energy&fuels

pubs.acs.org/EF

Article

Fast QoI-Oriented Bayesian Experimental Design with Unified Neural Response Surfaces for Kinetic Uncertainty Reduction

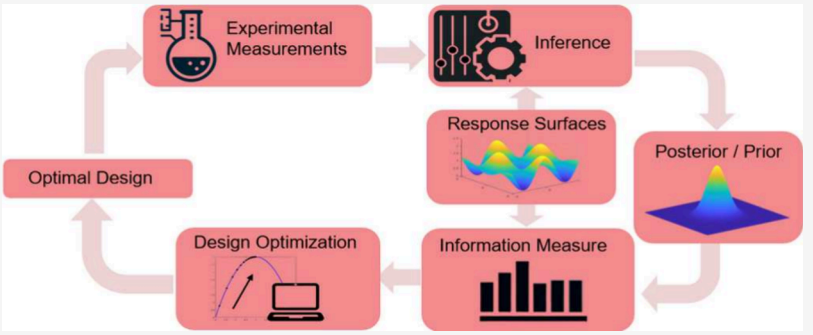
Published as part of Energy & Fuels virtual special issue "2024 Energy and Fuels Rising Stars".

Huaibo Chen, Qiaofeng Li, and Sili Deng*

Cite This: *Energy Fuels* 2024, 38, 15630−15641

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

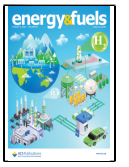


ABSTRACT: In the realm of combustion and reacting flow modeling, the calibration of the kinetic model parameters often relies on experimental data. However, not all data obtained under different experimental conditions (pressure, temperature, equivalence ratio, etc.) hold equal weight or feasibility for effective model calibration. Consequently, experimental design emerges as an important topic in combustion kinetics, aiming at identifying the most informative conditions computationally. In this work, we built a Bayesian experimental design framework enabling the highly efficient uncertainty reduction of kinetic parameters and model predictions. Our contributions are 3-fold. First, inspired by previous works aiming at uncertainty reduction of prediction or selected parameters, we proposed two new optimization objectives via model linearization oriented directly to quantities of interest (QoI), *parameter-oriented* and *prediction-oriented* design, for uncertainty reduction of specific parameters and prediction targets, respectively. We conducted theoretical analyses to link Bayesian information gain with dimensionless sensitivity (referred to as impact numbers) and to demonstrate the necessity of implementing QoI-oriented Bayesian experimental design (QBED). Second, neural network response surfaces with both kinetic parameters and experimental conditions as inputs were applied to the experimental design so that a single unified response surface can provide fast, differentiable predictions under a wide range of conditions. It not only facilitates gradient-based design but also accelerates enumeration-based design by parallel computing. Third, we integrated the posterior approximation by linearizing response surfaces with gradient ascent for design optimization. Comparisons with the enumeration-based method demonstrate that gradient-based design usually has a higher average information gain, while enumeration-based design, when assisted by the unified response surface, shows a faster computational speed with acceptable suboptimality. Comprehensive numerical experiments were conducted on the ignition delay times and laminar flame speeds of methanol. Statistical analysis was performed to prove the effectiveness of our methods. The dynamic evolution of uncertainty reduction was unraveled and is well supported by the insights from impact numbers. The proposed method can finish one design-inference iteration in 0.5 s in the 3-D design space and 1.6 s in the 9-D space on an NVIDIA GeForce RTX 2080 Ti graphics processing unit. The QBED source code was made available online.

1. INTRODUCTION

Kinetic parameters in combustion chemistry models are usually determined by theoretical calculation or direct measurements, both of which can introduce a large uncertainty into the parameters. To mitigate the uncertainty, a common approach is to calibrate these parameters against global combustion measurements, such as laminar flame speeds (LFS), ignition delay times (IDT), and species time-histories.[1] However, not all measurements are equally informative: some would be very helpful to reduce the posterior uncertainty, while others are not. Then, an important question arises: how to pick these

Received: May 15, 2024
Revised: July 12, 2024
Accepted: July 15, 2024
Published: July 29, 2024



more informative experimental conditions, a concept commonly known as experimental design or design of experiments (DoE).

Empirically, we have two simple principles to evaluate the information provided by an experiment. The first principle involves selecting experimental conditions where the model prediction for a specific target is highly sensitive to the parameters being inferred. This principle is evident in direct measurements of rate constants, which are typically conducted under conditions where the model prediction is extremely sensitive and primarily depends on one or two reaction rate constants.[2] The second principle is to select conditions with minimal measurement uncertainty. Experimentalists consistently strive to reduce the uncertainty of instruments[3] and to eliminate the interference of nonchemical factors, or nonideal effects in measurements.[4] In the past, combustion researchers have been passionate about determining the accurate values of kinetic parameters. These two empirical principles have been generally adopted by experimentalists. However, from the 1980s, the community gradually realized that developing a sufficiently accurate kinetic model is neither possible nor necessary.[1] A practical approach is to develop models with a well-quantified uncertainty. As uncertainty quantification (UQ) draws increasing attention in the combustion community, many systematic DoE algorithms for uncertainty reduction, instead of ambiguous, empirical principles, have been developed or implemented by combustion engineers[5] in recent years. At the same time, as an important topic of statistics, many DoE algorithms have been developed by statisticians and computer scientists.[6,7]

The earliest attempt for DoE in the combustion community is sensitivity analysis,[2,8] and it has dominated the design of kinetic experiments for several decades. However, it cannot provide a decision when multiple parameters or multiple targets are included. Frenklach and co-workers[9] developed the first DoE algorithm for combustion kinetics besides sensitivity analysis, where the posterior interval of parameters was evaluated. Mosbach et al.[10] designed experiments of compression engines using the determinant of the Fisher information matrix. Huan and Marzouk[11] combined polynomial chaos response surfaces and double-loop Monte Carlo (MC) sampling to evaluate expected information gain (EIG). This is the first work in which Bayesian experimental design was applied to combustion problems. vom Lehn et al.[12] evaluated EIG using the posterior covariance matrix analytically derived by the method of uncertainty minimization using polynomial chaos expansion (MUMPCE).[13] Wang et al.[14] assumed that both kinetic uncertainty and systematic errors would linearly impact the measurement so that the information gain can be analytically derived. It is worth noting that both Wang et al.[14] and vom Lehn et al.,[12] where the parameter-prediction relation was linearized to avoid MC sampling, can be considered as an approximated version of Huan and Marzouk.[11] Sheen and Manion[15] developed a method called experimental design by differential entropy, where a subset of potential experimental conditions was selected by assessing information flux in and out of targets. Li et al.[16] proposed sensitivity entropy as a measure of information gain, where the sensitivity entropy could measure the degree to which the sensitivities of different parameters are concentrated or dispersed.

An experimental design framework for uncertainty reduction usually includes three components. (1) Information measure, which quantifies the information one experiment provides given a design, such as the EIG in ref 11 (computed by double-loop MC) and sensitivity entropy in ref 16. (2) Design optimization methods that find the optimal condition in the design space by maximizing the information measure. The aforementioned research (except for ref 11, where gradient-free optimization was employed) first constructed a set of candidate conditions, evaluated the information gain for each candidate condition, and chose the optimal condition as the one with the highest information gain. (3) Response surfaces, which map the parameters to model predictions directly instead of performing physical model evaluations to reduce the computational cost, since model prediction will be repeatedly evaluated. Response surfaces have generally been used in both UQ and DoE works, such as polynomials in ref 12, polynomial chaos expansion in ref 11, and neural networks (NNs) in refs 16 and 17. In this study, we aim to identify and bridge gaps for these three components accordingly.

The gap in the information measure part is that most algorithms focus on reducing the uncertainty of the entire model,[12] or of several unspecified parameters.[16] This is an obvious *goal mismatch* from the common needs of reducing the uncertainty of specific parameters or prediction targets. Some studies only calibrated parameters of interests while fixed parameters out of interests,[11] leading to an *overconfident* estimation of uncertainty (see the analysis in Section 2.5). Some algorithms have been proposed to deal with this issue, such as the focused optimal Bayesian experimental design[18] for uncertainty reduction of selected parameters, goal-oriented or prediction-oriented experimental design for uncertainty reduction of model prediction.[19,20] In this paper, our first contribution is to propose fast estimators via model linearization for *parameter-oriented* and *prediction-oriented* design, collectively termed as *QoI-oriented Bayesian experimental design (QBED)*, based on the determinant of the posterior covariance matrix used in ref 12. QBED entails a minimal increase in computational cost and requires only minor changes to the existing measure but shows better performance than model-oriented design in various scenarios. Prediction-oriented design can also be used to determine alternative conditions for reducing the prediction uncertainty of targets that cannot be reached by experimental facilities. For example, data measured under low temperatures can be used to reduce the prediction uncertainty at high temperatures. Some algorithms have been developed to deal with this task,[21,22] but our method takes priors and measurement noise into consideration (instead of only sensitivity) and directly optimize the prediction uncertainty.

