

## Review

# Knowledge-driven learning, optimization, and experimental design under uncertainty for materials discovery

Xiaoning Qian,<sup>1,3,\*</sup> Byung-Jun Yoon,<sup>1,3</sup> Raymundo Arróyave,<sup>2</sup> Xiaofeng Qian,<sup>2</sup> and Edward R. Dougherty<sup>1</sup>

<sup>1</sup>Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX 77843, USA

<sup>2</sup>Department of Materials Science & Engineering, Texas A&M University, College Station, TX 77843, USA

<sup>3</sup>Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973, USA

\*Correspondence: [xqian@ece.tamu.edu](mailto:xqian@ece.tamu.edu)

<https://doi.org/10.1016/j.patter.2023.100863>

**THE BIGGER PICTURE** Thanks to the rapid advances in artificial intelligence, AI for science (AI4Science) has emerged as one of the new promising research directions for modern science and engineering. In this review, we focus on recent efforts to develop knowledge-driven Bayesian learning and experimental design methods for accelerating the discovery of novel functional materials as well as enhancing the understanding of composition-process-structure-property relationships. We specifically discuss the challenges and opportunities in integrating prior scientific knowledge and physics principles with AI and machine learning (ML) models for accelerating materials and knowledge discovery. The current state-of-the-art methods in knowledge-based prior construction, model fusion, uncertainty quantification, optimal experimental design, and symbolic regression are detailed in the review, along with several detailed case studies and results in materials discovery.

## SUMMARY

Significant acceleration of the future discovery of novel functional materials requires a fundamental shift from the current materials discovery practice, which is heavily dependent on trial-and-error campaigns and high-throughput screening, to one that builds on knowledge-driven advanced informatics techniques enabled by the latest advances in signal processing and machine learning. In this review, we discuss the major research issues that need to be addressed to expedite this transformation along with the salient challenges involved. We especially focus on Bayesian signal processing and machine learning schemes that are uncertainty aware and physics informed for knowledge-driven learning, robust optimization, and efficient objective-driven experimental design.

## INTRODUCTION

Accelerating the development of novel functional materials with desirable properties is a worldwide imperative because it can facilitate advances in diverse fields across science, engineering, and biomedicine with significant potential contributions to economic growth. For example, the US Materials Genome Initiative (MGI) calls for cutting the cost and time for bringing new materials from discovery to deployment by half by integrating experiments, computer simulations, and data analytics.<sup>1,2</sup> However, the current prevailing practice in materials discovery primarily relies on trial-and-error experimental campaigns or high-throughput virtual screening approaches by computational simulations, neither of which can efficiently explore the huge materials design space to develop materials that possess targeted functional properties.

To fundamentally shift the current trial-and-error practice to an efficient informatics-driven practice, there have been increasing research efforts to develop signal processing (SP) and machine learning (ML) methods that may ultimately enable autonomous materials discovery and expedite the discovery of novel materials at a substantially reduced cost and time.<sup>3–5</sup> When applying SP and ML methods in materials science, several unique challenges arise, which include (1) a limited amount of data (if any) for investigating and exploring new materials systems, (2) data of varying and inconsistent quality because of technical limitations and a lack of common profiling prototypes, (3) significant complexity and uncertainty in existing computational simulation and surrogate models,<sup>6</sup> and (4) incomplete domain knowledge.

To cope with the aforementioned challenges and to effectively discover novel functional materials with the desired target properties, robust decision-making strategies are critical for efficient



exploration of the immense materials design space through effective learning, optimization, and experimental design under significant uncertainty. Directly applying existing data-driven SP and ML methods falls short of achieving these goals, and comprehensive theoretical and methodological developments tailored to address these unique challenges are crucial. Some salient issues that need to be addressed include the following:

- (1) Knowledge-based prior construction: mapping scientific knowledge into a prior distribution that reflects model uncertainty to alleviate the issues stemming from data scarcity
- (2) Model fusion: updating the prior distribution to a posterior distribution with multiple uncertain models and data sources of different data quality
- (3) Uncertainty quantification (UQ): quantification of the cost of uncertainty relative to one or more objectives for efficient materials discovery
- (4) Optimization under uncertainty (OUU): derivation of an optimal operator from the posterior distribution
- (5) Optimal experimental design (OED): efficient experimental design and data acquisition schemes to improve the model to explore the materials design space more effectively
- (6) Knowledge discovery: closing the knowledge gap in the current model, such as composition-process-structure-property (CPSP) relationships relevant to the materials discovery objectives, based on newly acquired data or increased model knowledge

In this article, we will review the recent advances related to the aforementioned research issues. Especially, we will provide an in-depth review of Bayesian SP and ML approaches for knowledge-driven learning, objective-based UQ, and efficient experimental design for materials discovery under substantial model and data uncertainties. The core foundation underlying these strategies is a Bayesian framework that enables mathematical representation of the model and data uncertainties, encoding available domain knowledge into a Bayesian prior, seamlessly integrating experimental (or simulation) data with the domain knowledge to obtain a posterior, quantifying the impact of the uncertainty on the objective, and effective design of strategies that can reduce this uncertainty. It is important to note that the guiding principle of the aforementioned Bayesian framework is to have a (knowledge-based) prior represent an uncertainty class of *models*. The prior characterizes the state of our knowledge about the model representing the system, based on which we can design operators to achieve the scientific objectives. Artificial intelligence for science (AI4Science) has emerged as an enormous modern research field. Because of the rapidly evolving nature of this field, it is challenging to provide a comprehensive review of all ongoing research efforts, and the readers are strongly encouraged to refer to additional resources, including recent publications in AI4Science<sup>7–10</sup> as well as those in AI/ML-augmented materials discovery.<sup>11–13</sup>

In the following sections, we first introduce the UQ framework that encompasses the various components in knowledge-driven learning, optimization, and experimental design. This will be followed by in-depth discussion of the individual research themes,

where we will review the latest research results along these directions.

### Bayesian learning, UQ, and experimental design

Engineering generally aims at optimization to achieve operational objectives when studying complex systems. Because all but very simple systems must account for randomness, modern engineering may be defined as the study of optimal operators on random processes. Besides the mathematical and computational challenges that arise with classical system identification (learning) and operator optimization (control or filtering, for example) problems, such as nonstationary processes, high dimensions, and nonlinear operators, another profound issue is model uncertainty. For instance, with linear filtering there may be incomplete knowledge regarding the covariance functions or power spectra in the case of Wiener filtering. In such cases, not only must optimization of the operator (i.e., filter in this example) be relative to the original cost function but also relative to an uncertainty class of random processes. This naturally leads to the need for postulation of a new cost function that integrates the original cost function with the model uncertainty. If there is a prior (or posterior) distribution governing the likelihood of a model within the uncertainty class, then one can choose an operator that minimizes the *expected cost* over all possible models in the uncertainty class. In what follows, we first lay out the mathematical foundations pertinent to quantifying and handling model uncertainty and then review relevant existing literature, with recent efforts focusing on materials science research.

### Mathematical backgrounds

The design of optimal operators can take different forms depending on the random process constituting the scientific model and the operator class of interest. The operators might be filters, classifiers, or controllers. The underlying random process might be a random signal/image for filtering, a feature-label distribution for classification, or a Markov process for control. Optimal operator design involves a mathematical model representing the underlying (materials) system and a class of operators from which the best operator that minimizes the cost function reflecting the objective should be selected. It takes the general form

$$\psi_{\text{opt}} = \underset{\psi \in \Psi}{\operatorname{argmin}} C(\psi), \quad (\text{Equation 1})$$

where  $\Psi$  is the operator class, and  $C(\psi)$  is the cost of applying operator  $\psi$  on the system. The genesis of such an operator design formulation can be traced back to the Wiener-Kolmogorov theory in SP for optimal linear filters developed in the 1930s,<sup>14,15</sup> where the operational objective is to recover the underlying signals given noisy observations with the minimum mean squared error (MSE). In this class of filtering problems, the operators mentioned above are filters. The underlying system can be modeled by a joint random process  $(X(t), Y(s))$ ,  $t \in T, s \in S$ . Optimal filtering involves estimating the signal  $Y(s)$  at time  $s$  via a filter  $\psi$  given observations  $\{X(t)\}_{t \in T}$ . A filter  $\psi \in \Psi$  is a mapping on the space of possible observed signals, and a cost function takes the form  $C(Y(s), \hat{Y}(s))$ , with  $\hat{Y}(s) = \psi(X)(s)$ . For fixed  $s \in S$ , an optimal filter is defined by Equation 1 with  $C(\psi) = C(Y(s), \psi(X)(s)) = E[(Y(s) - \psi(X)(s))^2]$ . Similar

operator design formulations have been adopted in control<sup>16</sup> and, more recently, in ML,<sup>17</sup> where the corresponding operators are controllers that can desirably alter the system behavior or predictive models for system properties of interest (e.g., classifiers). For example, the operator may be a predictor that tries to characterize the property of a given material based on input features (such as its composition and structure).

When the true model is not known with certainty, it would be prudent to consider the entire uncertainty class  $\Theta$  of possible models that contains that true model  $\theta \in \Theta$ , where  $\theta$  may be typically a parameter vector specifying the model rather than aiming at accurate inference of the true model. Given  $\Theta$ , the goal would then be to design a robust operator that guarantees good performance over all possible models. For example, there have been significant research efforts taking a minimax strategy to design robust operators:

$$\psi_{\text{minimax}}^{\Theta} = \underset{\psi \in \Psi}{\text{argmin}} \max_{\theta \in \Theta} C_{\theta}(\psi), \quad (\text{Equation 2})$$

Where  $C_{\theta}(\psi)$  characterizes the cost of the operator  $\psi$  for model  $\theta$ . Taking filtering as an example,  $C_{\theta}(\psi) = C(Y(s; \theta), \psi(X)(s; \theta))$ , where  $\theta$  denotes the model parameters for the signal and observation random processes. Such a minimax robust strategy is risk averse because it aims to find an operator whose worst performance over an uncertainty class of models  $\Theta$  is the best among all operators in  $\Psi$ <sup>18,19</sup>. Minimax robustness has been applied in many optimization frameworks; for example, for filtering<sup>20–23</sup> with a general formulation in the context of game theory,<sup>24</sup> as well as recently in ML.<sup>25,26</sup> One critical downside of minimax robustness is that, in avoiding the worst-case scenario, the average performance of the designed operator can be poor, in particular when the prior knowledge about the uncertainty class  $\Theta$  is available and the worst-case model is unlikely. There has been extensive research on alleviating this potential issue by developing risk measures, such as conditional value at risk in a recently proposed risk quadrangle scheme,<sup>18</sup> to achieve better trade-off between the attainment of the operational objective and the aversion of potential risk because of uncertainty.

Unlike such minimax robust strategies, we focus on Bayesian robust strategies that try to optimize the expected performance in the presence of uncertainty. This leads to the design of the intrinsically Bayesian robust (IBR) operator, which is defined as

$$\psi_{\text{IBR}}^{\Theta} = \underset{\psi \in \Psi}{\text{argmin}} E_{\pi} [C_{\theta}(\psi)], \quad (\text{Equation 3})$$

where the expectation is with respect to a prior probability distribution  $\pi(\theta)$  of the uncertain model  $\theta \in \Theta$ . While not as risk averse as minimax robust operators, these Bayesian robust operators guarantee optimal performance on average. The prior  $\pi(\theta)$  probabilistically characterizes our prior knowledge as to which models are more likely to be the true model than the others. If there is no prior knowledge beyond the uncertainty class itself, then a uniform (non-informative) prior may be used.

### Related works

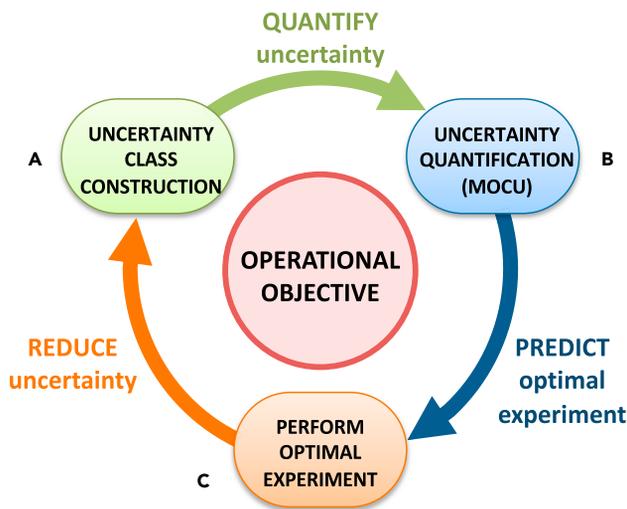
Before we delve into the Bayesian framework for learning, UQ, and experimental design, here we provide a literature review of related topics. We first review the history of operator design, in

particular related to filtering, classification, and control. For optimal operator design in filtering, Kalman-Bucy recursive filtering was proposed in the 1960s<sup>27</sup> after the Wiener filter.<sup>14,15</sup> Optimal control began in the 1950s, as did classification as now understood. In all three areas, it was quickly recognized that often the underlying scientific model would not be known—hence the development of adaptive linear/Kalman filters and adaptive controllers.<sup>28,29</sup> Classification became dependent on classification rules that make no effort to estimate the true feature-label distribution.<sup>17</sup> From the perspective of model uncertainty classes, control theorists delved into Bayesian robust control for Markov decision processes in the work of Bellman and Kalaba,<sup>30</sup> Silver,<sup>31</sup> and Martin<sup>30–32</sup> in the 1960s, but computation was prohibitive, and adaptive methods prevailed. Optimal linear filtering was approached via minimax in the late 1970s in the work of Kuznetsov,<sup>20</sup> Kassam and Lim,<sup>21</sup> Poor,<sup>22</sup> and Verdu and Poor.<sup>24</sup> Model-constrained Bayesian robust (MCBR) MSE linear filtering and classification appeared in the early 2000s.<sup>33,34</sup>

When considering uncertainty in optimization, there has been extensive research in designing different risk metrics for UQ. For example, different values at risk<sup>18</sup> and quantities of interest (QoIs)<sup>19</sup> have been proposed based on different statistics when modeling random processes or the corresponding model parameters as random variables, including the ones based on prediction variance<sup>35</sup> and predictive entropy.<sup>36,37</sup> With these risk metrics, different robust operator design strategies have been studied to derive risk-averse operators that can achieve good performance.<sup>19,25,26,35,37–39</sup> While introducing additional risk metrics enables balancing the trade-off between the operational objectives and the potential risk (or regret) because of uncertainty, incorporating different metrics with different strategies can be subjective. For example, there may be large predictive variance or entropy, but it may not always directly affect the operational objectives and, thereafter, consequent decision-making.

