IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 7, 2018, accepted November 29, 2018, date of publication December 13, 2018,
date of current version January 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2886576

Experimental Design via Generalized Mean Objective Cost of Uncertainty

SHAHIN BOLUKI, (Student Member, IEEE), XIAONING QIAN, (Senior Member, IEEE), AND EDWARD R. DOUGHERTY, (Fellow, IEEE)
Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA

Corresponding author: Shahin Boluki (s.boluki@tamu.edu)

This work was supported in part by the NSF through the projects *CAREER: Knowledge-driven Analytics, Model Uncertainty, and Experiment Design* under Grant NSF-CCF-1553281; *DMREF: Accelerating the Development of Phase-Transforming Heterogeneous Materials: Application to High Temperature Shape Memory Alloys* under Grant NSF-CMMI-1534534; and Elements: Software: Autonomous, Robust, and Optimal In-Silico Experimental Design Platform for Accelerating Innovations in Materials Discovery under Grant NSF-OAC-1835690.

ABSTRACT The mean objective cost of uncertainty (MOCU) quantifies the performance cost of using an operator that is optimal across an uncertainty class of systems as opposed to using an operator that is optimal for a particular system. MOCU-based experimental design selects an experiment to maximally reduce MOCU, thereby gaining the greatest reduction of uncertainty impacting the operational objective. The original formulation applies to finding optimal system operators, where optimality is with respect to a cost function, such as mean-square error; and the prior distribution governing the uncertainty class relates directly to the underlying physical system. Here, we provide a generalized MOCU and the corresponding experimental design. We, then, demonstrate how this new formulation includes as special cases MOCU-based experimental design methods developed for materials science and genomic networks, when there is experimental error. Most importantly, we show that the classical knowledge gradient and efficient global optimization procedures are specific implementations of MOCU-based experimental design under their modeling assumptions.

INDEX TERMS Bayesian methods, experimental design, complex systems, uncertainty.

I. INTRODUCTION

Optimal experimental design is critical for autonomously learning physical models. This is because experiments can be costly and time-consuming, such as the ones in biology and materials design. It is desirable to help design the experiments that reduce the uncertainty pertaining to the ultimate operational objective, be it control, filtering, classification, drug design, materials design, or some other operational goal.

From the Bayesian perspective, Lindley's paradigm posits a general framework for Bayesian experimental design [1]. Two standard procedures within this paradigm are the Knowledge Gradient (KG) [2], [3] and Efficient Global Optimization (EGO) [4], which provide (one-step) optimal experimental design under Gaussian belief and observation noise (KG only) for an offline ranking and selection problem. A more recently introduced method is based on the mean objective cost of uncertainty (MOCU), which quantifies the performance cost of using an operator that is optimal across an uncertainty class of systems as opposed to an operator that is optimal for a particular system within the class [5]. MOCU-based experimental design selects an experiment that maximally reduces MOCU, thereby optimally reducing uncertainty with respect to the operational objective [6]. For instance, if one wishes to design a Wiener filter when the relevant power spectra are not fully known but belong to an uncertainty class of power spectra, then the problem is to design a linear filter that is optimal relative to both mean-square error (MSE) and the probability mass over the uncertainty class. An optimal experiment maximally reduces MOCU relative to uncertainty in the relevant power spectra [7].

Here we consider a generalized formulation of MOCU that is neither necessarily dependent on the particularities of the underlying system model nor necessarily involves a design problem focused on operators. We show that the corresponding generalized experimental design encompasses existing formulations in signal processing, genomics, and materials discovery, and that it fits within Lindley's paradigm for Bayesian experimental design. Within this generalized framework we examine the connection and differences of MOCU-based formulations with other Bayesian experimental design

methods. In particular, we show that the generalized MOCU generates the same policies as Knowledge Gradient and Efficient Global Optimization under their modeling assumptions. Not only does the generalized MOCU framework unify disparate problems, it opens up Bayesian experimental design for reduction of objective related uncertainty, as demonstrated by materials discovery using Ginzburg-Landau theory.

II. GENERALIZED MOCU

We first formulate experimental design in terms of generalized MOCU and then give the standard method by simply defining the terms in the generalized model appropriately. In this paper, the lower case Greek letters denote random variables or distribution functions and capital Greek letters denote the corresponding domain space. We assume a probability space Θ with probability measure π, a set Ψ, and a function $C : \Theta \times \Psi \to [0, \infty)$, where Θ, π, Ψ, and C are called the *uncertainty class*, *prior distribution*, *action space*, and *cost function*, respectively. Elements of Θ and Ψ are called *uncertainty parameters* and *actions*, respectively. For any $\theta \in \Theta$, an *optimal action* is an element $\psi_\theta \in \Psi$ such that $C(\theta, \psi_\theta) \leq C(\theta, \psi)$ for any $\psi \in \Psi$. An *intrinsically Bayesian robust (IBR) action* is an element $\psi^{\Theta}_{\mathrm{IBR}} \in \Psi$ such that $\mathrm{E}_\theta[C(\theta, \psi^{\Theta}_{\mathrm{IBR}})] \leq \mathrm{E}_\theta[C(\theta, \psi)]$ for any $\psi \in \Psi$.

Whereas $\psi^{\Theta}_{\mathrm{IBR}}$ is optimal over Θ, for $\theta \in \Theta$, ψ_θ is optimal relative to θ. The *objective cost of uncertainty* is defined by the performance loss of applying $\psi^{\Theta}_{\mathrm{IBR}}$ instead of ψ_θ on θ:

$$\mathrm{U}_\Psi(\Theta) = C(\theta, \psi^{\Theta}_{\mathrm{IBR}}) - C(\theta, \psi_\theta). \tag{1}$$

Averaging this cost over Θ gives the *mean objective cost of uncertainty (MOCU)*:

$$\mathrm{M}_\Psi(\Theta) = \mathrm{E}_\theta[C(\theta, \psi^{\Theta}_{\mathrm{IBR}}) - C(\theta, \psi_\theta)]. \tag{2}$$

The action space is arbitrary so long as the cost function is defined on $\Theta \times \Psi$. It can be a set of filters defined on a random process with C being mean-square error or a set of drug interventions with C quantifying patient condition.

As noted in [5], MOCU can be viewed as the minimum expected value of a Bayesian loss function that maps an operator to its differential cost (for using the given operator instead of an optimal operator). The minimum expectation is attained by an optimal robust operator that minimizes the average differential cost. In decision theory, this differential cost is called the *regret*, which is defined as the difference between the maximum payoff (for making an optimal decision) and the actual payoff (for the decision that has been made). From this perspective, MOCU can be viewed as the minimum expected regret for using a robust operator.