In the design optimization part, most previous works used an enumeration-based method on grid points, which can be inefficient when the design space is high-dimensional. In the machine learning and statistics communities, several gradient-based experimental design algorithms have been developed.[23−26] However, in these studies, the information measure involves time-consuming MC sampling or variational inference. We combine gradient-based optimization with a highly efficient information measure developed by vom Lehm et al.,[12] enabling fast and fully differentiable evaluation of the expected information gain (EIG). All the matrix operations are done in PyTorch environment,[27] and hence fully differentiable.

For the construction of response surfaces, previous work usually built an individual response surface for each discretized condition, excluding the possibility of applying gradient-based

https://doi.org/10.1021/acs.energyfuels.4c02299
Energy Fuels 2024, 38, 15630−15641

methods in high-dimensional problems and accelerating enumeration-based methods via parallel computing on advanced hardware. Thus, a *single*, *unified* response surface is necessary. However, building a unified response surface has several challenges. First, since the inputs must include design conditions, the output can change across several orders of magnitude under different conditions (such as IDT), and such a change can be highly nonlinear and even nonmonotonic (such as LFS versus equivalence ratio). Second, we cannot use active learning algorithms such as the one proposed by Oh et al.,[28] to make training samples focus on the posterior region, since the posterior region is not known *a priori*. Inspired by Zhang et al.,[17] where thermodynamic conditions are also included as inputs of a response surface for UQ, a unified response surface that can provide fast, accurate prediction was applied to the DoE task for the first time. Neural networks are chosen due to their great expressive power and compatibility with the parallelism of graphics processing units (GPU).

2. METHODS

2.1. Pipeline. The pipeline of the kinetic experimental design in this work is shown in Figure 1. The steps in this pipeline can be

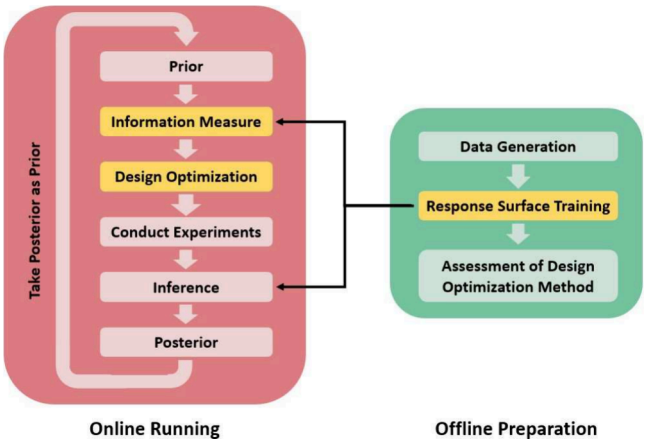


Figure 1. Pipeline of kinetic experimental design in this work. The three components to which our contribution corresponds are highlighted as Information Measure, Design Optimization, and Response Surface Training.

divided into offline preparation and online running. In the offline stage, we need to generate training samples and train the NN-based response surface so that we can replace the physical model by surrogate models for fast evaluation in the online stage. Then before we start the online stage, preliminary numerical experiments are needed to determine which design optimization should be used. We provide two choices, point-wise enumeration and gradient ascent. A proper sample size for enumeration and a proper number of initializations for gradient ascent should also be determined. In the online stage, we perform sequential experimental design. Starting from the prior distribution, we evaluate the information provided by an experimental measurement according to a predefined information measure and then maximize such as measure within the design space. After each design time, we collect experimental measurements (synthetic data generated by response surfaces in this paper) at the design point (according to eq 1). Then, these data are used to perform inference to obtain the posterior distribution. The inference method we employ is MUMPCE, but our approach is different from the method in ref 17 in two aspects. First, we do not eliminate inconsistent data since we generate synthetic data using the presumed model plus Gaussian noise; second, we do not freeze any parameters. In the next iteration of design inference, the posterior distribution is used as the prior. Such a design-inference iteration is repeated a given number of times (5 in this work). The response surfaces are used in both design and inference. Our main contributions lie in the three steps highlighted in Figure 1. It is worth noting that if design cannot be conducted along with the experimentation, our pipeline can also design all experiments at once by just taking n in eq 11 equal to the number of experiments that will be conducted.

2.2. Preliminaries. We present the preliminaries for Bayesian DoE following Huan and Marzouk[11] and vom Lehm et al.[12]

In the Bayesian model calibration, we assume that the experimental measurement is generated by model prediction at ground truth parameters plus noise:

$$\boldsymbol{y} = M(\boldsymbol{k}, \boldsymbol{d}) + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{k}$ is the vector of kinetic parameters, $\boldsymbol{y}$ is the vector of experimental measurements, $\boldsymbol{d}$ is the vector of design (including conditions such as temperature, pressure, equivalence ratio, etc.), and $M(\cdot)$ is the model prediction. $\boldsymbol{\epsilon}$ is the measurement noise, which usually follows a uniform Gaussian distribution. The Bayes' Theorem with design $\boldsymbol{d}$ can be expressed as

$$p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d}) = \frac{p(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{d})p(\boldsymbol{k})}{p(\boldsymbol{y}|\boldsymbol{d})} = \frac{p(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{d})p(\boldsymbol{k})}{\int_{\boldsymbol{K}} p(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{d})p(\boldsymbol{k})\,\mathrm{d}\boldsymbol{k}} \tag{2}$$

where $\boldsymbol{K}$ is the support of $\boldsymbol{k}$. Here, we assume that the prior distribution $p(\boldsymbol{k})$ is irrelevant to $\boldsymbol{d}$. Kullback–Leibler (K-L) divergence is commonly employed to quantify the difference between two distributions from an information-theoretic standpoint. The K-L divergence of prior and posterior distributions serves as a metric to gauge the information gain:

$$D_{\mathrm{KL}}[p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})\|p(\boldsymbol{k})] = \int_{\boldsymbol{K}} p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})\ln\left[\frac{p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})}{p(\boldsymbol{k})}\right]\mathrm{d}\boldsymbol{k} \tag{3}$$

K-L divergence is always non-negative, and it equals zero if and only if two distributions are identical. The larger the K-L divergence is, the larger the difference between the two distributions. Since we cannot ascertain the experimental data $\boldsymbol{y}$ beforehand, we can weight the K-L divergence by the probability density function (PDF) of $\boldsymbol{y}$, resulting in the *expected information gain* (EIG) $U(\boldsymbol{d})$:

$$\begin{aligned} U(\boldsymbol{d}) &= \int_{\boldsymbol{Y}} p(\boldsymbol{y}|\boldsymbol{d})D_{\mathrm{KL}}[p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})\|p(\boldsymbol{k})]\mathrm{d}\boldsymbol{y} \\ &= \int_{\boldsymbol{Y}}\int_{\boldsymbol{K}} p(\boldsymbol{y}|\boldsymbol{d})p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})\ln\left[\frac{p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})}{p(\boldsymbol{k})}\right]\mathrm{d}\boldsymbol{k}\,\mathrm{d}\boldsymbol{y} \end{aligned} \tag{4}$$

In addition, EIG can also be viewed as the expected entropy difference between prior and posterior distribution:[29]

$$\begin{aligned} U(\boldsymbol{d}) &= \int_{\boldsymbol{Y}}\int_{\boldsymbol{K}} p(\boldsymbol{y}|\boldsymbol{d})p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})\ln p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})\mathrm{d}\boldsymbol{k}\,\mathrm{d}\boldsymbol{y} - \int_{\boldsymbol{Y}}\int_{\boldsymbol{K}} p(\boldsymbol{k}|\boldsymbol{d}) \\ &\quad \times p(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{d})\ln p(\boldsymbol{k})\mathrm{d}\boldsymbol{k}\,\mathrm{d}\boldsymbol{y} \end{aligned} \tag{5}$$

$$= \int_{\boldsymbol{Y}} -p(\boldsymbol{y}|\boldsymbol{d})H[p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})]\mathrm{d}\boldsymbol{y} + \int_{\boldsymbol{Y}} p(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{d})H[p(\boldsymbol{k})]\mathrm{d}\boldsymbol{y} \tag{6}$$

$$= \int_{\boldsymbol{Y}} -p(\boldsymbol{y}|\boldsymbol{d})H[p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})]\mathrm{d}\boldsymbol{y} + H[p(\boldsymbol{k})] \tag{7}$$

$$= \int_{\boldsymbol{Y}} p(\boldsymbol{y}|\boldsymbol{d})\{H[p(\boldsymbol{k})] - H[p(\boldsymbol{k}|\boldsymbol{y}, \boldsymbol{d})]\}\mathrm{d}\boldsymbol{y} \tag{8}$$

where $H[p(\boldsymbol{\theta})] = -\int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta})\ln p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ is the entropy of a distribution with PDF $p(\boldsymbol{\theta})$.