Bayesian learning and experimental design offers one solution for robust design under uncertainty.<sup>19,40–44</sup> In this framework, UQ can be naturally measured by the loss of performance because of the utilization of a robust operator to cope with uncertainty. This leads to an experimental design strategy where experiments are selected to optimally reduce this performance loss, following the early thinking of Bayesian robust filtering and control.<sup>15,30–32</sup> Such an experimental design framework, rooted in the foundation of modern engineering, closes the loop from scientific knowledge on a complex system, models for the complex system under uncertainty, data generated by the system, and experiments to enhance the current system knowledge to better attain the objectives. In this paper, we focus on this closed-loop framework, which distinguishes itself from (1) other existing schemes that are purely data driven<sup>45–48</sup> or (2) experimental design frameworks based on high-throughput simulations, such as II 4U<sup>49</sup> and DAKOTA.<sup>50</sup>

Data-driven frameworks heavily depend on the availability of data, upon which “black box” surrogate models are trained. They typically model the operators (used to achieve the objectives) of interest rather than modeling the system itself when designing experiments. For example, in materials discovery, many existing methods rely on Bayesian optimization (BO),



**Figure 1.** Illustration of the knowledge-driven optimal experimental design (OED) cycle for materials discovery

which uses Gaussian processes (GPs) as surrogate models to directly approximate the target materials properties as “black box” functions.<sup>51–53</sup> While BO may be useful for optimizing the properties, the acquired data do not improve our knowledge regarding the materials system. As a consequence, there is often a scientific gap in making prior assumptions on these “black box” models and their uncertainty.<sup>54</sup> To better integrate scientific knowledge, such as materials’ process-structure-property relationships, as detailed under “Knowledge-driven prior construction,” the model uncertainty should be directly imposed on the system model that incorporates inter-relationships among the underlying random processes. For simulation-based frameworks, including II 4U and DAKOTA, UQ, sensitivity analysis, and experimental design are mostly based on forward model simulations, which do not provide a natural way to propagate the data generated by the selected experiments back to the system to fill the gap in our system knowledge and to improve the current model, which is precisely what our proposed paradigm aims to do. The emphasis here is that (1) the uncertainty is placed directly on the underlying random process (i.e., current knowledge regarding the materials system) and not on surrogate models that reflect operational performance on this uncertain process and that (2) the experimental design is centered around attaining specific objectives. A wide range of approaches can emerge, depending on the assumptions made regarding the uncertainty class, action space, and experiment space. Popular Bayesian experimental design policies, such as knowledge gradient (KG)<sup>46,47</sup> and efficient global optimization (EGO),<sup>45</sup> are special cases in this framework under their modeling assumptions. These approaches often adopt generic surrogate models with the uncertainty placed on the reward function; therefore, there is no direct connection between the prior model assumptions and the underlying process/system.

Because of these characteristics, Bayesian frameworks have been increasingly used to address a wide range of materials discovery problems.<sup>55–59</sup> BO’s ability to balance the exploration

and exploitation is ideally suited in materials discovery tasks because queries to the materials design space (either through computations or experiments) are extremely resource intensive. Most approaches focused on materials discovery are myopic in the sense that increased knowledge of the materials space being explored is not necessarily part of the objective. In other cases, Bayesian learning is used to increase knowledge of the physics underlying observed physical phenomena without much attention being put on improving the materials’ performance relative to the existing state of the art.<sup>60–64</sup> In materials discovery applications, the complexity and stochasticity because of substantial model and data uncertainty call for SP and ML approaches in a Bayesian setting that can provide a unified closed-loop framework for objective-based learning and optimal design of robust operators and effective experiments under uncertainty. This is illustrated in Figure 1.

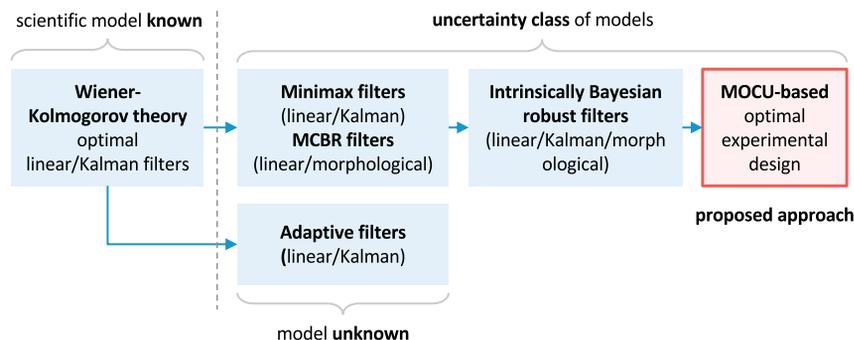
### IBR operator and mean objective cost of uncertainty (MOCU)-based UQ

In this section, we focus on the objective-based UQ (objective-UQ) framework using the MOCU,<sup>65,66</sup> which measures the expected loss with respect to the final operational objective because of the model uncertainty. Uncertainty is directly imposed on the model representing the underlying system and not on the parameters of the operator, as typically done in the ML community. Because the uncertainty is on the system model, reduction of this uncertainty inevitably leads to improving our knowledge regarding the system, leaving no discrepancy between what is learned (through data acquisition or experiments) about the model and what we know about the underlying system (and the relevant science).

Consider a stochastic model  $\mathcal{M}$  with uncertainty class  $\Theta$  composed of possible parameter vectors. Let  $C$  be a cost function and  $\Psi$  a class of operators on  $\mathcal{M}$ . For each operator  $\psi \in \Psi$ ,  $C_\theta(\psi)$  denotes the cost of applying  $\psi$  on the model parametrized by  $\theta \in \Theta$ . An IBR operator on  $\mathcal{M}$  is an operator  $\psi_{\text{IBR}} \in \Psi$  so that the expected value over  $\Theta$  of the cost  $C_\theta(\psi)$  is minimized by  $\psi_{\text{IBR}}$  as formulated in Equation 3,<sup>67</sup> the expected value being with respect to a prior probability distribution  $\pi(\theta)$  capturing model uncertainty over  $\Theta$ . Here, each parameter vector  $\theta \in \Theta$  corresponds to a model, and  $\pi(\theta)$  quantifies the likelihood that a model is  $\Theta$  and therefore reflects prior knowledge. If there is no prior knowledge beyond the uncertainty class itself, then it is taken to be uniform with all models being equally likely. Given a data sample  $S$  sampled independently from the full model, the IBR theory can be used with a posterior distribution  $\pi^*(\theta) = \pi(\theta|S)$ , giving the optimal Bayesian operator. Because of the optimality of the IBR operator  $\psi_{\text{IBR}}$  over  $\Theta$ ,  $E_\Theta[C_\theta(\psi_{\text{IBR}})] \leq E_\Theta[C_\theta(\psi)]$  for any operator  $\psi$ . For  $\theta \in \Theta$ , the objective cost of uncertainty relative to  $\theta$  is the difference between  $C_\theta(\psi_{\text{IBR}})$  and  $C_\theta(\psi_\theta)$ . Averaging this loss differential provides our basic UQ, the MOCU.<sup>65</sup>

$$\text{MOCU}(\Theta) = E_\theta[C_\theta(\psi_{\text{IBR}}) - C_\theta(\psi_\theta)], \quad (\text{Equation 4})$$

where  $\psi_\theta$  denotes the optimal operator with respect to the model specified by the model parameter  $\theta$ . The expectation is computed with respect to the distribution  $\pi(\theta)$  of the model  $\theta$  in the uncertainty class  $\Theta$ .



**Figure 2. Illustration of the historical context of the intrinsically Bayesian robust (IBR) framework and the concept of mean objective cost of uncertainty (MOCU)**

While the entropy of the prior (or posterior) has been commonly used to measure model uncertainty, entropy, however, does *not* focus on the objective. In other words, there may be large entropy, but it may not directly affect the operational objective because it may not affect the expected cost in Equation 13. Unlike entropy, MOCU aims to quantify the uncertainty that practically “matters” as it pertains to a specific objective (Figure 1).

IBR (with a prior) and optimal Bayesian (with an updated posterior given observed data) operator design have been applied in modern engineering, statistics, and ML. Based on different operators (for example, for filtering, classification, and control) and their corresponding cost functions, the research focus has been mostly on solving the corresponding inference and optimization problems, known as Bayesian learning or Bayesian inverse problems.<sup>68–70</sup> When systems understanding and operator design are the objectives of modeling complex systems, Bayesian experimental design and decision-making are often with respect to the uncertainty class of models and the cost function related to the operator of interest. More importantly, MOCU provides a natural measure for the cost of uncertainty that quantifies the potential operator performance degradation because of uncertainty, directly focusing on operational objectives. Therefore, this IBR-MOCU framework not only provides the robust operator design and objective-oriented UQ but also leads to experimental design to choose an experiment to optimally reduce performance loss by adding to existing scientific knowledge. The IBR-MOCU paradigm follows in line from the early thinking of Wiener and Kolmogorov, and it extends and unifies previous work on robust filtering, classification, and control. The historical context of the IBR-MOCU framework is depicted in Figure 2.

In the following sections, we focus on recent developments on the corresponding components of this IBR-MOCU framework, including prior construction, model fusion, OED, and automated feature engineering for knowledge discovery in the context of materials science applications.

### Knowledge-driven prior construction

The first challenge of applying SP/ML methods in the MOCU framework to materials science is modeling and quantifying uncertainty because there rarely exist sufficient data for satisfactory system identification because of the enormous search space and the complicated CPSP relationships.<sup>4</sup> Small samples are commonplace in materials applications, in particular when the research focus is to discover novel complex functional mate-

rials. Thereafter, if prior knowledge, such as physics principles, may help constrain the SP/ML model space, it is critical to utilize these in systems modeling.<sup>71–73</sup> While Bayesian methods naturally model the uncertainty because of their distribution-based nature to treat model parameters as random variables, the salient obstacle confronting Bayesian methods is how to appropriately impose model prior.

Regarding prior construction, Jaynes<sup>74</sup> has remarked, “... there must exist a general formal theory of determination of priors by logical analysis of prior information—and that to develop it is today the top priority research problem of Bayesian theory.” However, the most common practice of Bayesian methods is to adopt either non-informative or conjugate prior for computational convenience. When there are limited data or strong scientific prior knowledge, it is precisely then that the formal structure as commented by Jaynes<sup>74</sup> is critical for appropriate prior construction.

In this section, we first briefly review traditional prior construction methods and then focus on the formal structure for prior construction involving a constrained optimization, in which the constraints incorporate existing scientific knowledge augmented by slackness variables. The constraints tighten the prior distribution in accordance with prior knowledge while at the same time avoiding inadvertent over-restriction of the prior, an important consideration with small samples.

### Traditional priors

Starting from Jeffreys’<sup>75</sup> non-informative prior, there was a series of information-theoretic and statistical methods: maximal data information priors (MDIP),<sup>76</sup> non-informative priors for integers,<sup>77</sup> entropic priors,<sup>78</sup> reference (non-informative) priors obtained through maximization of the missing information,<sup>79</sup> and least informative priors.<sup>80</sup> As discussed in the literature,<sup>81–83</sup> the principle of maximum entropy can be seen as a method of constructing least informative priors,<sup>84,85</sup> though it was first introduced in statistical mechanics for assigning probabilities. Except in the Jeffreys’<sup>75</sup> prior, almost all of the methods are based on optimization: maximizing or minimizing an objective function, usually an information theoretic one. The least informative prior<sup>80</sup> is found among a restricted set of distributions, whereas the feasible region is a set of convex combinations of certain types of distributions. Zellner<sup>86</sup> proposed several non-informative and informative priors for different problems. All of these methods emphasize the separation of prior knowledge and observed sample data.

*A priori* knowledge in the form of graphical models (e.g., Markov random fields) has also been widely utilized to either constrain the model space (for example, in covariance matrix estimation in Gaussian graphical models)<sup>87,88</sup> or impose regularization terms.<sup>89</sup> In these studies, using a given graphical model

illustrating the interactions between variables, different problems have been addressed; e.g., constraints on the matrix structure<sup>87,90,91</sup> or known independencies between variables.<sup>88,92</sup> Nonetheless, these studies rely on a fundamental assumption: the given prior knowledge is complete and hence provides one single solution. However, in many applications, the given prior knowledge is uncertain, incomplete, and may contain errors. Therefore, instead of interpreting the prior knowledge as a single solution (e.g., a single deterministic covariance matrix), we aim to construct a prior distribution on an uncertainty class.

### CPSP relationships in materials science

A central tenet in the field of materials science and engineering is that the processing history controls the material's internal structure, which, in turn, controls the effective (macroscale) properties or performance characteristics exhibited by the material. Exploration and exploitation of the materials space thus necessitate the generation of CPSP linkages.<sup>93,94</sup> Given the multiscale nature of the material's structures,<sup>93</sup> such (abstract) sets of CPSP linkages can be visualized as a large connected and nested network of models that mediate the flow of information about the material's state and behavior up and down the scales.

Any single model in this large network of models can be formally expressed as  $f(\mu, \varphi)$ , where  $\mu$  represents the appropriate CPSP variables (i.e., related to process history, material structure, or material property), and  $\varphi$  denotes variables describing the physics controlling the material phenomenon of interest. Established domain knowledge can be used to construct a prior on  $\varphi$ . Seeking  $f(\mu, \varphi)$  allows us to explicitly capture physics in formulating our ML/AI models. This allows us to use physics-based simulation data to train  $f(\mu, \varphi)$  by independently varying  $\mu$  and  $\varphi$ . Given the enormous challenges associated with the development of concurrent multiscale CPSP relationships, materials analysis tends to be carried out (most of the time) at different, not necessarily strongly coupled scales. At the mesoscale level and beyond (i.e., larger than the atomic scale), several efforts have been made to predict materials' behavior by using data-driven approaches. Most successful efforts at this scale have exploited low-dimensional representation of microstructure information to build effective property models.<sup>95</sup> To date, however, there is not much work on the direct use of physical principles to constrain the models used to establish these CPSP linkages. In this regard, more success has been achieved when considering the structure-property connections at the atomic scale.