Suppose there is a set Ξ, called the *experiment space*, whose elements, ξ, called *experiments*, are jointly distributed with the uncertainty parameters θ. To avoid overly complex notation, we denote both an experiment and its outcome by ξ. More specifically, when used in conditioning the probability spaces and distributions, ξ represents an outcome, and when in a minimization/maximization argument, it corresponds to an experiment. Given $\xi \in \Xi$, the conditional distribution $\pi(\theta|\xi)$ is the *posterior distribution* relative to ξ and $\Theta|\xi$ denotes the corresponding probability space, called the *conditional uncertainty class*. Relative to $\Theta|\xi$, we define IBR actions $\psi^{\Theta|\xi}_{\mathrm{IBR}}$ and the conditional (remaining) MOCU,

$$\mathrm{M}_\Psi(\Theta|\xi) = \mathrm{E}_{\theta|\xi}[C(\theta, \psi^{\Theta|\xi}_{\mathrm{IBR}}) - C(\theta, \psi_\theta)], \tag{3}$$

where the expectation is with respect to $\pi(\theta|\xi)$. Taking the expectation over ξ gives the expected remaining MOCU,

$$\begin{aligned}\mathrm{D}_\Psi(\Theta, \xi) &= \mathrm{E}_\xi[\mathrm{M}_\Psi(\Theta|\xi)] \\ &= \mathrm{E}_\xi[\mathrm{E}_{\theta|\xi}[C(\theta, \psi^{\Theta|\xi}_{\mathrm{IBR}}) - C(\theta, \psi_\theta)]],\end{aligned} \tag{4}$$

which is called the *experimental design value*. An optimal experiment $\xi^* \in \Xi$ minimizes $\mathrm{D}_\Psi(\Theta, \xi)$, i.e.,

$$\xi^* = \operatorname*{argmin}_{\xi \in \Xi} \mathrm{D}_\Psi(\Theta, \xi). \tag{5}$$

ξ^* also minimizes the difference between the expected remaining MOCU and the current MOCU (maximizes the absolute difference):

$$\begin{aligned}\xi^* &= \operatorname*{argmin}_{\xi \in \Xi} \mathrm{D}_\Psi(\Theta, \xi) - \mathrm{M}_\Psi(\Theta) \\ &= \operatorname*{argmin}_{\xi \in \Xi} \mathrm{E}_\xi[\mathrm{E}_{\theta|\xi}[C(\theta, \psi^{\Theta|\xi}_{\mathrm{IBR}}) - C(\theta, \psi_\theta)]] \\ &\qquad - \mathrm{E}_\theta[C(\theta, \psi^{\Theta}_{\mathrm{IBR}}) - C(\theta, \psi_\theta)] \\ &= \operatorname*{argmin}_{\xi \in \Xi} \mathrm{E}_\xi[\mathrm{E}_{\theta|\xi}[C(\theta, \psi^{\Theta|\xi}_{\mathrm{IBR}})]] - \mathrm{E}_\theta[C(\theta, \psi^{\Theta}_{\mathrm{IBR}})].\end{aligned} \tag{6}$$

In the standard formulation, MOCU depends on a class of operators applied to a parameterized physical model in which θ is a random vector whose distribution depends on a physical characterization of the uncertainty. For instance, in a gene regulatory network, uncertainty arises regarding regulations and experimental design decides which unknown regulations should be determined via experiments so as to minimize the cost of uncertainty relative to the objective of minimizing the long-run likelihood of the cell being in a cancerous state [5], [6], [8]. Θ is an uncertainty class of system models parameterized by a vector θ governed by a probability distribution $\pi(\theta)$ and Ψ is a class of operators on the models whose performances are measured by C. For each operator ψ, $C(\theta, \psi)$ is the cost of applying ψ on model $\theta \in \Theta$. Initially proposed for optimal intervention in Markovian regulatory networks [5] and optimal robust classification [9], IBR operators have been designed for linear and morphological filters [10] and Kalman filters [11].

It is often the case in standard optimal operator design that an optimal operator can be formulated in terms of characteristics of the underlying system – for instance, a Wiener filter is expressed in terms of power spectra and a Kalman filter makes use of the Kalman gain matrix. When this is the case, under appropriate probabilistic assumptions it is often possible to express an IBR operator in precisely the same form, with the original characteristics replaced by *effective characteristics*, these being averages of the model-specific characteristics over the uncertainty class. The IBR Wiener filter takes the same form as the Wiener filter, with the

effective power spectra replacing the model-specific power spectra. The IBR Kalman filter is generated by a similar set of recursive equations, with the Kalman gain matrix replaced by the effective Kalman gain matrix (and other characteristics replaced by their effective counterparts).

As originally formulated [6], experimental design involves k experiments $T_1, \ldots, T_k$, where experiment T_i exactly determines the uncertain parameter θ_i in $\theta = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta$. The *conditional uncertainty vector* $\theta|\theta_i$ is composed of all uncertain parameters other than θ_i, with θ_i now determined by T_i. $\Theta|\theta_i$ is the *reduced uncertainty class* given θ_i. The IBR operator for $\Theta|\theta_i$, the remaining MOCU given θ_i, and the experimental design value take the forms $\psi_{\mathrm{IBR}}^{\Theta|\theta_i}$, $\mathrm{M}_\Psi(\Theta|\theta_i)$, and $\mathrm{D}(\theta_i) = \mathrm{E}_{\theta_i}[\mathrm{M}_\Psi(\Theta|\theta_i)]$, respectively. The *optimal experiment* T_{i^*} is specified by $i^* = \arg\min_{i=1,\ldots,k} \mathrm{D}(\theta_i)$.

Returning to the generalized MOCU formulation, there is wide flexibility in experimental design, depending on the assumptions regarding the uncertainty class, action space, and experiment space, leading to many existing Bayesian experimental design formulations. Bayesian experimental design has a long history, in particular, utilizing the expected gain in Shannon information [12]–[15]. In 1972, Lindley [1] proposed a general decision theoretic approach incorporating a two-part decision involving the selection of an experiment followed by a terminal decision. Supposing λ is a design selected from a family Λ and $\mathbf{X}$ is a data vector, and leaving out the terminal decision, an optimal experiment is given by

$$\lambda^* = \arg\max_{\lambda\in\Lambda} \mathrm{E}_{\mathbf{X}}[\mathrm{E}_\Theta[U(\theta, \mathbf{X}, \lambda)|\mathbf{X}, \lambda]\,|\lambda], \tag{7}$$

where U is a utility function (see [16] for the full decision-theoretic optimization).