The goal of Bayesian DoE is to find the optimal experimental condition within the design space $\mathcal{D}$:

$$\boldsymbol{d}^* = \arg\max_{\boldsymbol{d}\in\mathcal{D}} U(\boldsymbol{d}) \tag{9}$$

In Huan and Marzouk,[11] eq 4 is estimated through double-loop Monte Carlo (MC) sampling, which is asymptotically unbiased but computationally expensive. For the details of this algorithm, readers

https://doi.org/10.1021/acs.energyfuels.4c02299
Energy Fuels 2024, 38, 15630−15641

should refer to the original paper.[11] Rainforth et al.[30] have demonstrated that this double-loop (or nested) MC estimator has an error that scales with the computational cost at an order of $-\frac{1}{3}$. So prohibitively expensive computation is inevitable if one sticks to the full Bayesian formulation. In the combustion community, a simplified version of the full Bayesian DoE is often implemented. In this work, we adopted the criterion proposed by vom Lehn et al.,[12] where the posterior was evaluated using MUMPCE.

In the MUMPCE framework with Gaussian prior and independent and identically distributed (iid) additive Gaussian noises,[13] the posterior distribution is obtained by maximizing the numerator part of eq 2 to get the maximum a posteriori (MAP) estimation $\boldsymbol{k}^*$ first, and then by linearizing the response surface around MAP point, the posterior distribution would be Gaussian and the covariance matrix can be analytically derived:

$$\boldsymbol{\Sigma} = \left[\sum_{r=1}^{n} \frac{\boldsymbol{J}_r(\boldsymbol{k}^*, \boldsymbol{d}_r)\boldsymbol{J}_r(\boldsymbol{k}, \boldsymbol{d}_r)^{*T}}{(\sigma_r)^2} + \sum_{0}^{-1}\right]^{-1} \tag{10}$$

Here, $\boldsymbol{J}_r(\boldsymbol{k}^*, \boldsymbol{d}_r)$ is the $m \times p$ Jacobian matrix of model response evaluated at $\boldsymbol{k}^*$ and condition $\boldsymbol{d}_r$, i.e., $\partial \boldsymbol{M}(\boldsymbol{k}, \boldsymbol{d}_r)/\partial \boldsymbol{k}|_{k=k^*}$, where m is the number of parameters, p is the dimension of measurements for one experiment, n is total number of measurements. $\boldsymbol{\Sigma}_0$ is the covariance matrix of the prior distribution (for uniform prior distributions, the posterior is just the above equation with all-zero matrix for $\boldsymbol{\Sigma}_0^{-1}$), and σ_r is the standard deviation of the Gaussian noise. In Bayesian statistics, the technique of approximating posterior distribution by linearizing models around the MAP point is called Laplacian approximation, which provides a Gaussian distribution whose logarithm of PDF has a curvature same as that of the actual posterior distributions.[31]

vom Lehn et al.[12] did not draw samples from $p(\boldsymbol{y})$, evaluate information gain for each sample of $\boldsymbol{y}$, and take the average. Instead, they used the information gain where the measurement $\boldsymbol{y}$ is the prediction at the nominal values of kinetic parameters, as the EIG. In this case, the MAP estimation is exactly the same as the nominal values. Thus, the "expected" posterior covariance matrix is assumed to be

$$\boldsymbol{\Sigma}_E(\boldsymbol{d}) = \left[\sum_{r=1}^{n} \frac{\boldsymbol{J}_r(\boldsymbol{k}^0; \boldsymbol{d}_r)\boldsymbol{J}_r(\boldsymbol{k}^0; \boldsymbol{d}_r)^T}{(\sigma_r)^2} + \boldsymbol{\Sigma}_0^{-1}\right]^{-1} \tag{11}$$

and $\boldsymbol{d}_r$ is the condition for r-th experiment, $\boldsymbol{d} = \{\boldsymbol{d}_1, \boldsymbol{d}_2, ..., \boldsymbol{d}_n\}$. Note that the Jacobian matrix would change if we change the conditions under which the experiment is conducted. From eq 8 we can know that for a certain Gaussian prior, maximizing EIG is equivalent to minimizing the expected posterior entropy. The "expected" posterior is a Gaussian distribution with a covariance matrix, shown in eq 11. For a Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}$, the entropy is

$$H = \frac{m}{2}(1 + \ln(2\pi)) + \frac{1}{2}\ln\det\boldsymbol{\Sigma} \tag{12}$$

where m is the dimension of the Gaussian distribution. Thus, minimizing the expected posterior entropy is equivalent to minimizing the determinant of the expected posterior covariance matrix:

$$\boldsymbol{d}^* = \arg\min_{\boldsymbol{d}\in\mathcal{D}} \det \boldsymbol{\Sigma}_E(\boldsymbol{d}) \tag{13}$$

To gain more insight into eq 13, consider a Bayesian linear model with Gaussian prior $N(0, \boldsymbol{\Sigma}_0)$ and Gaussian noise $\epsilon \sim N(0, \boldsymbol{\Sigma}_\epsilon)$, where the measurement is given as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{k} + \epsilon$. In this case, the EIG in eq 4 is exactly $\ln[\det(\boldsymbol{\Sigma}_0)] - \ln[\det(\boldsymbol{X}\boldsymbol{\Sigma}_\epsilon{}^{-1}\boldsymbol{X}^T + \boldsymbol{\Sigma}_0^{-1})^{-1}]$.[32] Again maximizing EIG is equivalent to minimizing the second term. Then, the minimization in eq 13 is a linear approximation for the maximization of EIG of nonlinear model, where the parameter−prediction relation is linearized by the Jacobian matrix. Thus, as long as the physical model is linear enough around the nominal value, linear approaches of DoE should be close to a full Bayesian design. In fact, before ref 12, this estimator extended from Bayesian linear models has been generally used by statisticians.[32]

When the noise is multiplicative instead of additive, we just use a new Jacobian where the output is the logarithm of the original output and replace the noise level with the logarithm (same base as above) of the original noise level:

$$\boldsymbol{\Sigma}_E(\boldsymbol{d}) = \left[\sum_{r=1}^{n} \frac{\boldsymbol{J}_r(\boldsymbol{k}^0; \boldsymbol{d}_r)\boldsymbol{J}_r(\boldsymbol{k}^0; \boldsymbol{d}_r)^T}{(\ln\sigma_r)^2} + \boldsymbol{\Sigma}_0^{-1}\right]^{-1} \tag{14}$$

2.3. Gradient-Based Design. In our framework, since the constructed NN response surface is fully differentiable with respect to kinetic parameters and experimental conditions, the optimization problem in eq 13 or 14 can be solved by gradient-based optimization algorithms. Here, we employed Adam optimizer[33] due to its adaptive step sizes. The learning rate and weight decay are set as 0.01 and 0. Other parameters are kept as default values. Our experience shows that using gradient descent with a fixed step size suffers from divergence or very slow convergence if the step size is not picked carefully, while Adam can work very well with a learning rate fixed at 0.01. We transformed each condition linearly into the range $[-1, 1]$. Since we need to constrain our decision variable within the design space, we clamped each condition within the range of design space after each iteration (clamping within $[-1, 1]$ after linear transformation). Multiple initializations in optimization can avoid being trapped in the local minimum, and the effect of the number of initializations is discussed in Section 3. After the optimization finishes, the initialization with the largest EIG can be picked as the final optimal design.

2.4. QoI-Oriented Bayesian Experimental Design. In eq 13 or 14, the optimizer would minimize the volume of a hyper-ellipse described by the covariance matrix of all kinetic parameters. We call it a *model-oriented* experimental design. As mentioned in Section 1, in practice, minimizing the uncertainty of all kinetic parameters is often not the primary objective in model calibration using experimental measurements. Therefore, depending on the specific QoI, a modified objective function can be employed, which is termed *QoI-oriented Bayesian experimental design (QBED)*. First, consider the scenario where the aim is to minimize the uncertainty of one or several kinetic parameters. This approach is termed a *parameter-oriented* experimental design. In this case, only the rows and columns corresponding to the parameters of interest are selected before computing the determinants:

$$\boldsymbol{d}^* = \arg\min_{\boldsymbol{d}\in\mathcal{D}} \det\left[\left(\sum_{r=1}^{n} \frac{\boldsymbol{J}_r(\boldsymbol{0}; \boldsymbol{d}_r)\boldsymbol{J}_r(\boldsymbol{0}; \boldsymbol{d}_r)^T}{(\sigma_r)^2} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\right]_{I,I} \tag{15}$$

where I is the set of indices of parameters of interests. For example, if we only want to minimize the uncertainty of parameters 2 and 3, then $I = \{2, 3\}$, and $(\cdot)_{I,I}$ means picking the second and third rows and second and third columns of the matrix so that we would obtain a new 2×2 matrix.