From the atomic point of view, materials are fundamentally composed of atoms of similar or different types of chemical elements located on real-space sites. The equilibrium atomic structures of materials are reached through the minimization of total energy originated from the complex interaction among ions and electrons in the presence/absence of the external field. It consists of the Coulomb and kinetic energy of electrons and ions and the additional important contributions from quantum mechanical effects, such as (1) exchange energy because of the fermionic spin statistics of electrons, (2) static and dynamical correlation energy beyond the single Slater determinant approximated electronic wave functions, and (3) nuclear quantum effects when tunneling and delocalization of ions become important.<sup>96</sup> Recently, a graph convolutional neural network has been

applied to describe crystal and molecular structures of materials because atoms and bonds can be perfectly represented by graph nodes and edges, respectively. Recent examples include the crystal graph convolutional neural networks (CGCNN),<sup>97</sup> the improved CGCNN (iCGCNN),<sup>98</sup> the materials graph network (MEGNet),<sup>99</sup> etc. An underlying physical prior hypothesis is the locality of interactions; that is, the physical knowledge of interest can be learned from the local chemical interactions. For example, in the CGCNN,<sup>97</sup> the feature vector  $\mathbf{v}_i$  for atom  $i$  is updated via iterative convolution as

$$\mathbf{v}_i^{(t+1)} = \mathbf{v}_i^{(t)} + \sum_{j,k} \sigma(\mathbf{z}_{(ij)k}^{(t)} \mathbf{W}_f^{(t)} + \mathbf{b}_f^{(t)}) \odot g(\mathbf{z}_{(ij)k}^{(t)} \mathbf{W}_s^{(t)} + \mathbf{b}_s^{(t)}), \quad (\text{Equation 5})$$

where  $\mathbf{z}_{(ij)k}^{(t)} = \mathbf{v}_i^{(t)} \oplus \mathbf{v}_j^{(t)} \oplus \mathbf{u}_{(ij)k}$  is the concatenated neighbor vector consisting of atom  $i$ 's feature vector  $\mathbf{v}_i$ , feature vector  $\mathbf{v}_j$  of atom  $j$  located on the  $k$ -th bond of atom  $i$ , and the corresponding bond feature  $\mathbf{u}_{(ij)k}$ .  $\sigma$  is a sigmoid function, and  $g$  is a nonlinear softmax activation function.  $\mathbf{W}$  and  $\mathbf{b}$  denote the convolution weight matrix and bias of the corresponding layer, respectively. In these convolutional filters, the summation only runs through the local neighboring sites via local coordination determination<sup>97</sup> or Voronoi tessellation.<sup>98</sup> The results from these graph convolutional neural network approaches are promising because it is generally true that the physical interaction decreases as the distance of  $(i, j)$  atom pair (i.e., bond length), increases. This *a priori* physical knowledge is built inside these graph networks as an implicit constraint. While the bare Coulomb operator decays slowly with  $1/r$ , the destructive interference of electronic wave functions in many-particle systems leads to the nearsightedness of electronic matter in the absence of long-range ionic interactions;<sup>100,101</sup> i.e., local electronic properties, such as electron density, depend mostly on the effective external potential at nearby locations. However, for ionic systems, the long-range Coulomb interaction can have a non-negligible contribution to the total energy and atomic forces even when the  $(i, j)$  atom pair is separated far away, and further consideration to include these long-range interactions will be of great importance to more accurately describe the physical properties of ionic materials. In addition to these interaction-based physics principles, another important consideration when developing ML methods for materials systems is to make sure that the input feature and the derived descriptor representations should be invariant to the symmetries of the system, such as rotation, reflection, translation, and permutation of atoms of the same species. Kernel-based methods and topological invariants based on group theory have been recently investigated to help improve the accuracy of predictions in the ML modeling of solid state materials.<sup>102</sup>

### Maximal knowledge-driven prior (MKDIP) construction

Knowledge-driven prior construction utilizes first principles and expert domain knowledge to alleviate the model/data uncertainty and the small sample size issues through constraining the model space or deriving the uncertainty class of models based on physical and chemical constraints. Incorporating scientific knowledge to directly constrain Bayesian predictive models can achieve robust predictions, which would be impossible by using data alone. In materials science, there is a

substantial body of knowledge in the form of phenomenological models and physical theories for prior construction. Such knowledge can be used in choosing features or descriptors and constrain the model space for predicting novel materials with desired properties.

To translate more general materials knowledge into Bayesian learning, a general prior construction framework can be developed to map the known physical, chemical, and structural constraints into prior distributions in Bayesian learning. We have proposed such a framework, capable of transforming any source of prior information to prior probabilities given an uncertainty class of predictive models.<sup>103–106</sup> We call the final prior probability constructed via this framework an MKDIP. The new MKDIP construction constitutes two steps: (1) functional information quantification, where prior knowledge manifested as functional relationships is quantified as constraints to regularize the prior probabilities in an information theoretic way, and (2) objective-based prior selection, where, by combining sample data and prior knowledge, we build an objective function in which the expected mean log likelihood is regularized by the quantified information in step (1). As a special case, where we do not have any sample data, or where there is only one data point available for constructing the prior probability, the proposed framework is reduced to a regularized extension of the maximum entropy principle (MaxEnt).<sup>107</sup>

By introducing general constraints, which can appear as conditional statements based on expert domain knowledge or physics principles, the idea here is to maximally constrain the model uncertainty with respect to the prior knowledge characterized by these constraints. To give a simple example, assuming that we know *a priori*, based on physics principles, that certain microstructural properties  $R$  for a target material are determined by its composition  $X$ , we then can derive the corresponding constraint  $E_\pi[H_\theta(R|X)] \leq \xi$ , where  $H_\theta(R|X)$  denotes the conditional Shannon entropy of  $R$  given  $X$  under the probabilistic model determined by  $\theta$ . If our prior knowledge is correct, then  $H_\theta(R|X) \rightarrow 0$  for any appropriate model. Hence, under the uncertainty characterized by the prior distribution  $\pi(\theta)$ , we aim to derive the MKDIP with the expected conditional entropy as small as possible. Depending on different types of prior knowledge, we can write different forms of such constraints. Specifically, the MKDIP construction integrates materials science and statistical learning by (1) model prior knowledge quantification, where general materials knowledge, from physical theories or expert domain knowledge, is quantified via quantitative constraints or conditional probabilities and (2) optimization, where MKDIP construction requires solving the constrained optimization problems depending on different applications and data types of available observed measurements. When sufficient data exist, we can also split the data for prior construction and for updating the posterior, appropriately integrating prior knowledge and existing data.

In particular, MKDIP aims to derive the solution to the following optimization problem:

$$\operatorname{argmin}_{\pi \in \Pi} E_\pi[C_\theta(\xi, D)], \quad (\text{Equation 6})$$

where  $\Pi$  is the set of all proper priors, and  $C_\theta(\xi, D)$  is a cost function that depends on 1  $\theta$ , the random vector parameterizing the underlying probability distribution; (2)  $\xi$ , our state of (prior) knowl-

edge; and (3)  $D$ , partial observations. Alternatively, by parameterizing the prior probability as  $\pi(\theta; \gamma)$ , with  $\gamma \in \Gamma$  denoting the hyperparameters, the MKDIP can be found by solving

$$\operatorname{argmin}_{\gamma \in \Gamma} E_{\pi(\theta; \gamma)}[C_\theta(\xi, D)]. \quad (\text{Equation 7})$$

We have considered cost functions  $C_\theta$  that can be decomposed into three terms:<sup>106</sup>

$$C_\theta(\xi, D) = \lambda_1 [(1 - \beta)g_\theta^{(1)}(\gamma) + \beta g_\theta^{(2)}(D)] + \lambda_2 g_\theta^{(3)}(\xi),$$

where  $\beta$ ,  $\lambda_1$ , and  $\lambda_2$  are non-negative regularization parameters. Here,  $g^{(1)}(\cdot)$  denotes the information-theoretic cost, which can take different forms, including MaxEnt;<sup>107</sup>  $g^{(2)}(\cdot)$  is the cost that involves the partially observed data when they are available, including regularized MDIP and regularized expected mean log likelihood prior;<sup>103</sup> and, more critically,  $g^{(3)}(\cdot)$  denotes the knowledge-driven constraints that convert prior knowledge into functional constraints to further regularize the prior as detailed in Boluki et al.<sup>106</sup> Using this cost function, we formulate the MKDIP construction problem as the following optimization problem:

$$\operatorname{argmin}_{\gamma \in \Gamma} E_\theta [(1 - \beta)g_\theta^{(1)}(\gamma) + \beta g_\theta^{(2)}(D)] \quad (\text{Equation 8})$$

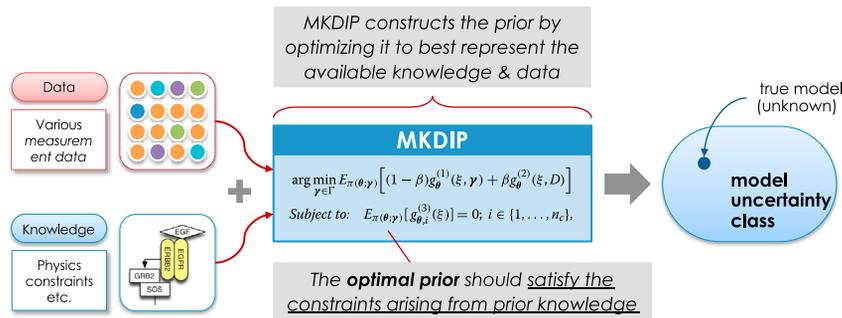
Subject to :  $E_\theta [g_{\theta, i}^{(3)}] > 0; i \in \{1, \dots, n_c\}$ ,

where  $g_{\theta, i}^{(3)}$ ,  $\forall i \in \{1, \dots, n_c\}$ , are constraints resulting from our state of knowledge  $\xi$  via the mapping  $\mathcal{T}: \xi \rightarrow E_\theta [g_{\theta, i}^{(3)}] > 0$ ,  $\forall i \in \{1, \dots, n_c\}$ ; for example, based on the aforementioned composition-structure relationship  $E_\pi[H_\theta(R|X)]$ . The overall MKDIP scheme is illustrated in Figure 3.

In contrast to non-informative priors, MKDIP aims to incorporate the available prior knowledge and uses *part* of the data to construct an informative prior. While, in theory, the observed data can be entirely used in the optimization problem in Equation 8, in practice one should be cautious to avoid overfitting to the given data. The MKDIP construction here introduces a formal procedure for incorporating prior knowledge. It allows the incorporation of the knowledge of functional relationships and any constraints on the conditional probabilities. Finally, we shall note that deriving the solution to the MKDIP optimization problem Equation 8 can be challenging because of the non-convexity of the objective function and constraints. Nevertheless, feasible and local optimal solutions, especially with the specific distribution families and constraint forms, can be derived.<sup>103</sup>

### Integrating prior knowledge in materials science

Xue et al.<sup>52</sup> have applied Bayesian learning and experimental design based on materials knowledge using results from the Landau-Devonshire theory for piezoelectric materials. In particular, a Bayesian regression model,<sup>54</sup> constrained by the Landau functional form and the constraints on morphotropic phase boundaries (MPBs), was developed to guide the design of novel materials with the functional response of interest and to help navigate the search space efficiently so that the desired composition can be achieved in a few trials. The Landau-Devonshire



**Figure 3. Illustration of knowledge-based prior construction via MKDIP**

theory has been widely used to reproduce phase diagrams for many piezoelectrics and to investigate their performance at the MPB. The ferroelectric nanodomain phases can be characterized by different polarization vectors,  $\vec{p} = \vec{n}\rho$ , where  $\vec{n} = [n_1, n_2, n_3]^T$  is a unit vector in the direction of polarization, and  $\rho$  is its magnitude.<sup>108</sup> The free energy,  $g$ , of the ferroelectric system (e.g.,  $BaTiO_3$ -based piezoelectrics) can be described by a Landau polynomial that depends on the modulus of the polarization vector ( $\rho$ ) and the polarization direction ( $\vec{n}$ ) at a given temperature  $\tau$ :

$$g(\rho, \vec{n}; \tau) = \frac{\alpha}{2} (n_1^2 + n_2^2 + n_3^2) \rho^2 + \frac{\beta_1}{4} (n_1^2 + n_2^2 + n_3^2)^2 \rho^4 + \frac{\beta_2}{4} (n_1^4 + n_2^4 + n_3^4) \rho^4 + \frac{\gamma_1}{6} (n_1^2 + n_2^2 + n_3^2)^3 \rho^6 + \frac{\gamma_2}{6} (n_1^6 + n_2^6 + n_3^6) \rho^6 + \frac{\gamma_3}{6} (n_1^2 n_2^2 n_3^2) \rho^6,$$

where the coefficients  $\alpha$ ,  $\beta$ 's, and  $\gamma$ 's are materials dependent and often determined from experiments; for example,  $\beta_2(\tau, X)$  depends on the temperature ( $\tau$ ) and composition ( $X$ ). The MPB is a phase boundary where the two phases (i.e., tetragonal [T] and rhombohedral [R] phases in  $BaTiO_3$ -based piezoelectrics) coexist and have degenerate free energy. Therefore, at MPB ( $\tau = \tau_{MPB}$  and  $X = X_{MPB}$ ),  $g_T = g_R$ , which leads to

$$\beta_2(\tau_{MPB}, X_{MPB}) + \frac{24\gamma_2 - \gamma_3}{27} \rho_{eq}^2(\tau_{MPB}, X_{MPB}) = 0,$$

$\rho_{eq}^2(\tau, X)$  denotes the polarization at equilibrium and has the functional form  $\rho(\tau - \tau_C(X))$ , where  $\rho$  is a constant, and  $\tau_C(X)$  is the composition-dependent Curie temperature. Based on these relationships (more details can be found in Xue et al.<sup>52</sup>), the MPB curve has the following quadratic form:

$$\tau_{MPB}(X) = \omega_1 X^2 + \omega_2 X + C_1,$$

where  $\omega_1$ ,  $\omega_2$ , and  $C_1$  are the corresponding model parameters to learn from experimental data. This serves as the prior knowledge to constrain our Bayesian regression model to map the material composition  $x$  to the MPB curves.

As illustrated in Figure 4, with the minimal collected data (only 20 characterized  $BaTiO_3$ -based piezoelectrics), the Bayesian regression model with the aforementioned functional constraints provides reliable phase boundaries and faithful uncertainty estimates. More importantly, we demonstrated our approach for finding  $BaTiO_3$ -based piezoelectrics with the desired target of a vertical

MPB. We have predicted, synthesized, and characterized a solid solution,  $(Ba_{0.5}Ca_{0.5})TiO_3$ - $Ba(Ti_{0.7}Zr_{0.3})O_3$ , with piezoelectric properties showing better temperature reliability than other  $BaTiO_3$ -based piezoelectrics in our initial training data.

When the prior knowledge, including different functional forms and constraints, is available, the MKDIP framework can

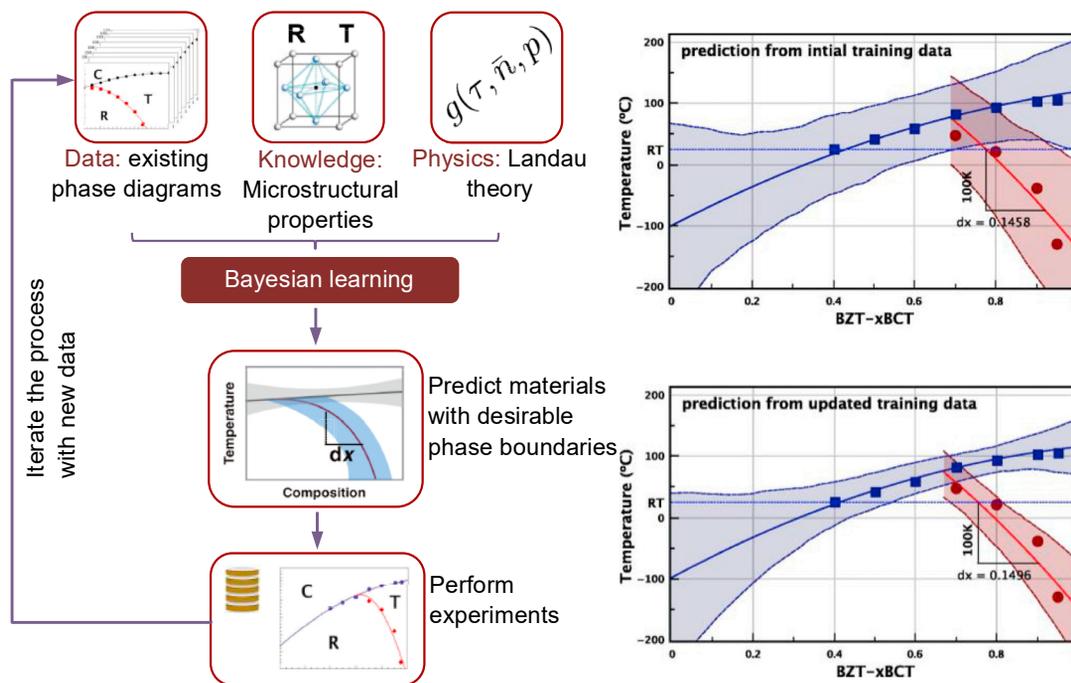
help take the best advantage of them to explicitly determine the predictive models as well as their corresponding predictors for specific functional responses of interest. Besides such explicit functional-form prior knowledge, which allows us to directly constrain predictive models, the existing prior knowledge on CPSP relationships may simply be in the form of correlation, conditional relationships, and inequality constraints. To enable users, especially materials domain experts, to easily explore and integrate existing phenomenological knowledge into Bayesian learning, infrastructure and friendly user interfaces should be developed to help prior construction via active knowledge acquisition from either materials scientists or even more recent large language foundation models as the unprecedented knowledge base.<sup>9,13</sup> The current practice is mostly hand crafted based on different problems and how data scientists work with their collaborating materials scientists. More interfacing efforts between data scientists and materials domain experts are required to achieve more synergistic collaboration in materials science.