With generalized MOCU, recalling that ξ represents both an experiment and its outcome, each experiment ξ corresponds to a data vector $\mathbf{X}|\xi$ and the expected remaining MOCU is

$$\begin{aligned}&\mathrm{E}_\xi[\mathrm{M}_\Psi(\Theta|\mathbf{X}, \xi)]\\ &\quad= \mathrm{E}_{\mathbf{X}|\xi}[\mathrm{E}_\Theta[C_{\theta|(\mathbf{X}|\xi)}(\psi_{\mathrm{IBR}}^{\Theta|(\mathbf{X}|\xi)}) - C_{\theta|(\mathbf{X}|\xi)}(\psi_{\theta|(\mathbf{X}|\xi)})]]\\ &\quad= \mathrm{E}_{\mathbf{X}|\xi}[\mathrm{E}_\Theta[\mathrm{U}_\Psi(\theta, \mathbf{X}, \xi; \Theta)]].\end{aligned} \tag{8}$$

From (8), the optimization of (5) can be expressed in the same form as (7), with ξ in place of λ and utility function $-\mathrm{U}_\Psi(\theta, \mathbf{X}, \xi; \Theta)$.

Hence, in descending order of generality, we have Lindley's procedure, generalized MOCU, and MOCU. The salient point regarding the latter is that the uncertainty is on the underlying random process, meaning the science, and its aim is to design a better operator on the underlying process.

With sequential experiments, the action space and experiment space can be time dependent, i.e., they can be different for each time step. Hereafter, in sequential experiment setups, the action space and experiment space at time step t, and the optimal experiment selected at t to be performed at the next time step are denoted by Ψ^t, Ξ^t, and $\xi^{*,t}$, respectively. Let $\pi(\theta|\xi^{:t})$ be the posterior distribution given the selected experiments' outcomes from the first time step through t, and $\Theta|\xi^{:t}$ denote the corresponding conditional uncertainty class. When experiments are selected sequentially and there is no fixed limited budget of experiments but instead the experimenter wants to stop the iterative procedure when only negligible knowledge regarding the objective can be gained from additional experiments, the form in (6) is useful because it incorporates the difference between the expected remaining MOCU and the current MOCU. The iterative procedure may be stopped if it falls below a threshold. While this procedure is optimal at each step, it is not optimal given a fixed number of experiments to be performed. This latter kind of finite-horizon optimal design using MOCU is treated in [17] using dynamic programming.

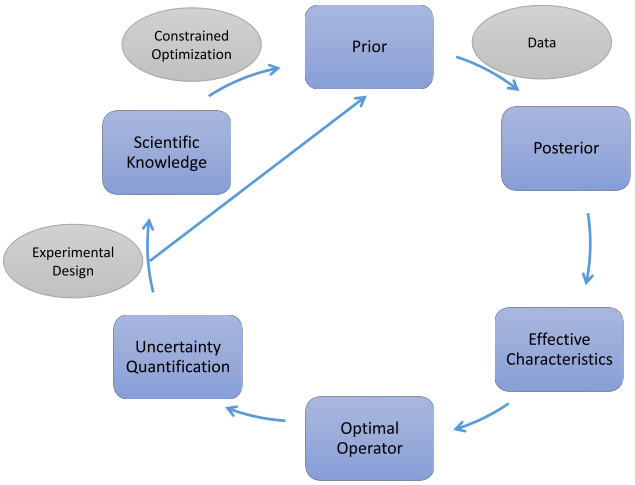


FIGURE 1. A design loop for designing optimal operators under uncertainty.

Sequential experiments can be understood in terms of a design loop for designing optimal operators under uncertainty. Referring to Fig. 1, and considering the standard MOCU formulation, the base of the design loop is construction of the prior. This can be done in numerous ways; however, a very general procedure can be used to derive the *Maximal Knowledge-driven Information Prior (MKDIP)* that minimizes an information-based cost function subject to constraints characterizing our prior knowledge [18], [19]. The prior can then be updated to a posterior using data. Assuming the existence of effective characteristics, following posterior update, these are computed and an IBR operator determined. Uncertainty is quantified by MOCU and, if desired, optimal experiments performed to produce new knowledge that can be used to supplement the original knowledge or directly condition the original prior, in either case producing a new prior to re-institute the design process. The design loop involves two optimizations, and therefore two cost functions, one for prior construction and one for operator design.

We next show how generalized MOCU can be applied in real-world applications, specifically in life and materials science research. We also show that the generalized MOCU experimental design includes other existing objective-based experimental-design formulations.

III. GENERALIZED MOCU IN APPLICATIONS

A. DYNAMICAL GENETIC NETWORKS

In [8], optimal objective-based experimental design is derived for networks with multiple dynamic trajectories, modeling in [8] being based on [20]. Briefly, the network's nodes and their corresponding values represent entities, proteins/chemicals or genes, and their corresponding concentration levels or expression levels, respectively. The values are assumed to be nonnegative integers. Each edge represents an interaction with its input, regulation, and output nodes. Each interaction can dynamically happen if all of its input and activator nodes are nonzero and its inhibitor nodes are zero. All interactions are known. When the network is in state x, it can have one or more possible interactions based on the node values, where if any takes place, the network transitions to a next state. When multiple interactions exist, if knowledge of the relative priorities of these competing interactions exist, we can completely determine the state trajectory of the network from an initial state x_0.

The assumption is that these relative priorities are not known but can be measured one at a time with experimental error. If the network has R of these competing interactions, i.e., interactions that can dynamically happen at the same time, then the uncertainty class consists of a set of R Boolean random variables, $\Theta = \{0, 1\}^R$, and $\theta = (\theta_1, \ldots, \theta_R)$, where $\theta_i \in \{0, 1\}_{i=1,\ldots,R}$. The i^{th} experiment can determine the value of θ_i with an experimental error having probability δ_i. Specifically, if θ_i is selected to be measured, with probability $1-\delta_i$ the outcome of the experiment is θ_i, and with probability δ_i is $1 - \theta_i$. Here, $\Xi = \{\xi_1, \ldots, \xi_R\}$, each experiment ξ_i corresponds to measuring θ_i, and

$$\xi_i|\theta_i = \begin{cases} \theta_i & \text{with probability} \quad 1-\delta_i, \\ 1-\theta_i & \text{with probability} \quad \delta_i. \end{cases} \tag{9}$$