The second scenario involves minimizing the prediction uncertainty of one or a series of targets,[19,20] referred to as *prediction-oriented* experimental design. Given the covariance matrix of parameters, we can obtain the covariance of (a series of) targets (such as IDT or LFS at several different conditions) analytically by linearizing the response surface (first-order version of that in MUMPCE[13]). In practice, we usually consider minimizing the absolute value of uncertainty of each target instead of the volume of hyper-ellipse of the targets' covariance matrix. Thus, here we minimize the sum of prediction standard deviations of all targets:

$$\boldsymbol{d} \in \min_{\mathcal{D}} \sum \operatorname{diag}\left[\boldsymbol{J}_t^T\left(\sum_{r=1}^{n} \frac{\boldsymbol{J}_r(\boldsymbol{0}; \boldsymbol{d}_r)\boldsymbol{J}_r(\boldsymbol{0}; \boldsymbol{d}_r)^T}{(\sigma_r)^2} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\boldsymbol{J}_t\right]^{1/2} \tag{16}$$

https://doi.org/10.1021/acs.energyfuels.4c02299
Energy Fuels 2024, 38, 15630−15641

where diag(·) means taking the diagonal elements of a matrix, and $\boldsymbol{J}_t$ means the Jacobian matrix of the targets whose uncertainty is desired to be reduced. We notice that this objective function is similar to the goal of A-optimality for linear models, where one minimizes the sum of variances of linear combinations of model parameters.[34] For targets that lie in a continuous condition range, we discretize the range into uniform points and then compute the Jacobian matrix on these points. For the covariance matrix, it is exactly the sum of standard deviations of all targets after picking diagonal elements, taking elementwise square root and summing them over.

2.5. Theoretical Analysis. In this subsection, we derive the relation between the Bayesian EIG and sensitivity analysis by leveraging the approximation algorithm described above. Suppose that we have a m-parameter, p-output model $\boldsymbol{M}(\boldsymbol{k}, \boldsymbol{d})$, where $\boldsymbol{k}$ is the vector of parameters and $\boldsymbol{d}$ is the vector of design. The Jacobian matrix with respect to parameters at their nominal values and a design $\boldsymbol{d}$ (i.e., $\nabla_k \boldsymbol{M}(\boldsymbol{k}, \boldsymbol{d})$ at nominal values) is $\boldsymbol{J} \in \mathbb{R}^{m\times p}$. The prior $p(\boldsymbol{k})$ is a Gaussian distribution with the covariance matrix $\boldsymbol{\Sigma}_0$. The measurement noise follows a Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}_\epsilon$.

Theorem 1: *By linearizing the model around the nominal values of parameters, the expected information gain in eq 4 can be approximated as*

$$\frac{1}{2}\ln\frac{\det(\boldsymbol{J}^T\boldsymbol{\Sigma}_0\boldsymbol{J}+\Sigma_\epsilon)}{\det(\boldsymbol{\Sigma}_\epsilon)} \tag{17}$$

This can be easily proven by the equivalent definition of EIG of mutual information[29] and basic properties of Gaussian random variables. The proof details are shown in the Supporting Information. Following a similar way, we can obtain the EIG for selected parameters.

Corollary 1.1: *By linearizing the model around the nominal values of parameters, the expected information gain in eq 4 for a subset of parameters I can be approximated as*

$$\frac{1}{2}\ln\frac{\det(\boldsymbol{J}^T\boldsymbol{\Sigma}_0\boldsymbol{J}+\boldsymbol{\Sigma}_\epsilon)}{\det(\boldsymbol{J}_{I^C}^T(\boldsymbol{\Sigma}_0)_{I^C,I^C}\boldsymbol{J}_{I^C}+\boldsymbol{\Sigma}_\epsilon)} \tag{18}$$

where I^C is the complement of I.

Here $(\boldsymbol{\Sigma}_0)_{I^C,I^C}$ means only retaining the rows and columns of the prior covariance corresponding to the parameters out of I, and $\boldsymbol{J}_{I^C}$ means only retaining the rows of the Jacobian matrix corresponding to the parameters out of I. The proof is also shown in the Supporting Information. In fact, this is exactly the EIG estimator used in ref 14, where the parameters out of interest are "latent random variables" in systematic uncertainty, such as reflected shock temperature or pressure, etc. This indicates that kinetic parameters that are of our interest can be interpreted as some random variables that would cause systematic measurement uncertainty.

To gain some insights, we can consider a single-output toy problem with iid Gaussian prior.

Corollary 1.2: *For a single-output model with iid Gaussian prior and Gaussian measurement noise, after linearizing the model around the nominal values of parameters, the expected information gain in eq 4 for a subset of parameters I is*

$$\frac{1}{2}\ln\frac{1+\Sigma_{i=1}^m f_i^2}{1+\Sigma_{i\in I^C} f_i^2} \tag{19}$$

where $f_i = \boldsymbol{J}_i\sigma_i/\sigma$.

Here, $\boldsymbol{J}_i$ is the sensitivity of output with respect to parameter i, σ_i is the standard deviation of the Gaussian prior of parameter i, and σ is the standard deviation of the Gaussian measurement noise. The proof is just substituting a specific form of Jacobian matrix, noise covariance, and prior variance into eq 18, which is also included in the Supporting Information

We can have several insights from the analysis above. (1) $f_i = \boldsymbol{J}_i\sigma_i/\sigma$, defined as impact numbers, play a key role in Bayesian EIG. It can represent the impact of the uncertainty of parameter i on the model prediction compared with measurement noise, which is similar to the signal-to-noise ratio in the scenario of model calibration. Intuitively, if the impact of the uncertain parameter i on model prediction is small compared with measurement noise, then this parameter is hard to calibrate. We note that this is similar to the impact factor defined in Tao et al.,[35] but impact numbers are dimensionless as it is divided by the noise level. (2) Increasing the impact number out of QoI can cause the opposite change for the EIG of the entire model and the EIG of QoI. Thus, model-oriented design can mismatch with the goal of reducing the uncertainty of specific parameters. For the scenario of reducing the uncertainty of prediction, suppose we have a target that is only related to parameters in subset I, and then the conditions identified by model-oriented design also mismatch with the goal. Thus, QoI-oriented Bayesian experimental design (QBED) is necessary. (3) Since fixing the parameters out of QoI (such as ref 11) is equivalent to setting all the impact numbers out of QoI as 0, it can lead to overconfident estimation of uncertainty of QoI. At the same time, eq 19 can provide a way to quantify the error of fixing some parameters *a priori*. Thus, fixing parameters with small impact numbers is still feasible as long as the error is at an acceptable level. (4) Some previous methods, such as choosing conditions where the prediction is only sensitive to the parameter of interests, or conditions with small sensitivity entropy[16] (i.e., the sensitivities to different reactions have a small degree of dispersion), can be considered as a heuristic approximated of Bayesian experimental design, but they can be misleading. For example, for a two-input model with parameter 1 as QoI, $f_1 = \sqrt{0.5}$, $f_2 = 0$ has a smaller EIG than $f_1 = f_2 = 2$, but the former obviously has a smaller sensitivity entropy.

2.6. Task Setup and Data Generation. In the case studies, our objective is to design experiments for laminar flame speed (LFS) and ignition delay time (IDT) measurements to minimize the uncertainty of a methanol combustion kinetic model developed by Zhang et al.,[36] which is also used in Zhou et al.[37] Pressure P, temperature T, and equivalence ratio ϕ are chosen as design conditions. Only pre-exponential factors are considered uncertain parameters for demonstration, while other kinetic parameters are fixed. The uncertainty factors UF_i are adopted from the Supporting Information of Zhou et al.[37] The prior distribution is log-normal:

$$x_i = \frac{\ln(k_i/k_{i,0})}{\frac{1}{2}\ln UF_i} \sim N(0,1) \tag{20}$$

where k_i is the random variable of each kinetic parameters, and $k_{i,0}$ is its nominal value. This mechanism includes 32 species and 197 elementary reactions. Plus, the high- and low-pressure limits of 16 falloff reactions, the dimension of parameters is 213. Two duplicate reactions, R95 and R96, were assigned with the same uncertainty factor. Observing the data set used by FFCM-1,[38] the $1-\sigma$ additive noise is around 10% of measured values for LFS and 20% for IDT. Thus, we assigned a multiplicative noise $\sigma_r = 1.1$ for LFS and $\sigma_r = 1.2$ for IDT.