### Bayesian model averaging (BMA) with experimental design

With a derived surrogate model, we would like to exploit it in combination with experiments to accelerate the development of new materials. However, often, because of incomplete prior knowledge, there are multiple feasible surrogate models within the uncertainty class. We further explore a Bayesian experimental design framework that is capable of adaptively selecting or aggregating competing models connecting materials composition and processing features to performance metrics through BMA.<sup>109,110</sup>

### Review on Bayesian model fusion

Bayesian model fusion methods have been studied extensively to achieve better predictive accuracy as well as robust risk and uncertainty estimates.<sup>70,109,111,112</sup> There are different Bayesian model ensemble strategies stemming from the Bayes' theorem from Bayesian inference,<sup>70</sup> including Bayesian model selection, Bayesian model combination, and BMA. They all start with an ensemble of candidate models as the uncertainty model class and then update the model posterior probabilities given observed data. The main difference among these different strategies lies in how the updated posterior probabilities guide the way to derive posterior predictive probabilities. For example, Bayesian model selection aims to identify the best predicting model(s) with different criteria, including the Bayesian information criterion (BIC) and Akaike information criterion (AIC).<sup>113,114</sup> Bayesian model combination often samples the best model



**Figure 4. Bayesian learning and experimental design constrained by the Landau functional for discovery of BaTiO<sub>3</sub>-based piezoelectrics as described in the text**

Shown are predicted (solid lines) and experimental (dots) phase diagrams for BZT-m50-n30, together with uncertainty estimates, from Bayesian regression. The solid lines show the mean phase boundaries, and the dashed lines mark the 95% confidence intervals. Notice the uncertainty reduction given more data.

subsets based on the updated model posterior, hoping to achieve better convergence.<sup>115</sup> In this paper, we focus on BMA, which essentially relies on the weighted ensemble of the models in the uncertainty class by the model posterior.<sup>109,112</sup> The theoretical properties of BMA have been studied in the literature. For example, BMA can achieve better prediction performance than any model in the uncertainty class.<sup>109,116</sup> The corresponding implementations addressing model uncertainty have also been investigated for more effective and efficient inference procedures.<sup>112</sup>

#### **BMA with MOCU for OED**

For Bayesian experimental design in general, there can be three categories of objective functions to guide the experimental design. In the first case, we have a parametric model where the parameters come from an underlying physical system. One such example is in biomedicine, where the objective function is the likelihood of the cell being in a cancerous state, given a state-space model based on genetic regulatory pathways.<sup>117</sup> Another example is in imaging (for example, for image reconstruction or filtering), where the parameters characterize the image appearance, and the objective function is an error measure between two images.

In the second category, the features are given, and the parameters come from a surrogate model used in place of the actual physical model but are believed to be appropriately related to the physical model. For example, in the materials science applications under “OED with MOCU,” the surrogate model is based on the time-dependent Ginzburg-Landau (TDGL) theory and simulates the free energy given dopant parameters, the objective function is the energy dissipation, and the action is to find an

optimal dopant and concentration.<sup>5</sup> To see how the approach in Dehghannasiri et al.<sup>5</sup> fits the above general theory, the reader can refer to Boluki et al.<sup>118</sup>

In the third category, we do not know the physical model, and we lack sufficient knowledge to posit a surrogate model with known features/forms relating to our objective. This case arises in many scenarios where the objective function is a “black box” function. Nevertheless, we can adopt a model, albeit one with known predictive properties. This model can be a kernel-based model, such as a GP.<sup>119</sup> Moreover, this model can consist of a set of possible parametric families, a kernel-based model with different possible feature sets, or even kernel-based models with different choices for the kernel function. In such scenarios, we do not *a priori* have any knowledge about which feature set or model family would be the best, and reliable model selection cannot be performed before starting the experiment design loop because of the limited number of observed samples. Considering the average prediction from models based on different feature sets or model families weighted by their posterior probability of being the correct model, namely BMA, is one possible approach.

In the context of materials discovery, we can frame the model averaging problem in a hierarchy to define a family of uncertain model classes in which, for example, different features contribute differently to functional property prediction differently. With such a hierarchical Bayesian model, BMA, essentially weighing all the possible models by their corresponding probability of being the true model, is embedded in BO for OED to realize a system not only capable of autonomously and adaptively learning the surrogate predictive models for the most

promising materials of desired properties but also utilizing the models to efficiently guide exploration of the design space. With more acquired data, the uncertainty of different models will be quantified, and improved predictive models as well as efficient experimental design can be attained.

Again, assume an uncertainty class  $\Theta$  with the probability measure  $\Pi$ , characterizing predictive models on a design space  $\mathcal{X}$ . The experimental design goal is to optimize an objective function  $f: \Theta \times \mathcal{X} \rightarrow \mathbb{R}$ . For example, we want to find a design  $\mathbf{x} \in \mathcal{X}$  that minimizes an unknown true objective function  $f(\mathbf{x}; \theta_t)$  over  $\mathcal{X}$ , where  $\theta_t \in \Theta$  denotes the true model. When there is no strong prior knowledge on functional forms of the objective function, often GP regression (GPR) is adopted and iteratively updated given data from performed experiments  $D_n$ :  $P(f|\mathbf{x}; D_n) \sim GP(y|\mu, K)$ , where  $\{\mu, K\}$  denote the corresponding mean and kernel parameters. To account for potential model uncertainty, BMA can be used for more robust modeling of the objective function:

$$P(f|\mathbf{x}; D_n) = \sum_{i=1}^L P(i|D_n)P(f|\mathbf{x}; D_n, \mu_i, K_i), \quad (\text{Equation 9})$$

where  $i$  is the index of the candidate models in the uncertainty class.

As explained under “Bayesian learning, UQ, and experimental design,” a robust design is an element  $\mathbf{x}^R \in \mathcal{X}$  that minimizes the average of the objective function across all possibilities in the uncertainty class relative to a probability distribution governing the corresponding space. This probability at each experimental design iteration is the posterior distribution given the observed data points available up to that step. Mathematically,

$$\mathbf{x}_n^R = \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} E_{\theta} [f(\mathbf{x}; \theta) | D_n], \quad (\text{Equation 10})$$

where  $D_n$  denotes the observed data till the  $n$ th iteration. MOCU in this context can be defined as the average gain in the attained objective between the robust design and the actual optimal designs across the possibilities:

$$\text{MOCU}_n^{\mathcal{X}}(\Theta) = E_{\theta} [f(\mathbf{x}_n^R; \theta) - f(\mathbf{x}_{\theta}^*; \theta) | D_n], \quad (\text{Equation 11})$$

where  $\mathbf{x}_{\theta}^*$  denotes the optimal action for a given model parameterized by  $\theta$ , including both GPR parameters and additional parameters from BMA. Note that, if we actually knew the true (correct) model, then we would simply take the optimal design for that model, and MOCU would be 0. Denoting the set of possible experiments by  $\Xi$ , the best experiment  $\xi_n^*$  at each time step (in one-step look-ahead scenario) is the one that maximally reduces the expected MOCU following the experiment; i.e.,

$$\xi_n^* = \underset{\xi \in \Xi}{\operatorname{argmin}} E_{\xi} [E_{\theta} [f(\mathbf{x}_{n+1}^R; \theta) | \xi, D_n]] - E_{\theta} [f(\mathbf{x}_n^R; \theta) | D_n]. \quad (\text{Equation 12})$$

In most cases in materials discovery, each experiment is synthesizing the corresponding materials design and measuring its actual properties (or their noisy versions). Thus, the experiment space is equivalent to the design space.

It is beneficial to recognize that MOCU can be viewed as the minimum expected value of a Bayesian loss function, where

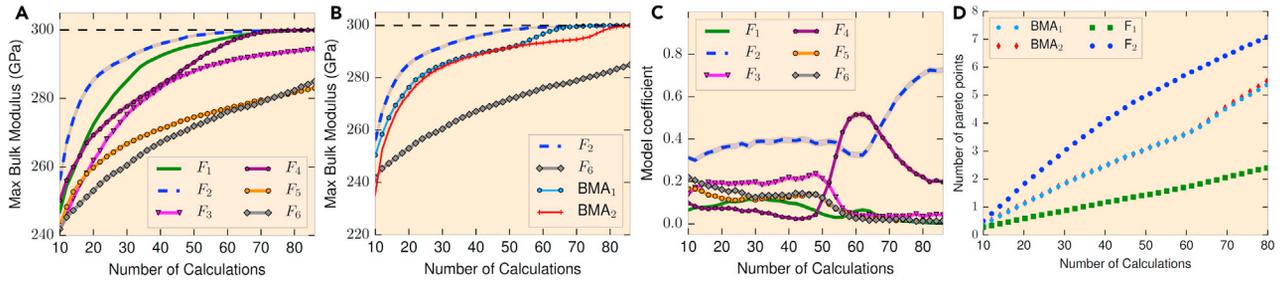
the Bayesian loss function maps an operator (the materials design in this context) to its differential objective value (for using the given operator instead of an optimal operator), and its minimum expectation is attained by an optimal robust operator that minimizes the average differential objective value. In decision theory, this differential objective value has been referred to as the regret. Under certain conditions, MOCU-based experimental design is, in fact, equivalent to KG and EGO.<sup>118</sup>

### BMA for materials science applications

We have integrated BMA with the MOCU-based experimental design to deploy an autonomous computational materials discovery framework that is capable of performing optimal sequential computational experiments to find optimal materials and updating the knowledge on materials system model at the same time. One of our recent exercises<sup>120</sup> consisted of implementing the BMA approach for robust selection of computational experiments to optimize properties of the MAX phase crystal system.<sup>121</sup> Employing BMA approaches using a set of GPR functions based on different feature sets, we demonstrated that the framework was robust against selection of poor feature sets because the approach considers all the feature sets at once, updating their relative statistical weights according to their ability to predict (successful) outcomes of unrealized simulations. More critically, we have demonstrated the effectiveness of our computational materials discovery platform for single and multiobjective optimization problems.

This framework has been used efficiently for objective-oriented exploration of materials design spaces (MDSs) through computational models and, more importantly to guide experiments by focusing on gathering data in sections of the MDS that will result in the most efficient path to achieving the optimal material within resource budgets. Additionally, the BO approach was successfully combined with BMA for autonomous and adaptive learning, which may be used to auto-select the best models in the MDS, thereby eliminating the requirement of knowing the best model *a priori*. Thus, this framework constitutes a paradigm shift in the approach to materials discovery by simultaneously (1) accounting for the need to adaptively build increasingly effective models for the accelerated discovery of materials while (2) accounting for the uncertainty in the models themselves. It enables a long-desired seamless connection between computation and experiments, each informing the other, while progressing optimally toward the target material.

In our implementation for MAX phase crystal systems, after training the GPs based on the current and previous observations, solving for the GP hyperparameters to maximize the marginal likelihood of the observed data, each GP provides a Gaussian distribution over the objective function value of each design. Averaging several GPs based on their posterior model probabilities is like mixing weighted Gaussian distributions over the objective value of each design. Based on the sum of weighted Gaussian distributions, the MOCU-based utility function or other acquisition functions, including expected improvement (EI) with a single objective<sup>45</sup> or expected hypervolume improvement (EHVI) with multiobjectives,<sup>122</sup> can be calculated for all possible designs, and the maximizer is chosen as the next experiment. In our experiments, six sets of basic



**Figure 5. Bayesian experimental design with BMA for MAX phases as described in the text**

- (A) The change of average maximum bulk modulus for the original six feature sets with the number of design iterations.  
 (B) The change of average maximum bulk modulus comparing BMA surrogates with the best and worst feature sets.  
 (C) The change of posterior model probabilities corresponding to six feature sets.  
 (D) The average number of sampled Pareto front points when considering bulk modulus and shear modulus.

compositional and structural features were chosen *a priori* without assuming any knowledge of their suitability for the underlying true model that generates data. We have investigated whether the updated model posterior in BMA captured the expected CPSP relationships.<sup>120</sup> The goal of experimental design is to discover MAX phases with maximum bulk modulus and minimum shear modulus, which were computed through density functional theory (DFT) calculations<sup>123,124</sup> for 1,500 randomly sampled MAX ternary carbide/nitride crystals. Among these DFT-calculated results, there were 10 MAX phases belonging to the Pareto front when considering the design goals. All of the reported performances of Bayesian experimental design were based on the average values of 1,500 runs starting from the random initial sets of 10 training samples. In Figure 5A, we show the change of the average maximum bulk modulus with the iterations of sequential experimental design. It is clear that, among six models with different features, the feature set  $F_2$  achieves the best experimental design performance because the average maximum bulk modulus is consistently higher than the other models. On the other hand,  $F_6$  has the worst performance. When adopting BMA (either based on first-order or second-order maximum likelihood inference [BMA1 or BMA2, respectively]), it is clear that BMA achieves robust performance even when some models may not have good predictive power (Figure 5B). With the increasing number of iterations, it is also clear that the posterior probability of the best model,  $F_2$ , gets higher (Figure 5C). Last but not least, as shown in Figure 5D, our BMA-based multiobjective experimental design can approach the Pareto front within a small number of sequential design iterations considering the vast MAX ternary carbide/nitride space. All of these experimental results on the maximization/minimization of mechanical properties of MAX phases suggest that BMA-based model fusion can lead to considerable reduction in the number of experiments/computations that need to be carried out to identify the desired solutions to this specific materials design problem.