An action blocks an interaction from happening, so the action space is $\Psi = \{\psi_1, \ldots, \psi_A\}$, where A is the number of interactions that can be blocked. Each action changes the dynamic trajectory of the network. If the set of possible state trajectories is denoted by $S^{\Theta}_{\psi_i}$ when the i^{th} action (ψ_i) is taken, then the probability of each trajectory $s \in S^{\Theta}_{\psi_i}$ is

$$P_{S^{\Theta}_{\psi_i}}(s) = \mathrm{E}_{x_0}\big[\mathrm{E}_{\theta}[\mathbb{1}_{s_{x_0,\theta}(\psi_i)=s}]\big], \tag{10}$$

where $\mathbb{1}_w$ is the indicator function ($\mathbb{1}_w = 1$ if w is true and is 0 otherwise), and $s_{x_0,\theta}(\psi_i)$ is the deterministic trajectory for a fixed initial state x_0 and θ, when action ψ_i is taken. Here, $S^{\Theta}_{\psi_i} = \cup_{x_0 \in X_0} \cup_{\theta \in \Theta} s_{x_0,\theta}(\psi_i)$, where X_0 denotes the set of all possible initial states. For each trajectory s, the dynamic performance cost $\varepsilon(s)$ is defined as the distance (in terms of any appropriate norm) of the steady-state vector corresponding to that trajectory (x^s_f) from a desired distribution v, i.e. $\varepsilon(s) = ||x^s_f - v||$. Thus, the cost function for a fixed θ and action ψ is the expected cost over the possible trajectories, $C(\theta, \psi) = \mathrm{E}_{S^{\Theta}_{\psi}}[\varepsilon(s)]$.

The IBR action for this problem is

$$\psi^{\Theta}_{\text{IBR}} = \arg \min_{\psi \in \{\psi_1, \ldots, \psi_A\}} \mathrm{E}_{\theta}[C(\theta, \psi)]. \tag{11}$$

According to (4) and (5), the optimal experiment can be derived as

$$\begin{aligned} \xi^* &= \operatorname*{argmin}_{\xi_i \in \Xi} \mathrm{E}_{\xi_i}[\mathrm{E}_{\theta|\xi_i}[C(\theta, \psi^{\Theta|\xi_i}_{\text{IBR}}) - C(\theta, \psi_\theta)]] \\ &= \operatorname*{argmin}_{\xi_i \in \Xi} \mathrm{E}_{\xi_i}[\mathrm{E}_{\theta_i|\xi_i}[\mathrm{E}_{\theta\setminus\theta_i}[C(\theta, \psi^{\Theta|\xi_i}_{\text{IBR}}) - C(\theta, \psi_\theta)]]] \\ &= \operatorname*{argmin}_{\xi_i \in \Xi} \mathrm{E}_{\theta_i}[\mathrm{E}_{\xi_i|\theta_i}[\mathrm{E}_{\theta\setminus\theta_i}[C(\theta, \psi^{\Theta|\xi_i}_{\text{IBR}}) - C(\theta, \psi_\theta)]]] \\ &= \operatorname*{argmin}_{\xi_i \in \Xi} \mathrm{E}_{\theta_i}[\mathrm{E}_{\xi_i|\theta_i}[\mathrm{E}_{\theta\setminus\theta_i}[C(\theta, \psi^{\Theta|\xi_i}_{\text{IBR}})]]], \end{aligned} \tag{12}$$

where "\" denotes set subtraction in the subscripts. The second line holds because only the posterior distribution of θ_i depends on experiment ξ_i; and the last equality follows from the independence of $C(\theta, \psi_\theta)$ from ξ_i. The last line is exactly the policy derived in [8] but there the policy derivation was based on adding the objective-based cost of experimental error to the previous notion of objective cost of uncertainty, whereas here we directly apply the generalized formulation of MOCU as we have formulated in Section II.

B. GUIDING SIMULATIONS IN MATERIALS DISCOVERY

In [21], optimal experimental design based on MOCU is 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 time-dependent Ginzburg-Landau (TDGL) theory. The TDGL model simulates the free energy for a specific dopant with a specified concentration, given the dopant's parameters. The assumption is that there is a set $D = \{d_1, \ldots, d_N\}$ of N potential dopants and each dopant d_i can be characterized by two parameters, its strength h_i and its range of stress disturbance r_i. The concentration of the dopants can be selected from a set $O = \{o_1, \ldots, o_P\}$ of P pre-specified values. The true values of these dopant parameters are unknown; however, there exists a prior distribution over the dopant parameters. In summary, we have $\Theta = H \times R$ and $\theta = [h, r]$, where $h = [h_1, \ldots, h_N]$ and $r = [r_1, \ldots, r_N]$, and H and R represent the sample spaces of h and r, respectively. Thus, $\theta_i = [h_i, r_i]$ fully characterizes dopant d_i.

Since the computational complexity of the TDGL model is enormous, the goal is to find an optimal dopant and concentration to minimize the simulated energy dissipation, with the least number of times running the TDGL model (least number of experiments). Following [21], for this purpose, a surrogate model $g(h, r, o)$ is trained based on fitting some initial data generated from the TDGL model. The surrogate model can approximately predict a dissipation energy for a specified dopant and concentration, and it is used as the cost function throughout the experimental design iterations. The TDGL model acts as the true underlying system, or Nature, and the surrogate model is the model of the true system. The action space is $\Psi = \{\psi_{d_i,o_j}\}_{d_i \in D, o_j \in O}$, where each action ψ_{d_i,o_j} is

using the i^{th} dopant with the j^{th} possible concentration. The cost function is $C(\theta, \psi_{d_i,o_j}) = g(h_i, r_i, o_j)$. The experiment space is $\Xi = \{\xi_{d_i,o_j}\}_{d_i \in D, o_j \in O}$, where ξ_{d_i,o_j} corresponds to obtaining a noisy measurement of the dissipation energy when using the i^{th} dopant with the j^{th} concentration. $\xi_{d_i,o_j} \sim f(\xi_{d_i,o_j}|\theta_i)$, where f is a probability distribution.

In this framework, the IBR action at time step t is

$$\begin{aligned}\psi_{\text{IBR}}^{\Theta|\xi^{:t}} &= \arg\min_{\psi\in\Psi} E_{\theta|\xi^{:t}}\big[C(\theta,\psi)\big] \\ &= \arg\min_{\psi_{d_i,o_j}\in\Psi} E_{\theta|\xi^{:t}}\big[g(h_i,r_i,o_j)\big]. \end{aligned} \tag{13}$$

From (4) and (5), the optimal experiment at time step t (to be performed at $t+1$) is

$$\begin{aligned}\xi^{*,t} &= \operatorname*{argmin}_{\xi\in\Xi} \mathrm{E}_{\xi}[\mathrm{E}_{\theta|\xi}[C(\theta,\psi_{\text{IBR}}^{\Theta|\xi,\xi^{:t}}) - C(\theta,\psi_\theta)]] \\ &= \arg\min_{\xi_{d_i,o_j}\in\Xi} \mathrm{E}_{\xi_{d_i,o_j}}[\mathrm{E}_{\theta|\xi_{d_i,o_j},\xi^{:t}}[C(\theta,\psi_{\text{IBR}}^{\Theta|\xi^{:t+1}})]], \end{aligned} \tag{14}$$

where the second equality is due to the independence of $C(\theta, \psi_\theta)$ from ξ_{d_i,o_j}. The last line of (14) is exactly the policy proposed in [21] for this materials science problem.