To minimize computational costs, we employ NN-based response surfaces to replace expensive physical simulations. The training data were generated from physical simulations. We used 300,000 samples to build the response surface of IDT, among which 243000 samples were for the training set, 30000 for the test set, and 27000 for the validation set. 106920 samples were used to build the response surface of LFS, among which 87480 samples were for the training set, 9,720 for the test set, and 9720 for the validation set. We take $\ln(k_i/k_{i,0})$ as the input of response surfaces, which follows zero-mean Gaussian distributions with standard deviation $\frac{1}{2}\ln UF_i$. In order to emphasize the accuracy in the region near the prior mean, for each response surface, we generated three hierarchical data subsets with different ranges for parameters, similar to Zhang et al.[17] Specifically, samples were generated by Latin hypercube sampling[39] from uniform distributions for both parameters and design (except temperatures in LFS). The space of the training samples is listed in Table 1. Please note that for prediction-oriented design, the design space is smaller than the space. For pressure, the uniform distribution was assumed for the logarithm of pressure. The parameter space is defined by a hypercube $\left[-\frac{k}{2}\ln UF_i, \frac{k}{2}\ln UF_i\right]$, where k = 1, 2, and 3 for three

https://doi.org/10.1021/acs.energyfuels.4c02299
Energy Fuels 2024, 38, 15630−15641

Table 1. Range of Training Samples

experimental targets	P [atm]	T [K]	ϕ
ignition delay times	1−30	800−1600	0.5−5
laminar flame speeds	0.5−10	300−500	0.8−1.5

subsets, respectively. In other words, the half-widths of the uniform distribution region for three subsets are σ_p, $2\sigma_p$, and $3\sigma_p$, respectively, where σ_p is the standard deviation of the prior distribution $\frac{1}{2}\ln \mathrm{UF}_i$. Clearly, the region near the prior mean is covered by all three sets. Different from Zhang et al.,[17] where three subsets were generated from a *Gaussian* distribution with standard deviation $0.2\sigma_p$, $0.6\sigma_p$, and σ_p respectively, our samples were generated from a uniform distribution, so that the response surfaces can still have good accuracy near the tail of the prior distribution. Then, each subset was split into training, validation, and test sets using the same ratio. We did not sample temperature from a uniform distribution in order to accelerate the solution of laminar flame from the initial guess in Cantera,[40] which is explained in the Data Generation Section of Supporting Information. The detailed setting of Cantera is also included in the Supporting Information.

2.7. Neural Network Response Surfaces. We used neural networks as response surfaces due to their great expressive power, especially in high-dimensional cases, and their good performance of parallel evaluation on a GPU. Gradient-based experimental design requires the gradient of information gain with respect to design variables, so design variables must be included as inputs. The network structure is shown in Figure 2, including three hidden layers with 32

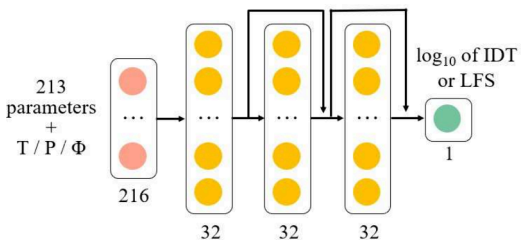


Figure 2. Schematic of a 3-hidden-layer ResNet structure for the unified response surface, including both design conditions and parameters as inputs.

neurons in each hidden layer. The input layer includes 216 neurons, corresponding to 213 kinetic parameters and three experimental conditions. In order to fit data across multiple orders of magnitude (six orders for IDT and one order for LFS) and minimize the multiplicative error during training, we took the logarithm of IDT and LFS as the outputs of response surfaces.

Residual neural network (ResNet)[41] structure was adopted for its good property of mitigating gradient explosion and gradient vanishing during back-propagation. Two short circuits were set in the second and third hidden layer. The SiLU[42] activation function was employed to avoid zero Hessian (see Supporting Information for explanations). Other details of the neural network and loss curves can also be found in the Supporting Information.

The assessment for accuracy of response surfaces followed the method in Zhang et al.,[17] where the mean and 95th percentile relative error of the *test set* were compared with a given criterion, as shown in Table 2. Relative error is defined as $|y_{NN} - y_G|/y_G$, where y_{NN} is the prediction of LFS or IDT by neural networks, y_G is the corresponding ground truth. Note that the outputs of neural networks are the logarithms of LFS and IDT, while the relative error uses the original value of LFS and IDT. The mean, 95th percentile, and maximum relative errors are shown in Tables 3 and 4. Note that the parameters of our data were sampled from uniform distributions covering a wider range than the Gaussian distributions in Zhang et al.[17] Thus, building such a response surface is more challenging than in the case of Zhang et al.[17] However, the performance of our neural networks still can satisfy the requirement, as shown in Table 2.

Table 2. Requirement of the Accuracy of Response Surfaces Proposed by Zhang et al.[17]

subsets	mean error (%)	95th percentile error (%)
1	1	2
2	2	5
3	3	10

Table 3. Accuracy of Our Response Surface for IDT

subsets	mean error (%)	95th percentile error (%)	maximum error (%)
1	0.5	1.9	1.5
2	0.8	2.9	39.8
3	1.6	5.8	76.5

Table 4. Accuracy of Our Response Surface for LFS

subsets	mean error (%)	95th percentile error (%)	maximum error (%)
1	0.7	1.5	3.5
2	1.1	2.8	6.9
3	2.0	5.1	16

The scatter plot of prediction versus ground truth for samples in test sets is shown in Figures S2 and S3 in the Supporting Information. We can clearly see that, for subsets 1 and 2, almost all samples lie on the diagonal line. For subset 3, only a small portion of samples deviated from the diagonal line.

3. NUMERICAL EXPERIMENTS: RESULTS AND DISCUSSION

In this section, we show the results of numerical experiments using the algorithms presented in Section 2. In previous studies,[12,16] the proposed algorithms were usually assessed by a single case. However, the results can be affected by randomness (or uncertainty) from two sources: noise and ground truth. The change of both can lead to different synthetic data used for inference and hence different information gain. However, noise differs among each realization, while the ground truth is not known *a priori*. To eliminate the randomness/uncertainty in assessment, we follow Shen and Huan,[43] where the pipeline was repeated lots of times and statistical analysis was done. Specifically, in each realization, we sample a parameter from a prior distribution as the ground truth and generate noisy synthetic data from such a "ground truth". Then inference was done based on the noisy data, and information gain for this case was computed from posterior uncertainty, defined as the part within the curly braces of eq 8, i.e., the difference of entropy of prior and posterior distribution after getting the synthetic experimental data. In all figures, the information gain was computed in the way described above. This process would be repeated 100 times. The average and its variance would be computed (equal to sample variance over sample size according to the central limit theorem) to determine the optimal design optimization strategies and to demonstrate the performance of our methods. For figures involving multiple iterations of design-inference (such as Figures 5 and 7), information gain refers to the entropy difference of the prior distribution before the first iteration and the current posterior distribution. All of the experiments were run on an NVIDIA GeForce RTX 2080 Ti GPU.

https://doi.org/10.1021/acs.energyfuels.4c02299
Energy Fuels 2024, 38, 15630−15641

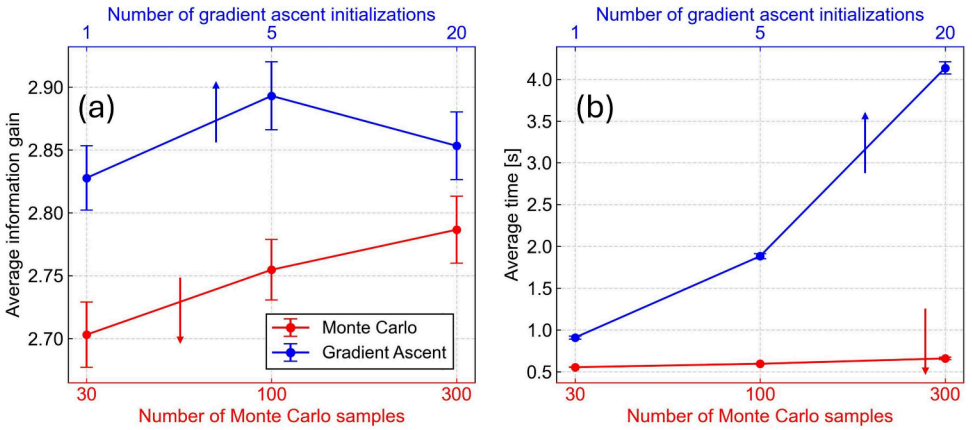


Figure 3. (a) Average information gain and (b) computation time for 100 realizations of the gradient-based method and Monte Carlo method for the prediction-oriented design of LFS with $n = 1$. Different numbers of initializations and samples are respectively considered for the gradient-based method and the Monte Carlo method. Error bar shows the $1 - \sigma$ interval of mean values, which are the standard deviation of 100 samples over $\sqrt{100}$ according to the central limit theorem.