Along these directions, we can develop robust Bayesian learning methods by model fusion that exploit correlations among sources/models. Together with a multiinformation source optimization framework driven by scientific knowledge, they will reliably and efficiently identify, given the current knowledge, the next best information source to query and guide the materials design.<sup>125</sup>

### OED with MOCU

In the context of OED, it has a long history in science and engineering as a properly designed experimental procedure that provides much greater efficiency than simply making random probes. Indeed, Francis Bacon's call for experimental design in 1620 is often taken to be the beginning of modern science.<sup>126</sup>

### MOCU-based OED

Because the MOCU<sup>65,66</sup> can be used to quantify the objective-based uncertainty, it provides an effective means to estimate the expected impact of potential experiments on the objective (i.e., operational goal) through the reduction of model uncertainty. Suppose we are given a set of potential experiments from which the next experiment could be chosen. Which among the possible experiments should be selected if we wish to optimally improve the operational performance of the operator based on the expected experimental outcome? A natural way to select the best possible experiment would be to choose the one that would lead to the minimum expected remaining MOCU after observing its outcome. To be more specific, let  $\xi \in \Xi$  be an experiment in the experimental design space  $\Xi$ . Given  $\xi$ , the MOCU conditioned on this experiment can be computed as

$$\text{MOCU}(\Theta|\xi) = E_{\theta|\xi} \left[ C_{\theta} \left( \psi_{\text{IBR}}^{\Theta|\xi} \right) - C_{\theta}(\psi_{\theta}) \right], \quad (\text{Equation 13})$$

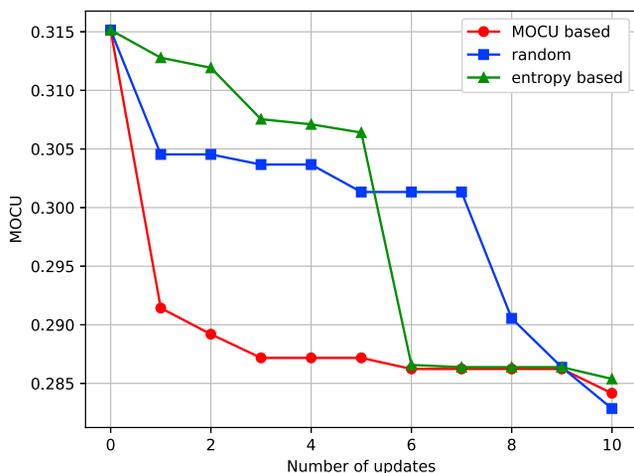
where  $\psi_{\text{IBR}}^{\Theta|\xi}$  is the IBR operator that is optimally robust for the uncertainty class of models  $\Theta|\xi$  that is now conditioned on this experiment  $\xi$ , and the expectation is taken with respect to the conditional distribution  $\pi(\theta|\xi)$ . The expected remaining MOCU can be evaluated by

$$R(\Theta|\xi) = E_{\xi} [\text{MOCU}(\Theta|\xi)], \quad (\text{Equation 14})$$

and the optimal experiment  $\xi^*$  is the one that minimizes the expected remaining MOCU in Equation 14 so that it satisfies

$$\xi^* = \underset{\xi \in \Xi}{\text{argmin}} R(\Theta|\xi). \quad (\text{Equation 15})$$

While this strategy does not guarantee that the selected experiment will indeed minimize the uncertainty impacting the objective among all experiments (because the experimental outcome is not known in advance with certainty), it will be



**Figure 6. Experimental design results based on a 5-oscillator Kuramoto model with uncertain coupling strength between the oscillators as described in the text**  
The MOCU-based OED scheme quickly reduces the model uncertainty that impacts the performance.

optimal on average. Recently, this MOCU-based experimental design scheme has been developed for a variety of systems and applications, which include enhancing the performance of gene-regulatory network intervention with partial network knowledge,<sup>117,127</sup> synchronization of an uncertain Kuramoto model that consists of interconnected oscillators with uncertain interaction strength,<sup>128,129</sup> optimal sequential sampling,<sup>130</sup> Bayesian classification through active learning,<sup>44,131</sup> and robust filtering of uncertain stochastic differential equation (SDE) systems.<sup>42</sup>

For materials discovery via OED guided by MOCU, as shown in Equation 15, optimization algorithms have to be developed based on the structure of the input design space as well as the properties of the MOCU computation based on different problem settings. For example, if we are investigating pool-based high-throughput screening or discovery problems with a finite set of candidates, either exhaustive search as in typical BO implementations<sup>5,53,131</sup> or dynamic programming algorithms based on KGs<sup>47,132–134</sup> can be developed for solving the optimization problems. When the input design space is continuous and the gradient of MOCU can be estimated, gradient-based local search algorithms can be implemented, as discussed in Zhao et al.<sup>135</sup> There are also other solution strategies that can be used to solve OED guided by MOCU, including sampling and genetic and other evolutionary algorithms.<sup>136</sup>

Figure 6 shows the performance of the MOCU-based OED strategy in reducing the uncertainty that impacts the synchronization cost of a Kuramoto model that consists of 5 oscillators, where the coupling strength between oscillators is uncertain and known only up to a range.<sup>128</sup> In this example, an experiment picks an oscillator pair and observes whether the selected oscillator pair is synchronized in the absence of external control. The observation can be used to reduce the range of the uncertain coupling strength between the oscillators. For this Kuramoto model, there exist  $\binom{5}{2}$  potential experiments in the experimental design space  $\Xi$ , and Figure 6 shows how MOCU de-

creases as a function of experimental updates. As can be seen, the MOCU-based OED strategy leads to a sharp reduction in uncertainty within a few updates, outperforming random selection (which selects one of the possible experiments from  $\Xi$  with uniform probability) or an entropy-based approach (which selects the experiment for the oscillator pair whose coupling strength has the largest uncertain range).

### OED for shape memory alloy (SMA)

In materials design, the MOCU-based OED strategy has been applied to a computational problem for shape memory alloy (SMA) design with desired stress-strain profiles for a particular dopant at a given concentration utilizing the TDGL theory.<sup>5</sup> The TDGL model simulates the free energy for a specific dopant with a specified concentration, given the dopant's parameters, which is considered an oracle in the experiments. Because the computational complexity of the TDGL model is enormous, an uncertain surrogate model is first trained to approximately predict a dissipation energy for a specified dopant and concentration. In particular, based on TDGL, a reciprocal function is adopted to model the energy dissipation at a specific temperature as a function of dopant potency, dopant spread, and dopant concentration. The experimental design goal is to discover SMAs with the minimum energy dissipation, and therefore this surrogate model is used as the cost function to define MOCU to efficiently guide throughout the experimental design iterations for an optimal dopant and concentration. With the MOCU defined based on this Landau mesoscale surrogate for SMAs as the cost function, the expected remaining MOCU, given the corresponding dopant and its corresponding concentration levels, can be computed by the definition in Equation 14. The optimal experiment can then be determined to minimize the expected remaining MOCU under model uncertainty as in Equation 15.

In the reported experiments,<sup>5</sup> MOCU-based OED was compared with the pure exploitation and random selection policies. Averaged over 10,000 simulations, our MOCU-based OED strategy, which strives to minimize the uncertainty in the model pertaining to the design objective, identified the dopant and concentration with the optimal dissipation after only two iterations on average, while either exploitation or random selection policies cannot find the optimal dopant even after 10 iterations. Getting optimal results after fewer iterations is especially crucial in materials discovery, where measurements by either high-throughput simulation models or synthesis and profiling experiments are expensive and time consuming.

### Automatic feature engineering (AFE)

Finally, with accumulated knowledge and data from experimental design based on objective-UQ using MOCU, we may help fill in the missing gap of the understanding in materials systems under study. In materials science, the fundamental paradigm is the existence of causal relationships connecting composition and processing (i.e., the modifications to a material's current state), structure (i.e., the multiscale arrangement of the material), and properties (i.e., the response of the material to an external stimulus); i.e., CPSP relationships. The navigation of this CPSP space is enormously resource intensive, regardless of whether this query is on physical experiments or computational ones. As a result, it typically takes more than 20 years to

identify, develop, and finally deploy one material in real-world applications—a key bottleneck for the MGI.<sup>1,3,4</sup> Attempting to use physics-agnostic models to build these relationships is limited by the scarcity of the training data itself. Moreover, one would be interested in discovering derived relationships that connect features to properties/behavior because these relationships can further be used to design/discover materials with optimal properties. Besides designing and discovering promising new materials with desired functional properties, identifying critical input features (related to composition, process, structure) that determine function properties as well as principled CPSP relationships can provide a systematic understanding of the underlying physics for different materials systems. Such knowledge can be explored and updated, as illustrated in the previous examples under “Integrating prior knowledge in materials science” and “BMA for materials science applications.” One such knowledge discovery strategy is AFE, which enables us to use physics constraints on learning surrogate models while facilitating the discovery of fundamental materials design rules at the same time.

Engineered features obeying physics principles provide valuable interpretability that is critical to help new knowledge discovery and consequent critical decision-making. It is worth noting that, in scientific ML (sciML) involving complex systems, training data tend to be scarce and noisy because obtaining data can be difficult, time consuming, and costly. Materials problems clearly reflect these challenges.

### Related work in feature engineering

Feature representation learning has been studied extensively in the SP/ML community, including “white box” methods based on specific basis families (Fourier and wavelet are two representatives) and data-driven “black box” methods, such as dictionary learning and deep learning.<sup>137–139</sup> Although “black box” deep AFE models<sup>140</sup> have shown great potential to improve the corresponding ML algorithm performance, we focus on feature engineering, aiming to derive features based on explicit functional forms in this survey. Desirable feature engineering should attain considerable improvement of prediction performance and generalizability as well as good interpretability with little manual labor. Among the existing methods, deep feature synthesis<sup>141</sup> extracts features based on explicit functional relationships without experts’ domain knowledge through stacking multiple primary features and implementing operations or transformations on them, but it suffers from efficiency and scalability problems because of its brute-force way to generate and select features. Kaul et al.<sup>142</sup> proposed *Autolearn* by regression-based feature learning through mining pairwise feature associations. While it avoids overfitting, to which deep learning-based FE methods are amenable, and improves the efficiency by selecting subsets of engineered features according to stability and information gain, it does not directly produce interpretable features. Khurana et al.<sup>143</sup> introduced *Cognito*, which formulates the feature engineering problem as a search on the transformation tree with an incremental search strategy to explore the prominent features and later extended the framework by combining reinforcement learning (RL) with a linear functional approximation<sup>144</sup> to improve the efficiency. A similar framework has recently been developed in Zhang et al.,<sup>145</sup> where the deep reinforcement learning (DRL)

policy is learned on a tree-like transformation graph. It improves the policy learning capability compared with *Cognito*. However, both frameworks do not explicitly incorporate available prior knowledge into the AFE procedures.

For AFE in materials science applications, we are interested in finding the actuating mechanisms of the materials’ functional properties of interest by identifying a set of physically meaningful variables and their relationships.<sup>146</sup> Such a set of physical variables with corresponding parameters that uniquely describe the materials’ properties of interest can be denoted as “descriptors.” Discovering descriptors in materials science can help better predict target functional properties with potential interpretability for a given complete class of materials.<sup>147</sup> Several methods have been developed, such as a method based on compressed sensing<sup>147</sup> and the more recent Sure Independent Screening and Sparse Operation (SISSO)<sup>148</sup> by brute-force search to generate and select subsets of generated features by sure independent screening<sup>149</sup> together with sparse operators such as least absolute shrinkage and selection operator (LASSO).<sup>150</sup> These methods pose a scalability challenge with the exponentially growing memory requirement to store intermediate features and high computational complexity to search for features.

### Physics-constrained AFE

In our recently developed AFE framework,<sup>151</sup> a feature generation tree (FGT) was constructed with physics constraints to explore the engineered feature (descriptor) space more efficiently based on first principles, which was demonstrated in several materials problems to be able to take advantage of prior chemical and physical knowledge of the materials systems under study.

Our FGT-based AFE framework focuses on sciML applications, where interpretability is critical to help consequent critical decision-making under data scarcity and uncertainty. Specifically, AFE strategies have been developed by combining FGT exploration with Deep Reinforcement Learning (DRL)<sup>152</sup> to address the interpretability and scalability challenges. Instead of employing a brute-force way to perform algebraic operations on the raw features in a given dataset and then selecting important descriptors, we combine the descriptor generating and selecting processes together by constructing FGTs and developing the corresponding tree exploration policies guided by a deep Q network (DQN). An efficient exploration of the prominent descriptors can be attained in the growing feature space based on the allowed algebraic operations. Our FGT-based AFE strategies construct interpretable descriptors based on a list of operations according to the DRL learned policies, which are more scalable and flexible with the performance-complexity trade-off with the help of adjustable batch size for generating intermediate features. More critical to materials science and other sciML problems, our FGT provides a flexible framework for incorporating prior knowledge (e.g., physics constraints) to generate and select features. This is important for knowledge discovery via interpretable learning with physics constraints under data scarcity and uncertainty because the space connecting intrinsic materials attributes/features to materials behavior is vast, sparse, and complex in nature.

In particular, let  $\mathbf{x}_0$  denote the finite set of  $p$  variables as raw or primary features  $\{\mathbf{x}_0^1, \dots, \mathbf{x}_0^p\}$  and  $y$  the target output vector. AFE

**Algorithm 1. DQN for AFE**

```

1: input: Primary features  $F_0$ , Action set  $O$ 
2: for  $d = 1, 2, \dots$  do
3:   Construct new DQN
4:   Clear Buffer
5:   for  $episode = 1, 2, \dots, N$  do
6:     for  $i = 0, 1, \dots$ , do
7:        $\varphi_i = \epsilon$ -Greedy Method( $F_i, \epsilon$ )
8:        $F_{i+1}^d, R_i, c_{i+1} = \text{FGT\_Grow} \{F_i^d, \varphi_i, c_i\}$ 
9:       Buffer  $\leftarrow \{F_i^d, F_{i+1}^d, \varphi_i, R_i, c_{i+1}\}$ 
10:      Train DQN with experience replay
11:      if  $R_i > \text{threshold}$  then
12:        goto Output
13:      end if
14:      if  $c_i \geq c_{max}$  then
15:        break
16:      end if
17:    end for
18:  end for
19:   $\mathbb{S} \leftarrow$  Candidate set  $\mathbb{S}^d$  with  $n$  features of highest  $R_i$ 
20: end for
21: Output: Optimal feature set  $F^*$  chosen from  $\mathbb{S}$ 

```

is to develop an algorithm to construct sets of engineered features as interpretable and predictive descriptors  $F_i = \{g_1(\mathbf{x}_0, c_1), g_2(\mathbf{x}_0, c_2), \dots\}$  based on explicit functional forms with allowed algebraic operations that accurately predict  $y$ . The set of algebraic operations  $\varphi$  in an operation set  $O$  can be constructed based on prior knowledge; for example, with the following unary and binary operations:  $O = \{\exp(\cdot), \log(\cdot), (\cdot)^2, (\cdot)^3, (\cdot)^{-1}, \sqrt{\cdot}, \sqrt[3]{\cdot}, +, -, \times, \div\}$ . For each function  $g(\mathbf{x}_0, c)$ ,  $c$  denotes the complexity of the corresponding generated descriptor—the number of algebraic operations. For example, the function  $\exp(\mathbf{x}_0^0) \times (\mathbf{x}_0^1)^2 + \sqrt{(\mathbf{x}_0^2)}$  has a complexity of 5. The operation set  $O$  can be pre-defined based on the prior knowledge about the system under study. If we denote the primary features  $\mathbf{x}_0$  by  $F_0$ , then  $F_i$  denotes the iteratively generated set of descriptors with the maximum allowed complexity  $c_i$ . Our goal is to find an optimal descriptor set  $F^*$  that maximizes the prediction performance score; for example, by classification or regression accuracy,  $A_L\{F', y\}$ :

$$F^* = \underset{\forall f^k \in F, f^k \in \mathbb{F}}{\operatorname{argmax}} A_L\{F, y\}, \quad (\text{Equation 16})$$

where  $L$  denotes the prediction model (for example, linear regression or Support Vector Machine (SVM) for interpretability with generated descriptors), and  $f^k$  is any descriptor (including primary features) in  $F$ , the set of all generated features with the maximum allowed complexity  $c_{max}$ .

