[15] G. E. P. Box and H. L. Lucas, *Design of experiments in non-linear situations*, Biometrika, 46 (1959), pp. 77–90.

[16] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler, *A computational framework for infinite-dimensional bayesian inverse problems part i: The linearized case, with application to global seismic inversion*, SIAM Journal on Scientific Computing, 35 (2013), pp. A2494–A2523.

[17] T. Bui-Thanh, K. Willcox, and O. Ghattas, *Model reduction for large-scale systems with high-dimensional parametric input space*, SIAM Journal on Scientific Computing, 30 (2008), pp. 3270–3288.

[18] J. Cahn and J. Hilliard, *Free energy of a nonuniform system. i. interfacial free energy*, J. Chem. Phys, 28 (1958), pp. 258–267.

[19] P. R. Cantwell, M. Tang, S. J. Dillon, J. Luo, G. S. Rohrer, and M. P. Harmer, *Grain boundary complexions*, Acta Materialia, 62 (2014), pp. 1–48.

[20] B. P. Carlin, J. B. Kadane, and A. E. Gelfand, *Approaches for optimal sequential decision analysis in clinical trials*, Biometrics, (1998), pp. 964–975.

[21] K. Chaloner and I. Verdinelli, *Bayesian experimental design: A review*, Statistical Science, 10 (1995), pp. 273–304.

[22] S. Chaturantabut and D. C. Sorensen, *Nonlinear model reduction via discrete empirical interpolation*, SIAM Journal on Scientific Computing, 32 (2010), pp. 2737–2764.

[23] J. A. Christen and C. Fox, *MCMC using an approximation*, Journal of Computational and Graphical statistics, 14 (2005), pp. 795–810.

[24] P. Conrad and Y. M. Marzouk, *Adaptive Smolyak pseudospectral approximations*, SIAM Journal on Scientific Computing, 35 (2013), pp. A2643–A2670.

[25] P. Conrad, Y. M. Marzouk, N. Pillai, and A. Smith, *Accelerating asymptotically exact MCMC for computationally intensive models via local approximations*, Journal of the American Statistical Association, submitted (2014). arXiv:1402.1694.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 2nd ed., 2006.

[27] T. Cui, Y. M. Marzouk, and K. Willcox, *Data-driven model reduction for the Bayesian solution of inverse problems*, International Journal for Numerical Methods in Engineering, 102 (2015), pp. 966–990.

[28] S. Davis et al., *Bayesian inference as a tool for analysis of first-principles calculations of complex materials: an application to the melting point of ti2gan*, Modelling and Simulation in Materials Science and Engineering, 21 (2013), p. 075001.

[29] M. Demkowicz, D. Bhattacharyya, I. Usov, Y. Wang, M. Nastasi, and A. Misra, *The effect of excess atomic volume on he bubble formation at fcc–bcc interfaces*, Applied Physics Letters, 97 (2010), pp. 161903–161903.

[30] M. Demkowicz and R. Hoagland, *Structure of kurdjumov–sachs interfaces in simulations of a copper–niobium bilayer*, Journal of Nuclear Materials, 372 (2008), pp. 45–52.

[31] M. Demkowicz, R. Hoagland, B. Uberuaga, and A. Misra, *Influence of interface sink strength on the reduction of radiation-induced defect concentrations and fluxes in materials with large interface area per unit volume*, Physical Review B, 84 (2011), p. 104102.

[32] M. Demkowicz, A. Misra, and A. Caro, *The role of interface structure in controlling high helium concentrations*, Current Opinion in Solid State and Materials Science, 16 (2012), pp. 101–108.

[33] M. J. Demkowicz, J. Wang, and R. G. Hoagland, *Interfaces between dissimilar crystalline solids*, Dislocations in solids, 14 (2008), pp. 141–205.

[34] M. Eldred, S. Giunta, and S. Collis, *Second-order corrections for surrogate-based optimization with model hierarchies*. AIAA Paper 2004-4457, in Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2004.

[35] D. Eyre, *An unconditionally stable one-step scheme for gradient systems*. Unpublished manuscript, University of Utah, Salk Lake City, UT, June 1998.

[36] I. Ford, D. M. Titterington, and K. Christos, *Recent advances in nonlinear experimental design*, Technometrics, 31 (1989), pp. 49–60.