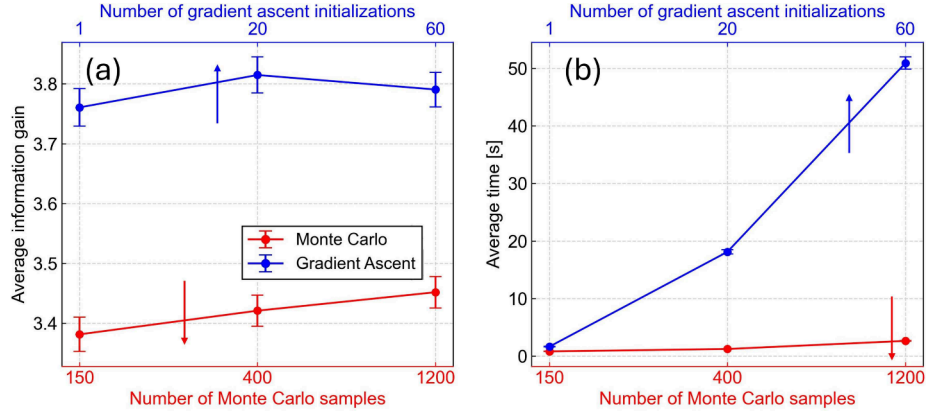


Figure 4. (a) Average information gain and (b) computation time for 100 realizations of gradient-based method and Monte Carlo method for the prediction-oriented design of LFS with $n = 3$. Different numbers of initializations and samples are respectively considered for the gradient-based method and the Monte Carlo method. Error bar shows the $1 - \sigma$ interval of mean values, which are the standard deviation of 100 samples over $\sqrt{100}$ according to the central limit theorem.

3.1. Assessment of Design Optimization Methods. In Section 2, we introduced two methods for design optimization: enumeration-based methods and gradient-based methods. In this subsection, we compare their performance for different sample sizes and different initializations on both high-dimensional design and low-dimensional design cases to determine the optimal strategies for different scenarios. In previous works, the enumeration was done on grid points within the design space. In this work, since we need to compare different design optimization methods, we evaluate information gain on Monte Carlo points to avoid the case where one of the predefined grid points is luckily very close to the global maximum of EIG. Specifically, in each design, we draw a given number of samples from the uniform distribution of design space, evaluate the EIG in each MC point, and pick the condition with the largest EIG as the optimal design. Since we only have three design conditions T, P and ϕ, we use batch design with $n = 3$ in eq 11, as a demonstration of high-dimensional cases, whose dimension is $3 \times 3 = 9$; and $n = 1$ as a demonstration of low-dimensional cases, whose dimension is 3. It is worth noting that designing three experiments simultaneously would be more computationally expensive than the problem of 9 design variables with the same size of response surface since the back-propagation would be done in parallel for three inputs (equivalent to improving the number of initializations, see Figure 3b). Thus, the computational time for a problem with 9 design variables and designing only one experiment would be shorter than the 9-D problem here. In this part, the number of iterations is chosen as 3 for $n = 1$ and 1 for $n = 3$. We only present the results of the prediction-oriented design of LFS to demonstrate the properties, while the results of the other three cases are shown in Figures S4 and S9 of the Supporting Information. We can clearly see that under different scenarios the optimal strategy can be different, and it also depends on users' trade-off between average information gain and computational time. Thus, we need to do a design optimization assessment before we start DoE for any specific case.

The information gain and running time of gradient ascent with different initializations and MC designs with different sample sizes under $n = 1$ are shown in Figure 3. Please note that the range of the y-axis of the left figure is small, which makes the $1 - \sigma$ intervals seem bigger than it is. For gradient ascent, we can conclude that beyond 5 initializations, the increase in the number of initializations would not have an obvious benefit. In fact, even if we only use one initialization,

https://doi.org/10.1021/acs.energyfuels.4c02299
Energy Fuels 2024, 38, 15630−15641

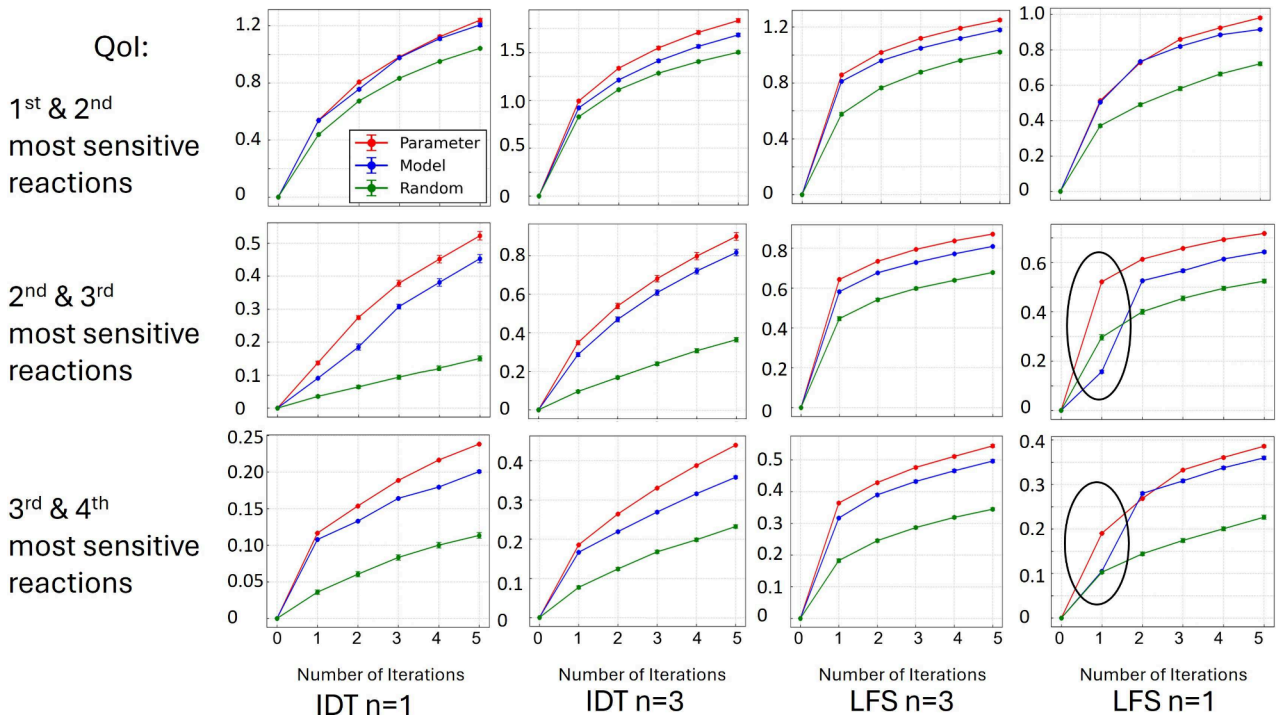


Figure 5. Average information gain of 100 runnings versus the number of iterations for parameter-oriented design (red), model-oriented design (blue), and random experiments (green), for both LFS and IDT scenarios, and with both $n = 1$ and $n = 3$. The first, second, and third rows show the results where the design target is the first and second most sensitive reactions, the second and third most sensitive reactions, and the third and fourth most sensitive reactions, respectively. The information gain is also in terms of the corresponding two target reactions in each case. Two circles highlight the two scenarios where the model-oriented design performs similarly to or even worse than random experiments, while the parameter-oriented design can outperform both.

the difference with 20 initializations is quite small. From eq 16, we know that for one initialization, the average prediction σ decreases to 1/2.83 = 0.353 of the prior one, while for 20 initializations this number is 0.351. We have shown that the landscape of information gain is usually nonconvex (see the landscape contours in Section 3.3 of ref 44.), so gradient-based optimization using one initialization does not have a guarantee to converge to the global minimum. Such a numerical experiment shows that although there are several local maxima in the landscape of information gain, the information gains of different maxima are very similar. In other words, it is arguably not necessary to arrive at the global maximum in some cases. For the MC design, the average information gain increases as the sample size increases, but similarly, the difference is not very large. From 30 to 300 samples, the average information gain is always lower than that of the gradient ascent. In terms of the running time, the number of sample sizes hardly affects the average running time of MC design since the information gain is evaluated in parallel. For gradient ascent, however, with an increase in the number of initializations, the optimization time would increase. On the one hand, since the stop criterion is based on the infinite norm of the difference between two successive iterations, a larger number of iterations for convergence is expected for a larger number of initializations; on the other hand, the time per iteration will increase as the number of initialization increases. Although the evaluation of gradient is also in parallel, it involves the computation of second-order derivative instead of just first-order derivative in evaluating EIG.