The combinatorial optimization problem in Equation 16 is NP hard. We solve it approximately by introducing the FGT to iteratively construct the descriptor space and transform the problem into a tree search problem for efficient AFE. Each node in the FGT represents a set of descriptors  $F_i$ , and each edge represents an operation  $\varphi$ . We denote  $(F^d)^* = \{(f^1)^*, (f^2)^*, \dots, (f^d)^*\}$  as the top  $d$  optimal features when we choose the cardinality of  $F^*$  as

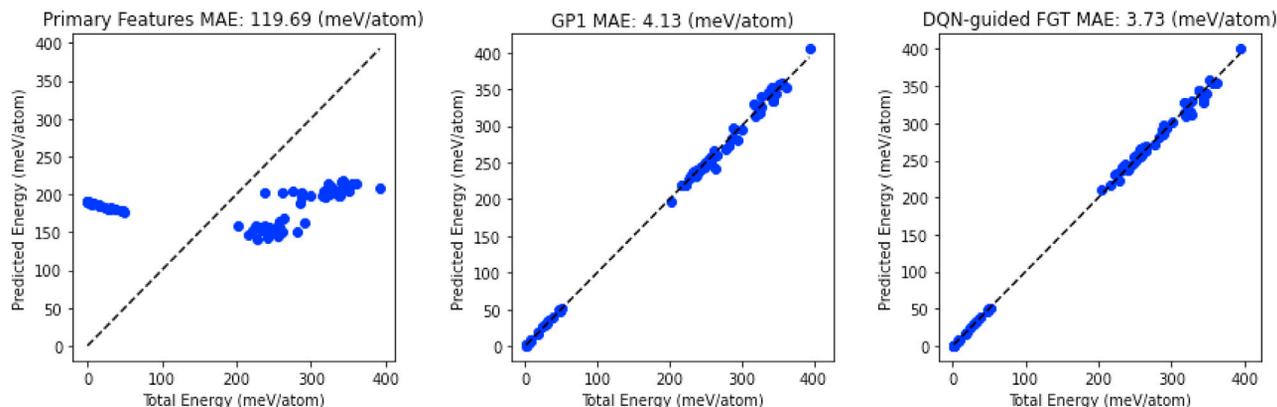
$d$  and  $(f^d)^*$  as the selected optimal feature for the  $d$ th dimension of  $(F^d)^*$ . The FGT exploration aims to search for the best descriptors  $(f^1)^*, (f^2)^*, \dots$  one by one based on the testing accuracy given the observed data. The corresponding complete AFE procedure constructs the feature subspace  $F^d$  sequentially as the search space of each  $(f^d)^*$  exploration, starting from the root node  $F_0$  with the primary feature set. At each node  $F_i$ , we would like to learn a generation policy  $\pi$  to choose an operation  $\varphi_i$  to generate the new descriptor set  $F_j (j > i)$  as the corresponding child node, with which the current optimal  $(F^d)'$  and  $A_L\{(F^d)', y\}$  will be updated accordingly. The FGT will grow by repeating the operations above until it reaches the maximum complexity  $c_{max}$ .

To learn the FGT generation policy  $\pi$ , we adopt a DQN with experience replay.<sup>152</sup> Formally, we define the states, actions and rewards as follows:

- state  $F_i^d$ , denoting a set of primary features or generated descriptors when looking for the  $d$ th optimal descriptor;
- action  $\pi(F_i^d) = \varphi_i$ , denoting an operation in the set  $O$ ;
- reward:  $R(F_i^d, \varphi_i) = \max_{F'} (1.001 - A_L\{F', y\})^{-1}$ , where  $0 \leq A_L\{F', y\} \leq 1$ .

The pseudo-code for learning DQN-based FGT exploration is given in Algorithm 1. To have a flexible exploration procedure for performance-complexity trade-off and incorporation of prior knowledge, each  $(f^d)^*$  in  $F^*$  can be chosen from the top  $n$  features with highest rewards in the corresponding feature subspace  $F^d$ , composing a candidate set  $\mathbb{S}^d$ . So  $(F^{d-1})^*$  can have multiple combinations according to the whole candidate sets  $\mathbb{S} = \{\mathbb{S}^1, \dots, \mathbb{S}^{d-1}\}$ , and  $F'$  also has multiple combinations according to different  $(F^{d-1})^*$  and  $f^d$ . Consequently the reward is computed as the maximum reward over  $F'$ .

Note that when we apply binary operations on  $F_i$ , beside the one feature in the  $F_i$ , we have to choose another feature in the



**Figure 7. Copper energy regression results for different interatomic potential models**

generated descriptor space, leading to the exponentially exploding number of new descriptors. To achieve appropriate performance-complexity trade-off, we introduce flexible batch sampling to randomly sample a feature subspace  $B$  from  $\mathbb{F}$  as a “batch set” each time and enumerate  $f^s$  only from  $B$  and take the maximum reward from all of the combinations as the reward. When prior knowledge is available as physics constraints on applying corresponding operations to specific feature groups, this batch sampling procedure can naturally take care of them.

### AFE to learn interatomic potential models for copper

First-principles DFT<sup>123,124</sup> has been extensively applied in materials science, physics, and chemistry. However, it is often constrained to simulate materials with 100–1,000 atoms for several thousands *ab initio* molecular dynamics steps, covering about 10 picoseconds. In contrast, classic interatomic potentials have been widely adopted in the past, allowing large-scale molecular dynamics simulation of millions of atoms for millions of time steps (that is, covering >10 nanoseconds). However, the construction of the functional form and the optimization of the corresponding parameters of classic potentials are highly nontrivial. Recently, developing and training neural network potentials based on first-principles DFT calculations of relatively small systems has become an important research direction in the atomistic simulation in combination with active learning.<sup>153–156</sup> Very recently, a feature engineering method has been pursued where genetic programming was applied to develop fast and accurate classic interatomic potentials with explicit functional forms from physically meaningful hypothesis space.<sup>136</sup> Particularly, genetic programming was applied to optimize the exact functional form of pairwise and many-body potentials as well as other potential forms, highlighting an important avenue toward the development of physics-constrained models with analytic functional form.

Different from the above genetic programming approach, we have adopted our FGT-based AFE and evaluated its ability to find potential models from data generated by DFT. To compare with the genetic programming symbolic regression approach,<sup>136</sup> we used the same 150 snapshots of 32-atom DFT molecular dynamics simulations on fcc copper in Hernandez et al.,<sup>136</sup> where each snapshot was generated every 100 steps with a time step of 1 fs. We adopted the same 150 snapshots, including 50 snapshots from *ab initio* molecular dynamics performed at 300 K in

the canonical (NVT) ensemble, 50 snapshots at 1,400 K in the NVT ensemble, and 50 snapshots at 1,400 K in the isothermal-isobaric (NPT) ensemble with pressure at 100 kPa. The 150 total energies calculated by Hernandez et al.<sup>136</sup> were considered the target output of interest,<sup>157</sup> with the random split of 125 structures and their corresponding total energies for training, and the remaining 125 structures and total energies for validation, for feature engineering.

We have compared our AFE with the recently developed physics-informed genetic programming method<sup>136</sup> to arrive at analytical many-body classical interatomic potential models. Figure 7 shows the plots of predicted total energy of these different copper structures vs. the simulated total energy based on primary features (left), genetic-programming-generated descriptors (center), and AFE-generated descriptors (right) on the same held-out testing data. With the same simulated molecular dynamics data and experimental setup in the paper, our AFE has achieved the total energy prediction with a mean absolute error (MAE) of 3.73 meV/atom within 12 h. By contrast, the reported model GP1 by genetic programming in Hernandez et al.<sup>136</sup> had a prediction MAE of 4.13 meV/atom after 360 CPU hours on the same training and test sets.

Our proposed AFE strategies approximate the expected future reward of engineered descriptors through DQN-based policy learning and replace the exhaustive feature generation by DQN-guided FGT exploration considering physics prior knowledge. Consequently, our AFE enhances scalability and computational efficiency without sacrificing prediction performance, as demonstrated in the reported experiment as well as other materials science experiments have demonstrated the potential of our DQN-guided FGT exploration in reducing the runtime and enhancing the scalability for AFE. More importantly, the engineered descriptors are interpretable with the corresponding lists of algebraic operations on the original primary features. Our physics-constrained AFE aims at generalizable learning under data scarcity and uncertainty. Interpretable instead of “black box” learning helps new knowledge discovery and better decision-making.

### Conclusions and future work

When facing real-world complex systems in various science and engineering domains—such as complex materials systems,

which are our focus in this paper—where acquisition of ample data is practically difficult and formidably expensive, currently existing ML methods fail to produce reliable and generalizable predictions. To cope with the current shortcomings of existing SP and ML schemes in materials discovery, there is a pressing need for novel methods that enable robust optimal decision-making under challenging conditions, such as small data size, enormous system complexity, nonstationarity, as well as data and model uncertainty. In this paper, we have presented recent efforts in knowledge-driven learning, optimization, and experimental design, and we have also provided their historical context against the rich research history in the SP community revolving around robust filtering and control, which can probably be dated back to the 1950s. Specifically, we have shown several examples in an objective-based UQ framework—via MOCU—to develop sciML methods to address the aforementioned challenges in accelerating materials discovery, focusing on learning and experimental design under uncertainty. The problems of studying complex systems will persist in diverse science and engineering disciplines, and we expect that the learning and optimization schemes based on objective-based UQ presented in this paper would provide a useful guideline for developing new sciML methods that more effectively incorporate scientific knowledge, design surrogate ML models that are better suited for the given systems under study, and devise computational solutions that are more scalable and efficient.

#### ACKNOWLEDGMENTS

The authors would like to thank the collaborators involved in the reviewed projects: Shahin Boluki, Guang Zhao, Mingzhou Fan, Ziyu Xiang, Tao Hu, Roozbeh Dehghannasiri, Nathan Wilson, and Daniel Willhelm. The authors were supported in part by National Science Foundation (NSF) awards CCF-1553281, IIS-1812641, OAC-1835690, DMR-1753054, DMR-2119103, CMMI-2226908, SHF-2215573, and IIS-2212419 as well as by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Mathematical Multifaceted Integrated Capability Centers program under award DE-SC0019303.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

#### REFERENCES

- Kaufman, L., and Ågren, J. (2014). CALPHAD, first and second generation—Birth of the materials genome. *Scripta Mater.* 70, 3–6. <https://doi.org/10.1016/j.scriptamat.2012.12.003>.
- McDowell, D.L., and Kaldindí, S.R. (2016). The materials innovation ecosystem: a key enabler for the materials genome initiative. *MRS Bull.* 41, 326–337. <https://doi.org/10.1557/mrs.2016.61>.
- Ghiringhelli, L.M., Vybiral, J., Levchenko, S.V., Draxl, C., and Scheffler, M. (2015). Big data of materials science: Critical role of the descriptor. *Phys. Rev. Lett.* 114, 105503. <https://doi.org/10.1103/PhysRevLett.114.105503>.
- Kim, C., Pilania, G., and Ramprasad, R. (2016). From organized high-throughput data to phenomenological theory using machine learning: The example of dielectric breakdown. *Chem. Mater.* 28, 1304–1311. <https://doi.org/10.1021/acs.chemmater.5b04109>.
- Dehghannasiri, R., Xue, D., Balachandran, P.V., Yousefi, M.R., Dalton, L.A., Lookman, T., and Dougherty, E.R. (2017). Optimal experimental design for materials discovery. *Comput. Mater. Sci.* 129, 311–322. <https://doi.org/10.1016/j.commatsci.2016.11.041>.
- S. Yip, ed. (2005). *Handbook of Materials Modeling* (Springer). <https://doi.org/10.1007/978-1-4020-3286-8>.
- Krenn, M., Pollice, R., Guo, S.Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., Dos Passos Gomes, G., Häse, F., Jinich, A., Nigam, A.K., et al. (2022). On scientific understanding with artificial intelligence. *Nat. Rev. Phys.* 4, 761–769. <https://doi.org/10.1038/s42254-022-00518-3>.
- Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature* 620, 47–60. <https://doi.org/10.1038/s41586-023-06221-2>.
- Zhang, X., Wang, L., Jacob, H., Luo, Y., Fu, C., Xie, Y., Liu, M., Lin, Y., Zhao, X., Yan, K., et al. (2023). Artificial intelligence for science in quantum, atomistic, and continuum systems. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2307.08423>.
- Choudhary, A., Fox, G., and Hey, T. (2023). *Artificial Intelligence for Science: A Deep Learning Revolution* (World Scientific Publishing Co. Pte Ltd). <https://doi.org/10.1142/13123>.
- Arróyave, R., Khatamsaz, D., Vela, B., Couperthwaite, R., Molkeri, A., Singh, P., Johnson, D.D., Qian, X., Srivastava, A., and Allaire, D. (2022). A perspective on Bayesian methods applied to materials discovery and design. *MRS Communications* 12, 1037–1049. <https://doi.org/10.1557/s43579-022-00288-0>.
- Fuhr, A.S., and Sumpter, B.G. (2022). Deep generative models for materials discovery and machine learning-accelerated innovation. *Front. Mater.* 9, 865270. <https://doi.org/10.3389/fmats.2022.865270>.
- Pyzer-Knapp, E.O., Pitera, J.W., Staar, P.W.J., Takeda, S., Laino, T., Sanders, D.P., Sexton, J., Smith, J.R., and Curioni, A. (2022). Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *npj Comput. Mater.* 8, 84. <https://doi.org/10.1038/s41524-022-00765-z>.
- Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (Wiley). <https://doi.org/10.7551/mitpress/2946.001.0001>.
- Kailath, T., Sayed, A.H., and Hassibi, B. (2000). *Linear Estimation* (Prentice-Hall).
- Betts, J.T. (2010). *Practical Methods for Optimal Control Using Nonlinear Programming*, 2nd ed. (SIAM Press). <https://doi.org/10.1137/1.9780898718577>.
- Dalton, L., and Dougherty, E.R. (2020). *Optimal Bayesian Classification* (SPIE Press).
- Rockafellar, R.T., and Uryasev, S. (2013). The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science* 18, 33–53. <https://doi.org/10.1016/j.sorms.2013.03.001>.
- Spantini, A., Cui, T., Willcox, K., Tenorio, L., and Marzouk, Y. (2017). Goal-oriented optimal approximations of Bayesian linear inverse problems. *SIAM J. Sci. Comput.* 39, S167–S196. <https://doi.org/10.1137/16m1082123>.
- Kuznetsov, V.P. (1976). Stable detection when signal and spectrum of normal noise are inaccurately known. *Telecommun. Radio Eng.* 30, 58–64.
- Kassam, S.A., and Lim, T.L. (1977). Robust Wiener filters. *J. Franklin Inst.* 304, 171–185. [https://doi.org/10.1016/0016-0032\(77\)90011-4](https://doi.org/10.1016/0016-0032(77)90011-4).
- Poor, H. (1980). Poor. On robust Wiener filtering. *IEEE Trans. Automat. Control* 25, 531–536. <https://doi.org/10.1109/TAC.1980.1102349>.
- Chen, Y., and Chen, B. (1994). Minimax robust deconvolution filters under stochastic parametric and noise uncertainties. *IEEE Trans. Signal Process.* 42, 32–45. <https://doi.org/10.1109/78.258119>.
- Verdu, S., and Poor, H. (1984). Minimax linear observers and regulators for stochastic systems with uncertain second-order statistics. *IEEE Trans. Automat. Control* 29, 499–511. <https://doi.org/10.1109/TAC.1984.1103576>.
- Li, T., Yi, X., Carmanis, C., and Ravikumar, P. (2017). Minimax Gaussian classification & clustering. In *Proceedings of the 20th International*