In the genomic application, the parameters of the cost function come from an underlying physical system. Another example of this kind is Karhunen-Loève compression, where the parameters characterize the image structure and the cost function measures the difference between the original and compressed images [22]. In the materials application, the surrogate model, instead of the actual physical model, is used for the experimental design.

A third possibility is that we do not possess a physical model and we lack sufficient knowledge to posit a surrogate model relating to our objective. Nevertheless, we can take an ad hoc approach and select a model with known predictive properties. This model can be a kernel-based model, for instance, a Gaussian Process Regression model [23]. More generally, the model can consist of a set of possible parametric families, or be a kernel-based model with different possible feature sets, or even kernel-based models with different choices for the kernel function. In [24] no knowledge is assumed regarding which feature set or model family would be the best. Instead, Bayesian model averaging is used based on different feature sets for a Gaussian process regression model, weighted by their posterior probabilities of being the correct model, where possible feature sets are selected based on domain knowledge. Assuming a single objective, generalized MOCU can be applied to all three scenarios.

IV. CONNECTION OF MOCU-BASED EXPERIMENTAL DESIGN WITH KG AND EGO

Knowledge Gradient (KG) [2], [3], which is used in different fields, from drug discovery to material design [25], [26], was originally introduced as a solution to an offline ranking and selection problem, where the assumption is that there are $A \geq 2$ actions (alternatives) that can be selected, i.e., $\Psi = \{\psi_1, \ldots, \psi_A\}$. Each action has an unknown true reward (sign-flipped cost) and at each time step an experiment provides a noisy observation of the reward of a selected action. There is a limited budget (B) of the number of measurements we can make before the time arrives to decide which action is the best, that being the one having the lowest expected cost (or the highest expected reward).

The assumption is that we have Gaussian prior beliefs over the unknown rewards, either independent Gaussian beliefs over the rewards when the rewards of different actions are uncorrelated, or a joint Gaussian belief when the rewards are correlated. In the independent case, for each action-reward pair $(\psi_i, \theta_{\psi_i})$, $\theta_{\psi_i} \sim N(m_{\psi_i}, \beta_{\psi_i})$. In the correlated case, the vector of rewards, $[\theta_{\psi_1}, \ldots, \theta_{\psi_A}]$, has a multivariate Gaussian distribution $N(m, \Sigma)$ with the mean vector $m = [m_{\psi_1}, \ldots, m_{\psi_A}]$ and covariance matrix Σ, with diagonal entries $[\beta_{\psi_1}, \ldots, \beta_{\psi_A}]$. If the selected action to be applied at t is ψ^t, then the observed noisy reward of ψ^t at that iteration is $\xi^t = \theta_{\psi^t} + \epsilon^t$, where θ_{ψ^t} is unknown and $\epsilon^t \sim N(0, \lambda_{\psi^t})$ is independent of the reward of ψ^t.

Here, the underlying system to learn is the unknown reward function and each possible model is fully described by a reward vector $\theta = [\theta_{\psi_1}, \theta_{\psi_2}, \ldots, \theta_{\psi_A}]$ in the uncertainty class Θ. For the independent case, $\pi(\theta) = \prod_{i=1}^{A} N(m_{\psi_i}, \beta_{\psi_i})$. For the correlated case, $\pi(\theta) = N(m, \Sigma)$. The experiment space is $\Xi = \{\xi_1, \ldots, \xi_A\}$, where experiment ξ_i corresponds to applying ψ_i and getting a noisy observation of its reward θ_{ψ_i}, that is, measuring θ_{ψ_i} with observation noise, where $\xi_i|\theta_{\psi_i} \sim N(\theta_{\psi_i}, \lambda_{\psi_i})$. In the independent case the state of knowledge at each time point t is captured by the posterior values of the means and variances for the rewards after incorporating observations $\xi^{:t}$ as $S^t = [(m^t_\psi, \beta^t_\psi)]_{\psi\in\Psi}$, and in the correlated case by the posterior vector of means and a covariance matrix after observing $\xi^{:t}$ as $S^t = (m^t, \Sigma^t)$, where $m^t = [m^t_{\psi_1}, \ldots, m^t_{\psi_A}]$ and the diagonal of Σ^t is the vector $[\beta^t_{\psi_1}, \ldots, \beta^t_{\psi_A}]$. The probability space $\Theta|\xi^{:t}$ is equal to $\Theta|S^t$ and the cost function is $C(\theta, \psi) = -\theta_\psi$.

For this problem, the IBR action at time step t is

$$\begin{aligned}\psi_{\text{IBR}}^{\Theta|\xi^{:t}} &= \operatorname*{argmin}_{\psi\in\Psi} \mathrm{E}_{\Theta|\xi^{:t}}\big[C(\theta,\psi)\big] = \operatorname*{argmin}_{\psi\in\Psi} \mathrm{E}_{\Theta|\xi^{:t}}\big[-\theta_\psi\big] \\ &= \operatorname*{argmax}_{\psi\in\Psi} \mathrm{E}_{\Theta|\xi^{:t}}\big[\theta_\psi\big] = \operatorname*{argmax}_{\psi\in\Psi} m^t_\psi. \end{aligned} \tag{15}$$

Again, by (4) and (5), the optimal experiment selected at time step t (to be performed at $t+1$) can be derived:

$$\begin{aligned}\xi^{*,t} &= \operatorname*{argmin}_{\xi_i\in\Xi} \mathrm{E}_{\xi_i|\xi^{:t}}[\mathrm{E}_{\theta|\xi_i,\xi^{:t}}[C(\theta,\psi_{\text{IBR}}^{\Theta|\xi^{:t},\xi_i})]] \\ &\qquad - \mathrm{E}_{\theta|\xi^{:t}}[C(\theta,\psi_{\text{IBR}}^{\Theta|\xi^{:t}})] \\ &= \operatorname*{argmin}_{\xi_i\in\Xi} \mathrm{E}_{\xi_i|\xi^{:t}}\Big[\mathrm{E}_{\theta|\xi^{:t+1}}\big[-\theta_{\psi_{\text{IBR}}^{\Theta|\xi^{:t+1}}}\big]\Big] \\ &\qquad - \mathrm{E}_{\theta|\xi^{:t}}\big[-\theta_{\psi_{\text{IBR}}^{\Theta|\xi^{:t}}}\big] \\ &= \operatorname*{argmax}_{\xi_i\in\Xi} \mathrm{E}_{\xi_i|\xi^{:t}}\Big[\mathrm{E}_{\theta|\xi^{:t+1}}\big[\theta_{\psi_{\text{IBR}}^{\Theta|\xi^{:t+1}}}\big]\Big] - \mathrm{E}_{\theta|\xi^{:t}}\big[\theta_{\psi_{\text{IBR}}^{\Theta|\xi^{:t}}}\big] \\ &= \operatorname*{argmax}_{\xi_i\in\Xi} \mathrm{E}_{\xi_i|\xi^{:t}}\Big[\max_{\psi'\in\Psi} m^{t+1}_{\psi'}\Big] - \max_{\psi'\in\Psi} m^t_{\psi'}. \end{aligned} \tag{16}$$