Comparing the two methods, MC design with 300 samples can achieve a slightly lower level of information gain as gradient ascent but with far less computation time. Thus, under a low-dimensional scenario, we prefer the MC design.

The sample-independence and initialization-independence study under $n = 3$ is shown in Figure 4. For gradient ascent, the number of initializations, ranging from 1 to 60, does not have an obvious effect on the average information gain compared with the standard deviation. It is highly possible that the different local minima have similar EIG. For MC, the increase of sample size from 150 to 1200, can increase the average information gain. Similar to the low-dimensional case, the average information gain of gradient ascent is always larger than that of the MC method. Even with only one initialization, gradient ascent can get a larger average information gain than that of MC with 1200 samples and at the same time use less time. Thus, in this work, we prefer gradient ascent under high-dimensional cases, considering both the time and performance. Again, the trade-off between computation time and information gain depends on specific scenarios and users' trade-offs. If speed is more important, then the MC method should be chosen. Here we fix the strategy for high-dimensional and low-dimensional cases, respectively, for demonstration.

3.2. Comparison with Model-Oriented Design and Random Experiments. In this subsection, we compare our proposed QBED against model-oriented design and random experiments (i.e., sampling one condition from a uniform distribution of the design space every time) for both LFS and IDT, and both high-dimensional ($n = 3$ in eq 10) and low-dimensional ($n = 1$) cases, to demonstrate the effectiveness of the proposed framework. According to the results in Section 3.1, under high-dimensional cases, the gradient-based method was used, while under low-dimensional cases, the MC method

https://doi.org/10.1021/acs.energyfuels.4c02299
Energy Fuels 2024, 38, 15630−15641

was used. The sample or initialization independence study of model-oriented design was also done to determine the threshold number of samples or initialization, exceeding which the information gain cannot be further increased. The results are shown in Figure S10 of Supporting Information. According to the results, and considering the computational budget, in all the following studies we will use five initializations for gradient ascent under high dimension and 300 samples for MC design under low dimension.

3.2.1. Parameter-Oriented Design. For parameter-oriented design, the design space is the same as the range of training samples listed in Table 1. For both IDT and LFS, the four reactions with the largest impact numbers are assessed. For IDT, they are $CH_3OH + HO_2 = CH_2OH + H_2O_2$, $CH_2OH + O_2 = CH_2O + HO_2$, $2HO_2 = H_2O_2 + O_2$, and $CH_2O + HO_2 = H_2O_2 + HCO$, from large impact number to small. For LFS, they are the following: (i) $CO + OH = CO_2 + H$, (ii) $H + O_2 = O + OH$, (iii) $CH_2OH = CH_2O + H$, and (iii) $CH_2OH + O_2 = CH_2O + HO_2$.

Figure 5 shows the comparison of parameter-oriented design, model-oriented design, and random experiments under 12 different cases and under different numbers of iterations. Each point shows the average of 100 cases and the error bar shows the standard deviation of the mean value computed by the central limit theorem. Different columns show LFS/IDT or $n = 1/n = 3$. Different rows show the different choices of our QoI. The first row shows the results where the target is the most and the second most sensitive reactions (all reactions were sorted in descending order by their impact number among $6 \times 6 \times 6$ condition grid points within design space), the second row for the second and the third most sensitive reactions, and the third row for the third and fourth most sensitive reactions. The information gain is defined in terms of the corresponding target reactions. From the results, we can clearly see that under all cases and all numbers of iterations, parameter-oriented design always shows the largest information gain. This is not beyond our expectation since the parameter-oriented design uses the expected posterior covariance volume of only QoI as the optimization objective, instead of the entire covariance matrix. This can avoid the mismatch between the conditions beneficial for the uncertainty reduction of the entire model and beneficial for only QoI. The two circles in Figure 5 show the two scenarios where the mismatch is very obvious so that the information gain of model-oriented design can be similar to, or even worse than random experiments.

This phenomenon can be explained by the sensitivity analysis under different conditions, which is shown in Figure S11 of the Supporting Information. For LFS, the condition with the largest sum of squares of impact numbers has very small impact numbers for reactions 2 and 3. Thus, the conditions chosen by model-oriented design show bad performance for the information gain of the second and third most sensitive reactions. For the target of the third and fourth sensitive reactions, the performance becomes better (similar to random experiments), since at these conditions with the largest sum of squares of impact numbers, the impact numbers for the fourth most sensitive reaction are pretty large compared with other conditions. Thus, the information gain at least can be comparable with random experiments. A similar phenomenon does not appear for the cases of IDT, which can also be explained by the sensitivity analysis of IDT shown in Figure S12 of the Supporting Information. The conditions with the largest sum of squares of impact numbers have large impact numbers for both the most and the third most sensitive reactions, while small sensitivity for the second and fourth most sensitive reactions. Thus, for the three targets here, none of the random designs are better than model-oriented experiments because in all three targets, at least one of the target reactions is sensitive under the optimal condition identified by the model-oriented design. However, we expect that if we set the target reactions as the second and the fourth most sensitive reactions, the model-oriented design would be worse than random experiments under $n = 1$ setting, which is verified by Figure S13 in the Supporting Information.

We then look at the evolution of uncertainty as the number of iterations increases for the case of LFS with $n = 1$, and the second and third most sensitive reactions are the target reactions. The $1 - \sigma$ and $2 - \sigma$ contours of the PDF for one case of 100 runnings are shown in Figure 6. We can clearly see

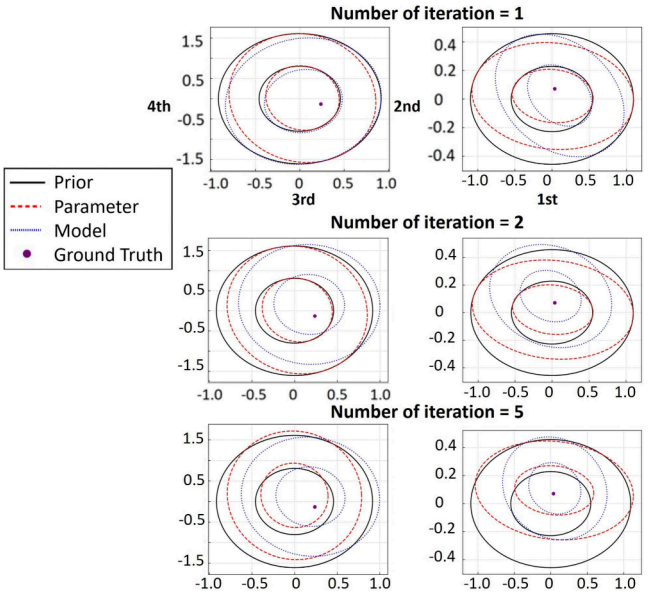


Figure 6. $1 - \sigma$ and $2 - \sigma$ PDF contours of four most sensitive reactions for prior (black) and posterior distributions after one, two, and five iterations done at conditions identified by parameter-oriented (red) and model-oriented (blue) design, in one of 100 cases for LFS with $n = 1$. Purple dots show the value of the "ground truth" for generating the noisy data. The QoI is the second and third sensitive reactions, corresponding to the horizontal axes of the left figures and the vertical axes of the right figures.

that after one experiment, the model-oriented design (blue contours) largely reduces the uncertainty for the most sensitive reaction, which is, unfortunately, not in our target reactions. The parameter-oriented design (red contours), however, can reduce the uncertainty of the second and the third most sensitive reactions more than the model-oriented design. After the second experiment, model-oriented design also reduce the second and third most sensitive reactions. This is because after reducing the most sensitive reaction, it is hard to further reduce it in the second experiment due to its decreased impact number, so the model-oriented design would reduce the uncertainty of the second, third, and even fourth most sensitive reactions. This explains why the model-oriented design surpasses random experiments and achieves a level a little lower than that of the parameter-oriented design at the second iteration, as shown in Figure 5. This also can explain why in the

https://doi.org/10.1021/acs.energyfuels.4c02299
Energy Fuels 2024, 38, 15630−15641

n = 3 case such phenomena are not observed, since within one iteration the three experiments identified by model-oriented design would reduce the uncertainty of several most sensitive reactions instead of just the first one. After five experiments, the posterior uncertainty for model-oriented design and parameter-oriented design looks very similar in Figure 6, which corresponds to the phenomenon that the difference of information gain for the two methods is reduced after the fifth experiment in Figure 5. Unsurprisingly, parameter-oriented design still outperforms model-oriented design.

3.2.2. Prediction-Oriented Design. In this subsection, we compared prediction-oriented design against model-oriented design on the performance of reducing the uncertainty of the prediction. The design space and prediction targets are shown in Table 5. Note that prediction targets cannot be reached within design space, so that the performance of identifying alternative conditions for unreachable targets can be assessed.