- Conference on Artificial Intelligence and Statistics, Volume 54 of Proceedings of Machine Learning Research, 1–9, A. Singh and J. Zhu, eds. <https://proceedings.mlr.press/v54/li17a.html>.
26. Bertsimas, D., and Paskov, I. (2020). Stable regression: On the power of optimization over randomization. *J. Mach. Learn. Res.* 21, 1–25. <http://jmlr.org/papers/v21/19-408.html>.
  27. Kalman, R.E., and Bucy, R.S. (1961). New results in linear filtering and prediction theory. *J. Basic Eng.* 83, 95–108. <https://doi.org/10.1115/1.3658902>.
  28. Mehra, R. (1972). Approaches to adaptive filtering. *IEEE Trans. Automat. Control* 17, 693–698. <https://doi.org/10.1109/TAC.1972.1100100>.
  29. Morris, J. (1976). The Kalman filter: A robust estimator for some classes of linear quadratic problems. *IEEE Trans. Inf. Theor.* 22, 526–534. <https://doi.org/10.1109/TIT.1976.1055611>.
  30. Bellman, R., and Kalaba, R. (1960). Dynamic programming and adaptive processes: Mathematical foundation. *IRE Trans. Automatic Control* 5, 5–10. <https://doi.org/10.1109/TAC.1960.6429288>.
  31. Silver, E.A. (1963). *Markovian Decision Processes with Uncertain Transition Probabilities or Rewards* (MIT Cambridge Operations Research Center). Technical report.
  32. Martin, J.J. (1967). *Bayesian Decision Problems and Markov Chains* (Wiley).
  33. Grigoryan, A.M., and Dougherty, E.R. (2001). Bayesian robust optimal linear filters. *Signal Process.* 81, 2503–2521. [https://doi.org/10.1016/S0165-1684\(01\)00144-X](https://doi.org/10.1016/S0165-1684(01)00144-X).
  34. Dougherty, E.R., Hua, J., Xiong, Z., and Chen, Y. (2005). Optimal robust classifiers. *Pattern Recogn.* 38, 1520–1532. <https://doi.org/10.1016/j.patcog.2005.01.019>.
  35. Anthony, C., and Atkinson. (2015). *Optimal Design, 1–17* (American Cancer Society). <https://doi.org/10.1002/9781118445112.stat04090.pub2>.
  36. Sebastiani, P., and Wynn, H.P. (2000). Maximum entropy sampling and optimal Bayesian experimental design. *J. Roy. Stat. Soc. B* 62, 145–157. <https://doi.org/10.1111/1467-9868.00225>.
  37. Musmann, S., and Liang, P. (2018). *On the relationship between data efficiency and error for uncertainty sampling* Jennifer Dy and Andreas Krause. In *Proceedings of the 35th International Conference on Machine Learning, volume 80 of Proceedings of Machine Learning Research* (Stockholm: PMLR), pp. 3674–3682. PMLR.
  38. Fedorov, V.V., and Leonov, S.L. (2014). *Optimal Design for Nonlinear Response Models* (Chapman and Hall/CRC Press). <https://doi.org/10.1201/b15054>.
  39. Duarte, B.P.M., Wong, W.K., and Atkinson, A.C. (2015). A semi-infinite programming based algorithm for determining t-optimum designs for model discrimination. *J. Multivariate Anal.* 135, 11–24. <https://doi.org/10.1016/j.jmva.2014.11.006>.
  40. Dennis, V. (1972). Lindley. *Bayesian Statistics, A Review* (SIAM).
  41. Huan, X., and Marzouk, Y. (2014). Gradient-based stochastic optimization methods in Bayesian experimental design. *Int. J. Uncertain. Quantification* 4, 479–510. <https://doi.org/10.1615/int.j.uncertaintyquantification.2014006730>.
  42. Zhao, G., Qian, X., Yoon, B.-J., Alexander, F.J., and Dougherty, E.R. (2020). Model-based robust filtering and experimental design for stochastic differential equation systems. *IEEE Trans. Signal Process.* 68, 3849–3859. <https://doi.org/10.1109/TSP.2020.3001384>.
  43. Foster, A., Martin, J., O’Meara, M., Yee, W.T., and Rainforth, T. (2020). A unified stochastic gradient approach to designing Bayesian-optimal experiments. . *Proceedings of the 23th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 108, pp. 2959–2969. <https://proceedings.mlr.press/v108/foster20a.html>.
  44. Zhao, G., Dougherty, E., Yoon, B.-J., Alexander, F., and Qian, X. (2021). *Uncertainty-aware active learning for optimal Bayesian classifier*. In *9th International Conference on Learning Representations (ICLR)*.
  45. Jones, D.R., Schonlau, M., and Welch, W.J. (1998). Efficient global optimization of expensive black-box functions. *J. Global Optim.* 13, 455–492. <https://doi.org/10.1023/a:1008306431147>.
  46. Frazier, P.I., Powell, W.B., and Dayanik, S. (2008). A knowledge-gradient policy for sequential information collection. *SIAM J. Control Optim.* 47, 2410–2439. <https://doi.org/10.1137/070693424>.
  47. Frazier, P., Powell, W., and Dayanik, S. (2009). The knowledge-gradient policy for correlated normal beliefs. *Inf. J. Comput.* 21, 599–613. <https://doi.org/10.1287/ijoc.1080.0314>.
  48. Denil, M., Agrawal, P., Kulkarni, T.D., Tom, E., Battaglia, P., and de Freitas, N. (2017). *Learning to perform physics experiments via deep reinforcement learning*. In *International Conference on Learning Representations (ICLR)*.
  49. Hadjidakis, P.E., Angelikopoulos, P., Papadimitriou, C., and Koumoutsakos, P. (2015). II4U: A high performance computing framework for Bayesian uncertainty quantification of complex models. *Journal of Computational Physics* 284, 1–21.
  50. Adams, B.M., Bohnhoff, W.J., Dalbey, K.R., Ebeida, M.S., Eddy, J.P., El-dred, M.S., Hooper, R.W., Hough, P.D., Hu, K.T., Jakeman, J.D., et al. (2021). *A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.15 User’s Manual*. Sandia Technical Report SAND2020-12495. <https://dakota.sandia.gov/content/about>.
  51. Xue, D., Balachandran, P.V., Hogden, J., Theiler, J., Xue, D., and Lookman, T. (2016). Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* 7, 11241. <https://doi.org/10.1038/ncomms11241>.
  52. Xue, D., Balachandran, P.V., Yuan, R., Hu, T., Qian, X., Dougherty, E.R., and Lookman, T. (2016). Accelerated search for BaTiO<sub>3</sub>-based piezoelectrics with vertical morphotropic phase boundary using Bayesian learning. *Proc. Natl. Acad. Sci. USA* 113, 13301–13306. <https://doi.org/10.1073/pnas.1607412113>.
  53. Peter, I.F., and Wang, J. (2016). Bayesian optimization for materials design. In *Information Science for Materials Discovery and Design* (Springer), pp. 45–75. [https://doi.org/10.1007/978-3-319-23871-5\\_3](https://doi.org/10.1007/978-3-319-23871-5_3).
  54. Qian, X., and Dougherty, E.R. (2016). Bayesian regression with network prior: Optimal Bayesian filtering perspective. *IEEE Trans. Signal Process.* 64, 6243–6253. <https://doi.org/10.1109/TSP.2016.2605072>.
  55. Ueno, T., Rhone, T.D., Hou, Z., Mizoguchi, T., and Tsuda, K. (2016). COMBO: An efficient bayesian optimization library for materials science. *Materials Discovery* 4, 18–21. <https://doi.org/10.1016/j.md.2016.04.001>.
  56. Seko, A., Maekawa, T., Tsuda, K., and Tanaka, I. (2014). Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids. *Phys. Rev. B* 89, 054303. <https://doi.org/10.1103/PhysRevB.89.054303>.
  57. Ju, S., Shiga, T., Feng, L., Hou, Z., Tsuda, K., and Shiomi, J. (2017). Designing nanostructures for phonon transport via Bayesian optimization. *Phys. Rev. X* 7, 021024. <https://doi.org/10.1103/PhysRevX.7.021024>.
  58. Gopakumar, A.M., Balachandran, P.V., Xue, D., Gubernatis, J.E., and Lookman, T. (2018). Multi-objective optimization for materials discovery via adaptive design. *Sci. Rep.* 8, 3738. <https://doi.org/10.1038/s41598-018-21936-3>.
  59. Khatamsaz, D., Molkeri, A., Couperthwaite, R., James, J., Arróyave, R., Srivastava, A., and Allaire, D. (2021). Adaptive active subspace-based efficient multifidelity materials design. *Mater. Des.* 209, 110001. <https://doi.org/10.1016/j.matdes.2021.110001>.
  60. Castillo, A.R., and Kalidindi, S.R. (2021). Bayesian estimation of single ply anisotropic elastic constants from spherical indentations on multi-laminate polymer-matrix fiber-reinforced composite samples. *Meccanica* 56, 1575–1586. <https://doi.org/10.1007/s11012-020-01154-w>.
  61. Marshall, A., and Kalidindi, S.R. (2021). Autonomous development of a machine-learning model for the plastic response of two-phase composites from micromechanical finite element models. *JOM* 73, 2085–2095. <https://doi.org/10.1007/s11837-021-04696-w>.

62. Honarmandi, P., Hossain, M.A., Arroyave, R., and Baxevanis, T. (2021). A top-down characterization of NiTi single-crystal inelastic properties within confidence bounds through Bayesian inference. *Shap. Mem. Superelasticity* 7, 50–64. <https://doi.org/10.1007/s40830-021-00311-8>.
63. Ladygin, V., Beniya, I., Makarov, E., and Shapeev, A. (2021). Bayesian learning of thermodynamic integration and numerical convergence for accurate phase diagrams. *Phys. Rev. B* 104, 104102. <https://doi.org/10.1103/PhysRevB.104.104102>.
64. Olivier, A., Shields, M.D., and Graham-Brady, L. (2021). Bayesian neural networks for uncertainty quantification in data-driven materials modeling. *Comput. Methods Appl. Mech. Eng.* 386, 114079. <https://doi.org/10.1016/j.cma.2021.114079>.
65. Yoon, B.-J., Qian, X., and Dougherty, E.R. (2013). Quantifying the objective cost of uncertainty in complex dynamical systems. *IEEE Trans. Signal Process.* 61, 2256–2266. <https://doi.org/10.1109/TSP.2013.2251336>.
66. Yoon, B.-J., Qian, X., and Dougherty, E.R. (2021). Quantifying the multi-objective cost of uncertainty. *IEEE Access* 9, 80351–80359. <https://doi.org/10.1109/ACCESS.2021.3085486>.
67. Dalton, L.A., and Dougherty, E.R. (2014). Intrinsically optimal Bayesian robust filtering. *IEEE Trans. Signal Process.* 62, 657–670. <https://doi.org/10.1109/TSP.2013.2291213>.
68. Box, G.E.P., and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis* (Wiley). <https://doi.org/10.1002/9781118033197>.
69. Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag). <https://doi.org/10.1007/978-1-4757-4286-2>.
70. Christopher, M. (2006). Bishop. *Pattern Recognition and Machine Learning* (Springer). <https://link.springer.com/book/9780387310732>.
71. Dougherty, E.R., Zollanvari, A., and Braga-Neto, U.M. (2011). The illusion of distribution-free small-sample classification in genomics. *Curr. Genom.* 12, 333–341. <https://doi.org/10.2174/138920211796429763>.
72. Dougherty, E.R., and Dalton, L.A. (2013). Scientific knowledge is possible with small-sample classification. *EURASIP J. Bioinf. Syst. Biol.* 2013, 10–12. <https://doi.org/10.1186/1687-4153-2013-10>.
73. Coveney, P.V., Dougherty, E.R., and Highfield, R.R. (2016). Big data need big theory too. *Phil. Trans. R. Soc. A* 374, 20160153. <https://doi.org/10.1098/rsta.2016.0153>.
74. Jaynes, E.T. (1980). *What Is the Question? Bayesian Statistics*.
75. Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. A Math. Phys. Sci.* 186, 453–461. <https://doi.org/10.1098/rspa.1946.0056>.
76. Zellner, A.; University of Chicago (1995). *Department of Economics. Past and Recent Results on Maximal Data Information Priors. Working Paper Series in Economics and Econometrics* (University of Chicago, Graduate School of Business, Department of Economics).
77. Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Ann. Stat.* 11, 416–431. <https://doi.org/10.1214/aos/1176346150>.
78. Rodriguez, C.C. (2002). Entropic priors for discrete probabilistic networks and for mixtures of Gaussian models. *AIP Conf. Proc.* <https://doi.org/10.1063/1.1477063>.
79. Berger, J.O., and Bernardo, J.M. (1992). On the development of reference priors. *Bayesian statistics* 4, 35–60.
80. Spall, J.C., and Hill, S.D. (1990). Least-informative Bayesian prior distributions for finite samples based on information theory. *IEEE Trans. Automat. Control* 35, 580–583. <https://doi.org/10.1109/CDC.1989.70640>.
81. Bernardo, J.M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Stat. Soc. B* 41, 113–128. <https://doi.org/10.1111/j.2517-6161.1979.tb01066.x>.
82. Kass, R.E., and Wasserman, L. (1996). The selection of prior distributions by formal rules. *J. Am. Stat. Assoc.* 91, 1343–1370. <https://doi.org/10.1080/01621459.1996.10477003>.
83. Berger, J.O., Bernardo, J.M., and Sun, D. (2012). Objective priors for discrete parameter spaces. *J. Am. Stat. Assoc.* 107, 636–648. <https://doi.org/10.1080/01621459.2012.682538>.
84. Jaynes, E.T. (1957). Information theory and statistical mechanics. *Phys. Rev.* 106, 620–630. <https://doi.org/10.1103/PhysRev.106.620>.
85. Jaynes, E. (1968). Jaynes. Prior probabilities. *IEEE Trans. Syst. Sci. Cybern.* 4, 227–241. <https://doi.org/10.1109/TSSC.1968.300117>.
86. Zellner, A. (1996). Models, prior information, and Bayesian analysis. *J. Econom.* 75, 51–68. [https://doi.org/10.1016/0304-4076\(95\)01768-2](https://doi.org/10.1016/0304-4076(95)01768-2).
87. Werner, K., Jansson, M., and Stoica, P. (2008). On estimation of covariance matrices with kronecker product structure. *IEEE Trans. Signal Process.* 56, 478–491. <https://doi.org/10.1109/TSP.2007.907834>.
88. Wiesel, A., Eldar, Y.C., and Hero, A.O. (2010). Covariance estimation in decomposable Gaussian graphical models. *IEEE Trans. Signal Process.* 58, 1482–1492. <https://doi.org/10.1109/TSP.2009.2037350>.
89. Eldar, Y.C. (2009). Generalized SURE for exponential families: Applications to regularization. *IEEE Trans. Signal Process.* 57, 471–481. <https://doi.org/10.1109/TSP.2008.2008212>.
90. Burg, J.P., Luenberger, D.G., and Wenger, D.L. (1982). Estimation of structured covariance matrices. *Proc. IEEE* 70, 963–974. <https://doi.org/10.1109/PROC.1982.12427>.
91. Wei, P., and Pan, W. (2012). Bayesian joint modeling of multiple gene networks and diverse genomic data to identify target genes of a transcription factor. *Ann. Appl. Stat.* 6, 334–355. <https://doi.org/10.1214/11-aos502>.
92. Wiesel, A., and Hero, A.O. (2010). Distributed covariance estimation in Gaussian graphical models. In *2010 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)* (IEEE), pp. 193–196.
93. Kalidindi, S.R. (2015). *Hierarchical Materials Informatics: Novel Analytics for Materials Data* (Elsevier). <https://doi.org/10.1016/C2012-0-07337-1>.
94. Ghoreishi, S.F., Molkeri, A., Srivastava, A., Arroyave, R., and Allaire, D. (2018). Multi-information source fusion and optimization to realize ICME: Application to dual-phase materials. *J. Mech. Des. N. Y.* 140. <https://doi.org/10.1115/1.4041034>.
95. Kalidindi, S.R. (2019). A Bayesian framework for materials knowledge systems. *MRS Communications* 9, 518–531. <https://doi.org/10.1557/mrc.2019.56>.
96. Markland, T.E., and Ceriotti, M. (2018). Markland and Michele Ceriotti. Nuclear quantum effects enter the mainstream. *Nat. Rev. Chem* 2, 0109. <https://doi.org/10.1038/s41570-017-0109>.
97. Xie, T., and Grossman, J.C. (2018). Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* 120, 145301. <https://doi.org/10.1103/PhysRevLett.120.145301>.
98. Park, C.W., and Wolverton, C. (Jun 2020). Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Phys. Rev. Mater.* 4, 063801. <https://doi.org/10.1103/PhysRevMaterials.4.063801>.
99. Chen, C., Ye, W., Zuo, Y., Zheng, C., and Ong, S.P. (2019). Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* 31, 3564–3572. URL <https://doi.org/10.1021/acs.chemmater.9b01294>. <https://doi.org/10.1021/acs.chemmater.9b01294>.
100. Kohn, W. (1996). Density functional and density matrix method scaling linearly with the number of atoms. *Phys. Rev. Lett.* 76, 3168–3171. <https://doi.org/10.1103/PhysRevLett.76.3168>.
101. Prodan, E., and Kohn, W. (2005). Nearsightedness of electronic matter. *Proc. Natl. Acad. Sci. USA* 102, 11635–11638. <https://doi.org/10.1073/pnas.0505436102>.
102. Schmidt, J., Marques, M.R.G., Botti, S., and Marques, M.A.L. (2019). Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* 5, 83–3960. <https://doi.org/10.1038/s41524-019-0221-0>.