The policy (16) derived by direct application of the generalized MOCU is exactly the same as the original KG policy

in [2], [3], and [27]. As KG is shown to be optimal when the horizon is a single measurement and asymptotically optimal (the number of measurements goes to infinity), the same holds for the MOCU-based policy for this problem.

Efficient Global Optimization (EGO) [4], which is based on expected improvement (EI), is widely used for black-box optimization and experimental design. As shown in [26], KG reduces to EGO when there is no observation noise and choosing the best action at each time step is limited to selecting from the set of actions whose rewards have been previously observed; that is, at each time step if we want to make a final decision as to the best action to be applied, it must be an action whose performance has been previously observed from the first time step up to that time. Thus, MOCU-based learning can also be reduced to EGO under its model assumptions. We will show this directly.

Consider the ranking and selection problem with no noise in the observations, so that $\epsilon^t = 0$ for all t. Each experiment ξ_i corresponds to applying ψ_i and observing the true value of θ_{ψ_i}. Moreover, the choice of the best action at each time step is confined to the set of actions whose rewards have been previously observed. Let Ψ^t denote this set: $\Psi^t = \{\psi^{t'}\}_{t'=1,\ldots,t}$. The IBR action at time t is

$$\psi_{\mathrm{IBR}}^{\Theta|\xi^{:t}} = \arg\min_{\psi\in\Psi^t} \mathrm{E}_{\Theta|\xi^{:t}}[-\theta_\psi] = \arg\max_{\psi\in\Psi^t} \theta_\psi, \qquad (17)$$

where the last equality is due to the fact that the reward of an action whose performance is already observed is known, since there is no observation noise. Let $Z^t = \{\xi^{t'}\}_{t'=1,\ldots,t}$ denote the set of experiments performed up to the current time t, where experiment $\xi^{t'}$ corresponds to $\psi^{t'}$ being applied at t' and its reward being observed, in other words, measurement of $\theta_{\psi^{t'}}$ at t'. Since there is no point in measuring an action's reward more than once, the next experiment is selected from the set of remaining experiments, so that the experiment space at time step t is $\Xi^t = \Xi\backslash Z^t$. From (4), (5), and (17), the optimal experiment selected at t is

$$\begin{aligned}
\xi^{*,t} &= \arg\min_{\xi_i\in\Xi^t} \mathrm{E}_{\xi_i|\xi^{:t}}\Big[\mathrm{E}_{\theta|\xi^{:t+1}}\big[-\theta_{\psi_{\mathrm{IBR}}^{\Theta|\xi^{:t+1}}}\big]\Big] \\
&\qquad - \mathrm{E}_{\theta|\xi^{:t}}\big[-\theta_{\psi_{\mathrm{IBR}}^{\Theta|\xi^{:t}}}\big] \\
&= \arg\max_{\xi_i\in\Xi\backslash Z^t} \mathrm{E}_{\theta_{\psi_i}|\xi^{:t}}\Big[\max\big(\theta_{\psi_i}, \max_{\psi'\in\Psi^t}\theta_{\psi'}\big)\Big] \\
&\qquad - \max_{\psi'\in\Psi^t}\theta_{\psi'} \\
&= \arg\max_{\xi_i\in\Xi\backslash Z^t} \mathrm{E}_{\theta_{\psi_i}|\xi^{:t}}\Big[\max\big(\theta_{\psi_i} - \max_{\psi'\in\Psi^t}\theta_{\psi'}, 0\big)\Big], \qquad (18)
\end{aligned}$$

which is exactly the EGO policy in [4].

There are fundamental differences between the general MOCU formulation and KG (or EGO): (1) with MOCU the experiment space and action space can be different, enabling more flexible experimental design compared to the assumption of the same experiment and action space in KG (or EGO); (2) MOCU considers the uncertainty directly on the underlying physical model, which allows direct incorporation of prior knowledge regarding the underlying system, whereas in KG (or EGO) the uncertainty is considered on the reward function and there is no direct connection between prior assumptions and the underlying physical model.

V. A SIMULATION STUDY TO COMPARE MOCU-BASED EXPERIMENTAL DESIGN AND KG

In this section, we perform a simulation study to illustrate the flexibility of MOCU-based experimental design compared to KG, especially the importance of the flexibility of dissecting the uncertainty class assumptions to better incorporate prior knowledge regarding the underlying model. Here we compare the experimental design performances by MOCU and KG based on a simulated quadratic function example with one input variable as the underlying reward function that we want to maximize: $f(\theta,\psi) = \theta_1\psi^2 + \theta_2\psi + \theta_3$, i.e. $C(\theta,\psi) = -f(\theta,\psi)$. The observation noise is additive Gaussian with the distribution $N(0,\theta_4^2)$. In this simulation model, θ_1, θ_2 , θ_3 and θ_4 are unknown parameters. We take $\Psi = \{\psi_1,\ldots,\psi_{20}\} = \{0.5, 1, 1.5, \ldots, 10\}$ as the set of actions (possible input values ψ). The corresponding experiment for each action is to apply ψ_i so that we can observe the outcome ξ_i (the reward):

$$\xi_i|\theta \sim N(\theta_1\psi_i^2 + \theta_2\psi_i + \theta_3, \theta_4^2). \qquad (19)$$

Note that as shown in Section IV, under model assumptions of KG, MOCU-based experimental design results in the same policy as KG. But here, as opposed to KG that directly models the rewards (and corresponding costs) of actions with Gaussian distributions with (prior) fixed parameter values (either known or estimated), MOCU-based experimental design computes the generalized MOCU by modeling the uncertainty of the reward function by incorporating the uncertainty over the underlying parameters, to guide the experimental design procedure.