[37] E. Fu, N. Li, A. Misra, R. Hoagland, H. Wang, and X. Zhang, *Mechanical properties of sputtered cu/v and al/nb multilayer films*, Materials Science and Engineering: A, 493 (2008), pp. 283 – 287. Mechanical Behavior of Nanostructured Materials, a Symposium Held in Honor of Carl Koch at the {TMS} Annual Meeting 2007, Orlando, Florida.

[38] E. Fu, A. Misra, H. Wang, L. Shao, and X. Zhang, *Interface enabled defects reduction in helium ion irradiated cu/v nanolayers*, Journal of Nuclear Materials, 407 (2010), pp. 178–188.

[39] L. D. Gabbay and S. Senturia, *Computer-aided generation of nonlinear reduced-order dynamic macromodels. i. non-stress-stiffened case*, Microelectromechanical Systems, Journal of, 9 (2000), pp. 262–269.

[40] T. Gerstner and M. Griebel, *Dimension-adaptive tensor-product quadrature*, Computing, 71 (2003), pp. 65–87.

[41] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer, 1991.

[42] J. Ginebra, *On the measure of the information in a statistical experiment*, Bayesian Analysis, 2 (2007), pp. 167–212.

[43] M. Grepl, Y. Maday, N. Nguyen, and A. Patera, *Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations*, Mathematical Modelling and Numerical Analysis (M2AN), 41 (2007), pp. 575–605.

[44] G. E. Hilley, R. Bürgmann, P.-Z. Zhang, and P. Molnar, *Bayesian inference of plastosphere viscosities near the kunlun fault, northern tibet*, Geophysical Research Letters, 32 (2005), pp. n/a–n/a.

[45] J. Hirth and J. Lothe, *Theory of Dislocations*, Wiley, 1992.

[46] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, *Bayesian model averaging: A tutorial*, Statistical Science, 14 (1999), pp. 382–417.

[47] P. Holmes, J. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, UK, 1996.

[48] X. Huan and Y. M. Marzouk, *Simulation-based optimal Bayesian experimental design for nonlinear systems*, Journal of Computational Physics, 232 (2013), pp. 288–317.

[49] X. Huan and Y. M. Marzouk, *Gradient-based stochastic optimization methods in Bayesian experimental design*, International Journal for Uncertainty Quantification, 4 (2014), pp. 479–510.

[50] H. Jeffreys, *An invariant form for the prior probability in estimation problems*, Proceedings of the Royal Society, (1946).

[51] D. R. Jones, M. Schonlau, and W. J. Welch, *Efficient global optimization of expensive black-box functions*, Journal of Global optimization, 13 (1998), pp. 455–492.

[52] A. Kashinath, A. Misra, and M. Demkowicz, *Stable storage of helium in nanoscale platelets at semicoherent interfaces*, Physical review letters, 110 (2013), p. 086101.

[53] M. C. Kennedy and A. O'Hagan, *Bayesian calibration of computer models*, Journal of the Royal Statistical Society. Series B (Statistical Methodology), 63 (2001), pp. 425–464.

[54] J. Kiefer and J. Wolfowitz, *Stochastic estimation of the maximum of a regression function*, The Annals of Mathematical Statistics, 23 (1952), pp. 462–466.

[55] J. Kristensen and N. J. Zabaras, *Bayesian uncertainty quantification in the evaluation of alloy properties with the cluster expansion method*, Computer Physics Communications, 185 (2014), pp. 2885–2892.

[56] H. Kushner and G. Yin, *Stochastic approximation and recursive algorithms and applications*, Applications of mathematics, Springer, 2003.

[57] O. P. Le Maître and O. M. Knio, *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*, Springer, 2010.

[58] J. Lewandowski and A. Greer, *Temperature rise at shear bands in metallic glasses*, Nature Materials, 5 (2006), pp. 15–18.

[59] J. Li and Y. M. Marzouk, *Adaptive construction of surrogates for the Bayesian solution of inverse problems*, SIAM Journal on Scientific Computing, 36 (2014), pp. A1163–A1186.

[60] N. Li, E. Fu, H. Wang, J. Carter, L. Shao, S. Maloy, A. Misra, and X. Zhang, *He ion irradiation damage in fe/w nanolayer films*, Journal of Nuclear Materials, 389 (2009), pp. 233–238.