103. Esfahani, M.S., and Dougherty, E.R. (2014). Incorporation of biological pathway knowledge in the construction of priors for optimal Bayesian classification. *IEEE ACM Trans. Comput. Biol. Bioinf* *17*, 202–218. <https://doi.org/10.1109/TCBB.2013.143>.
104. Esfahani, M.S., Shahrokh, E., and Dougherty, E.R. (2015). An optimization-based framework for the transformation of incomplete biological knowledge into a probabilistic structure and its application to the utilization of gene/protein signaling pathways in discrete phenotype classification. *IEEE ACM Trans. Comput. Biol. Bioinf* *12*, 1304–1321. <https://doi.org/10.1109/TCBB.2015.2424407>.
105. Boluki, S., Esfahani, M.S., Qian, X., and Dougherty, E.R. (2019). Constructing pathway-based priors within a constructing pathway-based priors within a Gaussian mixture model for Bayesian regression and classification. *IEEE ACM Trans. Comput. Biol. Bioinf* *16*, 524–537. <https://doi.org/10.1109/TCBB.2017.2778715>.
106. Boluki, S., Esfahani, M.S., Qian, X., and Dougherty, E.R. (2017b). Incorporating biological prior knowledge for Bayesian learning via maximal knowledge-driven information priors. *BMC Bioinf.* *18* (Suppl 14), 552. <https://doi.org/10.1186/s12859-017-1893-4>.
107. Guiasu, S., and Shenitzer, A. (1985). The principle of maximum entropy. *Math. Intel.* *7*, 42–48. <https://doi.org/10.1007/bf03023004>.
108. Heitmann, A.A., and Rossetti, G.A. (2014). Thermodynamics of ferroelectric solid solutions with morphotropic phase boundaries. *J. Am. Ceram. Soc.* *97*, 1661–1685. <https://doi.org/10.1111/jace.12979>.
109. Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. (1999). Bayesian model averaging: A tutorial. *Stat. Sci.* *4*, 382–417. <https://www.jstor.org/stable/2676803>.
110. Wasserman, L. (2000). Bayesian model selection and model averaging. *J. Math. Psychol.* *44*, 92–107. <https://doi.org/10.1006/jmps.1999.1278>.
111. Clarke, B. (2003). Comparing Bayes model averaging and stacking when model approximation error cannot be ignored. *J. Mach. Learn. Res.* *4*, 683–712. <https://doi.org/10.1162/153244304773936090>.
112. Clyde, M.A., Ghosh, J., and Littman, M.L. (2011). Bayesian adaptive sampling for variable selection and model averaging. *J. Comput. Graph Stat.* *20*, 80–101. <https://doi.org/10.1198/jcgs.2010.09049>.
113. George, E., and Foster, D. (2000). Calibration and empirical Bayes variable selection. *Biometrika* *87*, 731–747. <https://doi.org/10.1093/biomet/87.4.731>.
114. Yang, Y. (2003). Regression with multiple candidate models: Selecting or mixing? *Stat. Sin.* *13*, 783–809.
115. Monteith, K., Carroll, J.L., Seppi, K., and Martinez, T. (2011). Turning Bayesian model averaging into Bayesian model combination. In *The 2011 International Joint Conference on Neural Networks*, pp. 2657–2663. <https://doi.org/10.1109/IJCNN.2011.6033566>.
116. Madigan, D., and Raftery, A.E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Am. Stat. Assoc.* *89*, 1535–1546. <https://doi.org/10.2307/2291017>.
117. Dehghannasiri, R., Yoon, B.-J., and Dougherty, E.R. (2015). Optimal experimental design for gene regulatory networks in the presence of uncertainty. *IEEE ACM Trans. Comput. Biol. Bioinf* *12*, 938–950. <https://doi.org/10.1109/TCBB.2014.2377733>.
118. Boluki, S., Qian, X., and Dougherty, E.R. (2019). Experimental design via generalized mean objective cost of uncertainty. *IEEE Access* *7*, 2223–2230. <https://doi.org/10.1109/ACCESS.2018.2886576>.
119. Rasmussen, C.E., and Williams, C.K.I. (2005). *Gaussian Processes for Machine Learning* (Adaptive Computation and Machine Learning) (The MIT Press). <https://doi.org/10.7551/mitpress/3206.001.0001>.
120. Talapatra, A., Boluki, S., Duong, T., Qian, X., Dougherty, E., and Arroyave, R. (2018). Autonomous efficient experiment design for materials discovery with Bayesian model averaging. *Phys. Rev. Mater.* *2*, 113803. <https://doi.org/10.1103/PhysRevMaterials.2.113803>.
121. Barsoud, M.W. (2013). *MAX Phases: Properties of Machinable Ternary Carbides and Nitrides* (John Wiley & Sons). <https://doi.org/10.1002/9783527654581>.
122. Emmerich, M.T.M., André, H.D., and Jan, W.K. (2011). Hypervolume-based expected improvement: Monotonicity properties and exact computation. In *2011 IEEE Congress of Evolutionary Computation (CEC)* (IEEE), pp. 2147–2154.
123. Hohenberg, P., and Kohn, W. (1964). Inhomogeneous electron gas. *Phys. Rev.* *136*, B864–B871. <https://doi.org/10.1103/PhysRev.136.B864>.
124. Kohn, W., and Sham, L.J. (1965). Self-consistent equations including exchange and correlation effects. *Phys. Rev.* *140*, A1133–A1138. <https://doi.org/10.1103/PhysRev.140.A1133>.
125. Talapatra, A., Boluki, S., Honarmandi, P., Solomou, A., Zhao, G., Ghor-eishi, S.F., Molkeri, A., Allaire, D., Srivastava, A., Qian, X., et al. (2019). Experiment design frameworks for accelerated discovery of targeted materials across scales. *Front. Mater.* *6*, 82. <https://doi.org/10.3389/fmats.2019.00082>.
126. Bacon, F. (2000). *The New Organon* (Cambridge University Press). <https://doi.org/10.1017/CBO9781139164030.006>.
127. Dehghannasiri, R., Yoon, B.-J., and Dougherty, E.R. (2015). Efficient experimental design for uncertainty reduction in gene regulatory networks. *BMC Bioinf.* *16*, S2. <https://doi.org/10.1186/1471-2105-16-S13-S2>.
128. Hong, Y., Kwon, B., and Yoon, B.-J. (2021). Optimal experimental design for uncertain systems based on coupled differential equations. *IEEE Access* *9*, 53804–53810. <https://doi.org/10.1109/ACCESS.2021.3071038>.
129. Woo, H.-M., Hong, Y., Kwon, B., and Yoon, B.-J. (2021). Accelerating optimal experimental design for robust synchronization of uncertain kuramoto oscillator model using machine learning. *IEEE Trans. Signal Process.* *69*, 6473–6487. <https://doi.org/10.1109/TSP.2021.3130967>.
130. Broumand, A., Esfahani, M.S., Yoon, B.-J., and Dougherty, E.R. (2015). Discrete optimal Bayesian classification with error-conditioned sequential sampling. *Pattern Recogn.* *48*, 3766–3782. <https://doi.org/10.1016/j.patcog.2015.03.023>.
131. Zhao, G., Dougherty, E., Yoon, B.-J., Alexander, F., and Qian, X. (2021). Bayesian active learning by soft mean objective cost of uncertainty. In *24th International Conference on Artificial Intelligence and Statistics (AISTATS)*.
132. Ben-Gal, I., and Caramanis, M. (2002). Sequential DOE via dynamic programming. *IIE Trans.* *34*, 1087–1100. <https://doi.org/10.1023/A:1019670414725>.
133. Warren, B. (2011). *Powell. Approximate Dynamic Programming: Solving the Curses of Dimensionality*, 2nd edition (John Wiley & Sons, Inc.). <https://doi.org/10.1002/9781118029176>.
134. Huan, X., and Marzouk, Y.M. (2016). Sequential Bayesian optimal experimental design via approximate dynamic programming. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1604.08320>.
135. Zhao, G., Dougherty, E., Yoon, B.-J., Alexander, F., and Qian, X. (2021). Efficient active learning for Gaussian process classification by error reduction. In *35th Conference on Neural Information Processing Systems (NeurIPS)*.
136. Hernandez, A., Balasubramanian, A., Yuan, F., Mason, S.A.M., and Mueller, T. (2019). Fast, accurate, and transferable many-body interatomic potentials by symbolic regression. *npj Comput. Mater.* *5*, 112. <https://doi.org/10.1038/s41524-019-0249-1>.
137. Rubinstein, R., Bruckstein, A.M., and Elad, M. (2010). Dictionaries for sparse representation modeling. *Proc. IEEE* *98*, 1045–1057. <https://doi.org/10.1109/JPROC.2010.2040551>.
138. Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* *35*, 1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>.
139. Li, Y., Yang, M., and Zhang, Z. (2019). A survey of multi-view representation learning. *IEEE Trans. Knowl. Data Eng.* *31*, 1863–1883. <https://doi.org/10.1109/TKDE.2018.2872063>.
140. Long, W., Lu, Z., and Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowl. Base Syst.* *164*, 163–173. <https://doi.org/10.1016/j.knosys.2018.10.034>.

141. James, M.K., and Veeramachaneni, K. (2015). Deep feature synthesis: Towards automating data science endeavors. In 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA) (IEEE), pp. 1–10. <https://doi.org/10.1109/DSAA.2015.7344858>.
142. Kaul, A., Maheshwary, S., and Pudi, V. (2017). Autolearn—automated feature generation and selection. In 2017 IEEE International Conference on Data Mining (ICDM) (IEEE), pp. 217–226.
143. Khurana, U., Turaga, D., Samulowitz, H., and Srinivasan, P. (2016). Cognito: Automated feature engineering for supervised learning. In 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW) (IEEE), pp. 1304–1307.
144. Khurana, U., Samulowitz, H., and Turaga, D. (2018). Feature engineering for predictive modeling using reinforcement learning. Proc of AAAI 2018 32, 3407–3414.
145. Zhang, J., Jianye, H., Fogelman-Soulié, F., and Wang, Z. (2019). Automatic feature engineering by deep reinforcement learning. In Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, pp. 2312–2314.
146. Ghiringhelli, L.M., Vybiral, J., Levchenko, S.V., Draxl, C., and Scheffler, M. (2015). Big data of materials science: critical role of the descriptor. Phys. Rev. Lett. 114, 105503. <https://doi.org/10.1103/PhysRevLett.114.105503>.
147. Ghiringhelli, L.M., Vybiral, J., Ahmetcik, E., Ouyang, R., Levchenko, S.V., Draxl, C., and Scheffler, M. (2017). Learning physical descriptors for materials science by compressed sensing. New J. Phys. 19, 023017. <https://doi.org/10.1088/1367-2630/aa57bf>.
148. Ouyang, R., Curtarolo, S., Ahmetcik, E., Scheffler, M., and Ghiringhelli, L.M. (2018). SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. Phys. Rev. Mater. 2, 083802. <https://doi.org/10.1103/PhysRevMaterials.2.083802>.
149. Fan, J., and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. J. Roy. Stat. Soc. B 70, 849–911. <https://doi.org/10.1111/j.1467-9868.2008.00674.x>.
150. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc. B 58, 267–288. <https://www.jstor.org/stable/2346178>.
151. Xiang, Z., Fan, M., Guillermo Vazquez Tovar, Trehern, W., Yoon, B.-J., Qian, X., Arróyave, R., and Qian, X. (2021). Physics-constrained automatic feature engineering for predictive modeling in materials science. In the 35th AAAI Conference on Artificial Intelligence (AAAI 2021)).
152. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. Nature 518, 529–533. <https://doi.org/10.1038/nature14236>.
153. Podryabinkin, E.V., and Shapeev, A.V. (2017). Active learning of linearly parametrized interatomic potentials. Comput. Mater. Sci. 140, 171–180. <https://doi.org/10.1016/j.commatsci.2017.08.031>.
154. Zhang, L., Lin, D.-Y., Wang, H., Car, R., and Weinan, E. (Feb 2019). Active learning of uniformly accurate interatomic potentials for materials simulation. Phys. Rev. Mater. 3, 023804. <https://doi.org/10.1103/PhysRevMaterials.3.023804>.
155. Vandermause, J., Torrisi, S.B., Batzner, S., Xie, Y., Sun, L., Kolpak, A.M., and Kozinsky, B. (2020). On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. npj Comput. Mater. 6, 20. <https://doi.org/10.1038/s41524-020-0283-z>.
156. Wilson, N., Wilhelm, D., Qian, X., Arróyave, R., and Qian, X. (2022). Batch active learning for accelerating the development of interatomic potentials. Comput. Mater. Sci. 208, 111330. <https://doi.org/10.1016/j.commatsci.2022.111330>.
157. Chen, C., Deng, Z., Tran, R., Tang, H., Chu, I.-H., and Ong, S.P. (2017). Accurate force field for molybdenum by machine learning large materials data. Phys. Rev. Mater. 1, 043603. <https://doi.org/10.1103/PhysRevMaterials.1.043603>.