For both MOCU-based experimental design and KG, we assume that there is no prior knowledge on the model parameters $\theta = [\theta_1, \theta_2, \theta_3, \theta_4]$. For MOCU, the non-informative prior $\pi(\theta) \propto \theta_4^{-2}$ is used, which updates to a Gaussian-inverse-gamma distribution $\pi^*(\theta)$ when measurements become available as experiments are carried out in sequence. For KG, to model the rewards of actions directly with correlated Gaussian distributions, approximate beliefs are constructed at each experiment since the noise variance is unknown and no joint Gaussian prior distribution exists over the reward values of the actions. For this approximation, following [28] and [26], a Gaussian process regression (GPR) model [23] with a quadratic basis (mean) function and a squared exponential covariance matrix with additive Gaussian observation noise is trained using the measurements performed (experiment outcomes observed) up to that time step (by maximizing the marginal log-likelihood of the observations).

In our simulation, θ_1 is drawn from $U(-5,2)$ ($U(a,b)$ denotes the uniform distribution over the interval (a,b)); θ_2 is set to $-2\theta_1 r$, where r is drawn from $U(-2.5,13)$; θ_3 is sampled from $U(-5,5)$; and θ_4 is set to $\sigma(f)\times w$, where

$w \sim U(0.075, 0.7)$ and $\sigma(f)$ denotes the true standard deviation of the reward values of actions based on the given model parameters. Each simulation starts with four randomly selected actions, for which noisy observations of their rewards are simulated as initial training data to both MOCU-based experimental design and KG. The sequential experimental design procedures based on MOCU and KG are both continued for five iterations. For KG at each time step t, the (posterior) vector of means (m^t), the covariance matrix (Σ^t), and the noise variance are estimated by training a GPR model on the available measurements, and the next experiment is selected by (16). For MOCU-based experimental design at each time step t, the (posterior) Gaussian-inverse-gamma distribution after incorporating the available measurements is used in (6) to optimally select the next experiment.

To compare the performances, we check the average opportunity cost metric, defined as the difference between the true maximum of the reward among all the actions and the true reward of the action selected as the best one based on the two experimental design strategies. Note that this best action might be different from the next suggested experiment by each policy. The best action at each time step is the one that would be selected to be applied if the iterative experiments are stopped at that time. In other words, each experimental design policy suggests the next experiment, and after observing the outcome and based on its updated beliefs selects the best action (that would be applied if the iterative experiments were to stop) and the next experiment to be performed (if experimental budget is not exhausted). When following the MOCU-based policy, the next suggested experiment is the minimizer of the expected remaining MOCU, but the best action at each time step is the IBR action that maximizes (minimizes) the expectation of the reward (cost) with respect to the (posterior) Gaussian-inverse-gamma distribution of uncertain parameters based on the latest belief at that time step. When following the KG policy, the best action at each time step is the one that maximizes the (posterior) GPR mean value at that time step which might be different from the suggested next experiment by KG.

Fig. 2 illustrates the average opportunity cost for MOCU-based experimental design and KG over 1,000 simulation runs. As can be seen from the figure, as soon as the experimental design iterations begin MOCU-based policy consistently has the lower average opportunity cost compared to KG. This confirms that directly incorporating the model uncertainty (the uncertainty of model parameters in this simulation study as we assume that we have the model functional form) in the generalized MOCU framework results in a better experimental design policy. Note that at iteration 0 no experiment selection by any of the methods is performed, and only four randomly selected experiment outcomes are available. Since the flat (non-informative) prior is assumed for the parameters in the MOCU-based framework, the IBR action selection as the best action can be very conservative before beginning the experimental design procedure. The maximizer

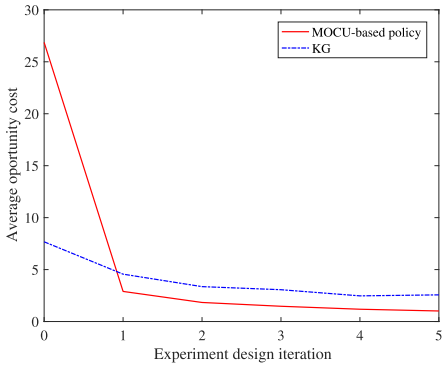


FIGURE 2. Average opportunity cost of MOCU-based policy compared with KG policy.

of the direct approximation of the reward function by GPR at iteration 0 is better than the IBR action for this simple simulation model. But as soon as the first experiment is selected by the policies, MOCU-based policy greatly reduces the uncertainty pertaining to the objective very sharply with the observed measurements and performs consistently better than KG.

VI. CONCLUSIONS

This paper presents a generalized MOCU framework, leading to the MOCU-based experimental design pertaining to the maximum uncertainty reduction of differential cost with respect to the actual operational objectives. The proposed framework fits into Lindley's utility paradigm [1] in classical Bayesian experimental design and is more flexible for the development of corresponding experimental design strategies for different real-world applications compared to the existing KG and EGO methods with their corresponding model assumptions. As we have shown in the simulation study (Section V) and in the recent applications to life and materials science (Sections III-A and III-B), our generalized MOCU framework, with the benefits from flexible dissection of the uncertainty class, action (operator) space, experiment space, and utility function depending on operational objectives, can lead to better objective-based uncertainty quantification and thereafter better experimental design to achieve desired objectives with smaller operational cost.

REFERENCES

[1] D. V. Lindley, *Bayesian Statistics: A Review*. Philadelphia, PA, USA: SIAM, 1972.

[2] P. I. Frazier, W. B. Powell, and S. Dayanik, "A knowledge-gradient policy for sequential information collection," *SIAM J. Control Optim.*, vol. 47, no. 5, pp. 2410–2439, 2008.

[3] P. Frazier, W. Powell, and S. Dayanik, "The knowledge-gradient policy for correlated normal beliefs," *INFORMS J. Comput.*, vol. 21, no. 4, pp. 599–613, 2009.

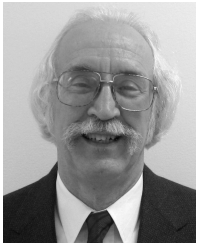
[4] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, no. 4, pp. 455–492, 1998.

[5] B.-J. Yoon, X. Qian, and E. R. Dougherty, "Quantifying the objective cost of uncertainty in complex dynamical systems," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2256–2266, May 2013.