[61] N. Li, M. Martin, O. Anderoglu, A. Misra, L. Shao, H. Wang, and X. Zhang, *He ion irradiation damage in al/ nb multilayers*, Journal of Applied Physics, 105 (2009), p. 123522.

[62] N. Li, J. Wang, J. Huang, A. Misra, and X. Zhang, *In situ tem observations of room temperature dislocation climb at interfaces in nanolayered al/nb composites*, Scripta Materialia, 63 (2010), pp. 363–366.

[63] D. V. Lindley, *Bayesian Statistics, A Review*, Society for Industrial and Applied Mathematics (SIAM), 1972.

[64] T. J. Loredo, *Rotating stars and revolving planets: Bayesian exploration of the pulsating sky*, in Bayesian Statistics 9: Proceedings of the Nineth Valencia International Meeting, Oxford University Press, 2010, pp. 361–392.

[65] T. J. Loredo and D. F. Chernoff, *Bayesian adaptive exploration*, in Statistical Challenges of Astronomy, Springer, 2003, pp. 57–69.

[66] D. J. MacKay, *Information theory, inference, and learning algorithms*, vol. 7, Citeseer, 2003.

[67] Y. M. Marzouk and H. N. Najm, *Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems*, Journal of Computational Physics, 228 (2009), pp. 1862–1902.

[68] Y. M. Marzouk, H. N. Najm, and L. A. Rahn, *Stochastic spectral methods for efficient Bayesian solution of inverse problems*, Journal of Computational Physics, 224 (2007), pp. 560–586.

[69] Y. M. Marzouk and D. Xiu, *A stochastic collocation approach to Bayesian inference in inverse problems*, Communications in Computational Physics, 6 (2009), pp. 826–847.

[70] J. E. Mehner, L. D. Gabbay, and S. D. Senturia, *Computer-aided generation of nonlinear reduced-order dynamic macromodels. ii. stress-stiffened case*, Microelectromechanical Systems, Journal of, 9 (2000), pp. 270–278.

[71] Y. Mishin, M. Asta, and J. Li, *Atomistic modeling of interfaces and their impact on microstructure and properties*, Acta Materialia, 58 (2010), pp. 1117–1151.

[72] A. Misra, M. Demkowicz, X. Zhang, and R. Hoagland, *The radiation damage tolerance of ultra-high strength nanolayered composites*, Jom, 59 (2007), pp. 62–65.

[73] A. Misra, J. Hirth, and R. Hoagland, *Length-scale-dependent deformation mechanisms in incoherent metallic multilayered composites*, Acta Materialia, 53 (2005), pp. 4817–4824.

[74] T. E. Mitchell, Y. C. Lu, A. J. G. Jr., M. Nastasi, and H. Kung, *Structure and mechanical properties of copper/niobium multilayers*, Journal of the American Ceramic Society, 80 (1997), pp. 1673–1676.

[75] J. I. Myung and M. A. Pitt, *Optimal experimental design for model discrimination*, Psychological Review, 116 (2009), pp. 499–518.

[76] J. A. Nelder and R. Mead, *A simplex method for function minimization*, The Computer Journal, 7 (1965), pp. 308–313.

[77] A. Noor and J. Peters, *Reduced basis technique for nonlinear analysis of structures*, AIAA Journal, 18 (1980), pp. 455–462.

[78] G. Odette, M. Alinger, and B. Wirth, *Recent developments in irradiation-resistant steels*, Annu. Rev. Mater. Res., 38 (2008), pp. 471–503.

[79] M. Parno, P. Conrad, A. Davis, and Y. M. Marzouk, *MIT uncertainty quantification (MUQ) library*. http://bitbucket.org/mituq/muq.

[80] V. Picheny, D. Ginsbourger, Y. Richet, and G. Caplin, *Quantile-based optimization of noisy computer experiments with tunable precision*, Technometrics, 55 (2013), pp. 2–13.

[81] Z.-Q. Qu, *Model Order Reduction Techniques with Applications in Finite Element Analysis: With Applications in Finite Element Analysis*, Springer Science & Business Media, 2004.

[82] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.

[83] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Verlag, 2004.

[84] K. J. Ryan, *Estimating expected information gains for experimental designs with application to the random fatigue-limit model*, Journal of Computational and Graphical Statistics, 12 (2003), pp. 585–603.