[6] R. Dehghannasiri, B. J. Yoon, and E. R. Dougherty, "Optimal experimental design for gene regulatory networks in the presence of uncertainty," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 12, no. 4, pp. 938–950, Jul. 2015.
[7] R. Dehghannasiri, X. Qian, and E. R. Dougherty, "Optimal experimental design in the context of canonical expansions," *IET Signal Process.*, vol. 11, pp. 942–951, Oct. 2017.
[8] D. N. Mohsenizadeh, R. Dehghannasiri, and E. R. Dougherty, "Optimal objective-based experimental design for uncertain dynamical gene networks with experimental error," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 1, pp. 218–230, Jan./Feb. 2018, doi: 10.1109/TCBB.2016.2602873.
[9] L. A. Dalton and E. R. Dougherty, "Optimal classifiers with minimum expected error within a Bayesian framework—Part I: Discrete and Gaussian models," *Pattern Recognit.*, vol. 46, no. 5, pp. 1301–1314, 2013.
[10] L. A. Dalton and E. R. Dougherty, "Intrinsically optimal Bayesian robust filtering," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 657–670, Feb. 2014.
[11] R. Dehghannasiri, M. S. Esfahani, and E. R. Dougherty, "Intrinsically Bayesian robust Kalman filter: An innovation process approach," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2531–2546, May 2017.
[12] M. Stone, "Application of a measure of information to the design and comparison of regression experiments," *Ann. Math. Statist.*, vol. 30, no. 1, pp. 55–70, 1959.
[13] M. H. DeGroot, "Uncertainty, information, and sequential experiments," *Ann. Math. Statist.*, vol. 33, no. 2, pp. 404–419, 1962.
[14] M. H. DeGroot, *Concepts of Information Based on Utility*. Dordrecht, The Netherlands: Springer, 1986, pp. 265–275.
[15] J. M. Bernardo, "Expected information as expected utility," *Ann. Statist.*, vol. 7, no. 3, pp. 686–690, 1979.
[16] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Stat. Sci.*, vol. 10, no. 3, pp. 273–304, 1995.
[17] M. Imani, R. Dehghannasiri, U. M. Braga-Neto, E. R. Dougherty, "Sequential experimental design for optimal structural intervention in gene regulatory networks based on the mean objective cost of uncertainty," *Cancer Informat.*, vol. 17, p. 1176935118790247, 2018, doi: 10.1177/1176935118790247.
[18] S. Boluki, M. S. Esfahani, X. Qian, and E. R. Dougherty, "Constructing pathway-based priors within a Gaussian mixture model for Bayesian regression and classification," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, to be published, doi: 10.1109/TCBB.2017.2778715.
[19] S. Boluki, M. S. Esfahani, X. Qian, and E. R Dougherty, "Incorporating biological prior knowledge for Bayesian learning via maximal knowledge-driven information priors," *BMC Bioinf.*, vol. 18, no. 14, p. 552, Dec. 2017.
[20] D. N. Mohsenizadeh, J. Hua, M. Bittner, and E. R. Dougherty, "Dynamical modeling of uncertain interaction-based genomic networks," *BMC Bioinf.*, vol. 16, no. 13, p. S3, 2015.
[21] R. Dehghannasiri *et al.*, "Optimal experimental design for materials discovery," *Comput. Mater. Sci.*, vol. 129, pp. 311–322, Mar. 2017.
[22] R. Dehghannasiri, X. Qian, and E. R. Dougherty, "Intrinsically Bayesian robust karhunen-Loève compression," *Signal Process.*, vol. 144, pp. 311–322, Mar. 2018.
[23] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning* (Adaptive Computation and Machine Learning series). Cambridge, MA, USA: MIT Press, 2006.
[24] A. Talapatra, S. Boluki, T. Duong, X. Qian, E. Dougherty, and R. Arróyave, "Autonomous efficient experiment design for materials discovery with Bayesian model averaging," *Phys. Rev. Mater.*, vol. 2, p. 113803, Nov. 2018.
[25] S. Chen, K.-R. G. Reyes, M. K. Gupta, M. C. McAlpine, and W. B. Powell, "Optimal learning in experimental design using the knowledge gradient policy with application to characterizing nanoemulsion stability," *SIAM/ASA J. Uncertainty Quantification*, vol. 3, no. 1, pp. 320–345, 2015.
[26] P. I. Frazier and J. Wang, "Bayesian optimization for materials design," *Information Science for Materials Discovery and Design* (Springer Series in Materials Science), vol. 225. 2016, pp. 45–57.
[27] Y. Wang, K. G. Reyes, K. A. Brown, C. A. Mirkin, and W. B. Powell, "Nested-batch-mode learning and stochastic optimization with an application to sequential multistage testing in materials science," *SIAM J. Sci. Comput.*, vol. 37, no. 3, pp. B361–B381, 2015.
[28] W. Scott, P. Frazier, and W. Powell, "The correlated knowledge gradient for simulation optimization of continuous parameters using Gaussian process regression," *SIAM J. Optim.*, vol. 21, no. 3, pp. 996–1026, 2011.

SHAHIN BOLUKI received the B.Sc. degree (Hons.) in electrical engineering from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2014. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. His current research interests include Bayesian learning, uncertainty quantification in complex systems, and statistical machine learning. He was named one of the three winners of the Young Scientist Excellence Award at the 14th Annual MCBIOS Conference, in 2017.

XIAONING QIAN (S'01–M'07–SM'17) received the Ph.D. degree in electrical engineering from Yale University, New Haven, CT, USA. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. He is also with the TEES-AgriLife Center for Bioinformatics and Genomic Systems Engineering and the Center for Translational Environmental Health Research, Texas A&M University. His research interests include machine learning and Bayesian computation and their applications in computational network biology, genomic signal processing, and biomedical signal and image analysis. He was a recipient of the National Science Foundation CAREER Award, the Texas A&M Engineering Experiment Station Faculty Fellow, and the Montague-Center for Teaching Excellence Scholar at Texas A&M University. His recent work on computational network biology has received the Best Paper Award at the 11th Asia Pacific Bioinformatics Conference, in 2013, and the Best Paper Award in the International Conference on Intelligent Biology and Medicine, in 2016.



EDWARD R. DOUGHERTY (M'05–SM'09–F'12) received the M.Sc. degree in computer science from the Stevens Institute of Technology, Hoboken, NJ, USA, the Ph.D. degree in mathematics from Rutgers University, New Brunswick, NJ, USA, and the Doctor Honoris Causa from the Tampere University of Technology, Finland. He is currently a Distinguished Professor with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, where he holds the Robert M. Kennedy '26 Chair in electrical engineering and is the Scientific Director of the Center for Bioinformatics and Genomic Systems Engineering. He is the author of 19 books, the editor of five others, and the author of over 350 journal papers. He is an SPIE Fellow. He was a recipient of the SPIE President's Award. At Texas A&M University, he was a recipient of the Association of Former Students Distinguished Achievement Award in Research, and has been named a Fellow of the Texas Engineering Experiment Station and a Halliburton Professor of the Dwight Look College of Engineering. He has served as an Editor for the SPIE/IS&T *Journal of Electronic Imaging*.