[85] C. Schwab and A. M. Stuart, *Sparse deterministic approximation of Bayesian inverse problems*, Inverse Problems, 28 (2012), p. 045003.

[86] S. Shao, H. Zbib, I. Mastorakos, and D. Bahr, *The void nucleation strengths of the cu–ni–nb-based nanoscale metallic multilayers under high strain rate tensile loadings*, Computational Materials Science, 82 (2014), pp. 435–441.

[87] A. Shapiro, *Asymptotic analysis of stochastic programs*, Annals of Operations Research, 30 (1991), pp. 169–186.

[88] L. Sirovich, *Turbulence and the dynamics of coherent structures. Part 1: Coherent structures*, Quarterly of Applied Mathematics, 45 (1987), pp. 561–571.

[89] A. Solonen, H. Haario, and M. Laine, *Simulation-based optimal design using a response variance criterion*, Journal of Computational and Graphical Statistics, 21 (2012), pp. 234–252.

[90] J. C. Spall, *An overview of the simultaneous perturbation method for efficient optimization*, Johns Hopkins APL Technical Digest, 19 (1998), pp. 482–492.

[91] A. Sutton and R. Balluffi, *Interfaces in crystalline materials*, Monographs on the physics and chemistry of materials, Clarendon Press, 1995.

[92] H. Ullmaier, *The influence of helium on the bulk properties of fusion reactor structural materials*, Nuclear Fusion, 24 (1984), p. 1039.

[93] J. van den Berg, A. Curtis, and J. Trampert, *Optimal nonlinear Bayesian experimental design: an application to amplitude versus offset experiments*, Geophysical Journal International, 155 (2003), pp. 411–421.

[94] A. Vattré, N. Abdolrahim, K. Kolluri, and M. Demkowicz, *Computational design of patterned interfaces using reduced order models*, Scientific reports, 4 (2014).

[95] Y.-C. Wang, A. Misra, and R. Hoagland, *Fatigue properties of nanoscale cu/nb multilayers*, Scripta materialia, 54 (2006), pp. 1593–1598.

[96] B. P. Weaver, B. J. Williams, C. M. Anderson-Cook, and D. M. Higdon, *Computational enhancements to Bayesian design of experiments using Gaussian processes*, Bayesian Analysis, (2015).

[97] Q. Wei, N. Li, N. Mara, M. Nastasi, and A. Misra, *Suppression of irradiation hardening in nanoscale v/ag multilayers*, Acta Materialia, 59 (2011), pp. 6331–6340.

[98] Q. Wei and A. Misra, *Transmission electron microscopy study of the microstructure and crystallographic orientation relationships in v/ag multilayers*, Acta Materialia, 58 (2010), pp. 4871–4882.

[99] B. Williams, D. Higdon, J. Gattiker, L. Moore, M. McKay, and S. Keller-McNulty, *Combining experimental data and computer simulations, with an application to flyer plate experiments*, Bayesian Analysis, 1 (2006), pp. 765–792.

[100] D. Xiu, *Efficient collocational approach for parametric uncertainty analysis*, Communications in Computational Physics, 2 (2007), pp. 293–309.

[101] D. Xiu and J. S. Hesthaven, *High-order collocation methods for differential equations with random inputs*, SIAM Journal of Scientific Computing, 27 (2005), pp. 1118–1139.

[102] L. Yarin, *The Pi-Theorem: Applications to Fluid Mechanics and Heat and Mass Transfer*, vol. 1, Springer, 2012.

[103] Y. Yoo, C. Jo, and C. Jones, *Compositional prediction of creep rupture life of single crystal ni base superalloy by bayesian neural network*, Materials Science and Engineering, (2001), pp. 22–29.

[104] D. Yuryev and M. Demkowicz, *Computational design of solid-state interfaces using o-lattice theory: An application to mitigating helium-induced damage*, Applied Physics Letters, 105 (2014), p. 221601.

[105] X. Zhang, E. Fu, A. Misra, and M. Demkowicz, *Interface-enabled defect reduction in he ion irradiated metallic multilayers*, Jom, 62 (2010), pp. 75–78.

[106] S. Zinkle and N. Ghoniem, *Operating temperature windows for fusion reactor structural materials*, Fusion Engineering and design, 51 (2000), pp. 55–71.