Table 5. Design Space and Prediction Targets for Prediction-Oriented Design Assessment

	P [atm]	T [K]	ϕ
design space of IDT	1−15	800−1200	0.5−2
prediction targets of IDT	30	1200−1600	5
design space of LFS	0.5−5	300−400	0.8−1.2
prediction targets of LFS	10	500	0.8−1.5

Figure 7 shows the average information gain in terms of the prediction uncertainty for both LFS and IDT with both n = 1

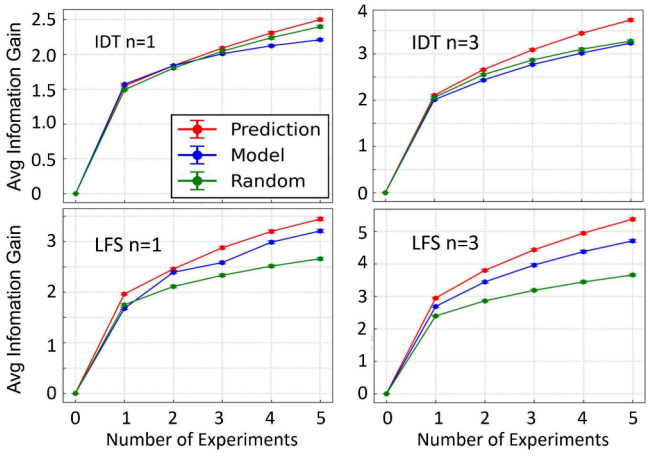


Figure 7. Average information gain of 100 runnings versus the number of iterations for prediction-oriented design (red), model-oriented design (blue), and random experiments (green), for both LFS and IDT and with both n = 1 and n = 3. The error bars show the 1 − σ bounds of the average computed by the central limit theorem.

(n is the number of experiments that are simultaneously designed, see eqs 10 and 11) and n = 3. In all the cases, the prediction-oriented design outperforms both model-oriented design and random experiments. It is worth noting that for IDT, the model-oriented design is worse than or similar to random experiments.

Figure 8 shows the prior uncertainty and posterior uncertainty obtained by different design methods, of different numbers of iterations, for LFS with n = 1. It is observed that after one experiment, the uncertainty of the lean side is similar for prediction- and model-oriented design, but on the rich side,

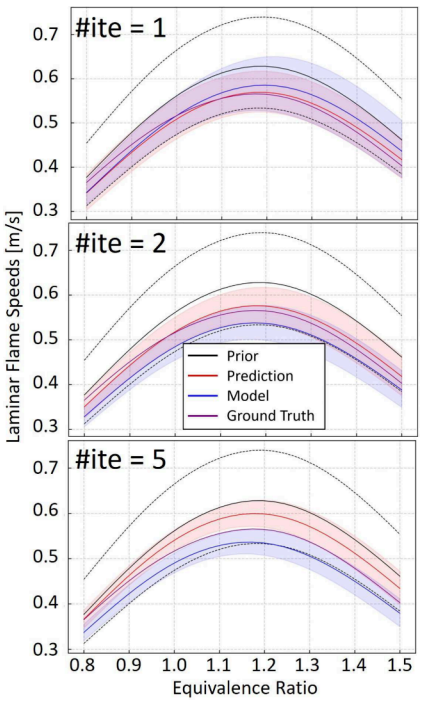


Figure 8. Prior and posterior 1-σ prediction uncertainty for LFS with n = 1 after one, two, and five iterations in one of 100 cases. Prior uncertainty is denoted by two dashed curves, while posteriors are denoted by shadows of different colors. The purple curve shows the ground truth prediction, while other solid lines show the mean values of prior and posterior models.

the prediction-oriented design obviously shows smaller uncertainty. It is worth noting that the design space is confined within ϕ = 0.8−1.2, so prediction-oriented design shows smaller uncertainty at conditions that cannot be accessed by experiments. After two experiments, on both the lean and rich sides, the uncertainty is similar for the two methods. After five experiments, the uncertainty of the two methods is still very similar in Figure 8. But from Figure 7 we can know that prediction-oriented design always outperforms model-oriented design.

Figure 9 shows the prediction uncertainty comparison for one case of IDT with n = 3. After one iteration (i.e., three experiments), the uncertainty of the prediction-oriented design shows slightly smaller uncertainty on both high-temperature and low-temperature ends (check the zoomed-in figures), but the difference is negligible. After five iterations (i.e., 15 experiments), we can see that the difference becomes larger.

4. CONCLUSION

In this paper, we built a fast QoI-oriented Bayesian experimental design (QBED) framework for the efficient uncertainty reduction of kinetic models. First, we proposed a fast parameter-oriented and a prediction-oriented information gain estimator based on the linearization of forward models. Theoretical analysis verified such a mismatch and provided insights for Bayesian DoE problems, while numerical experiments showed superior performance under all of the scenarios compared with the model-oriented design. Second, we combined the Laplacian approximation-based fast EIG estimator with the gradient ascent. The comparison with the enumeration-based method shows gradient ascent usually shows larger average information gain but runs slowly, while

https://doi.org/10.1021/acs.energyfuels.4c02299
Energy Fuels 2024, 38, 15630−15641

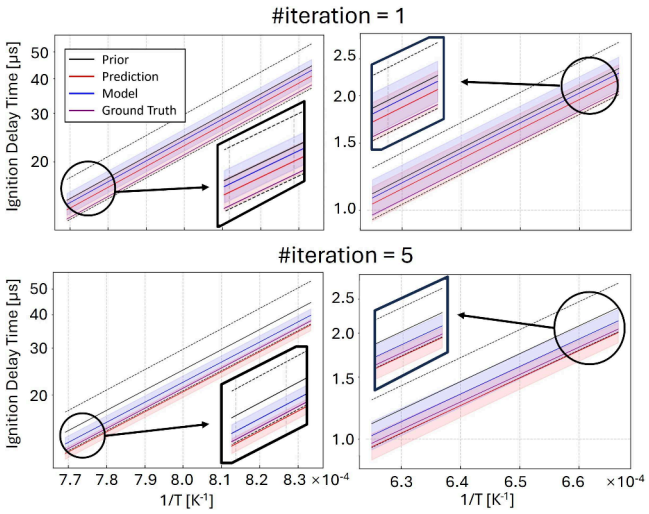


Figure 9. Prior and posterior $1-\sigma$ prediction uncertainty for IDT with $n = 3$ after one, two, and five iterations in one of the 100 cases. Prior uncertainty is denoted by two dashed curves, while posteriors are denoted by shadows of different colors. The purple curve shows the ground truth prediction, while other solid lines show the mean value of prior and posterior models. The left and right panels show the ranges of $T = 1200-1300$ and 1500−1600 K, respectively.

enumeration-baed methods usually show slightly suboptimal information gain but fast computational speed. Third, a single, unified response surface that can provide fast, differentiable predictions across the entire design space was built and largely accelerated the design algorithms. For high-dimensional cases, it renders gradient-based design optimization possible; for low-dimensional cases, it can make enumeration-based methods run faster utilizing the parallel evaluation of neural networks on GPU. Comprehensive numerical experiments with a large number of runnings were conducted to provide statistical evidence of the advantages of the proposed methods. The dynamical evolution of uncertainty as the number of experiments increased was unraveled and explained by the impact numbers. The proposed pipeline can finish one design in 0.5 s for the 3-D design space and in 1.6 s for the 9-D design space.

For experimentalists, this article provides a useful tool to design kinetic experiments for efficient uncertainty reduction. The entire pipeline was presented with detailed guidance. The code and user manual of the proposed framework are publicly available. For modelers, this paper can be a reference for utilizing advances in computer science, especially machine learning, to solve combustion problems. Even some commonly seen techniques, such as parallel evaluation of neural networks and autodifferentiation, can still be further exploited for suitable combustion tasks.

■ ASSOCIATED CONTENT

Data Availability Statement

The source code of QBED can be found in the repository: https://github.com/DENG-MIT/QoI-oriented-Bayesian-experimental-design.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.energyfuels.4c02299.

Supporting tables, figures, proofs, and details of the methods (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Sili Deng − *Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States;* orcid.org/0000-0002-3421-7414; Email: silideng@mit.edu

Authors

Huaibo Chen − *Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*

Qiaofeng Li − *Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.energyfuels.4c02299

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work is supported by the National Science Foundation (NSF) under Grant No. CBET-2143625. H.C. thanks Mr. Wendi Dong at Stanford University for helpful discussions.

■ REFERENCES

(1) Wang, H.; Sheen, D. A. Combustion kinetic model uncertainty quantification, propagation and minimization. *Prog. Energy Combust. Sci.* **2015**, *47*, 1−31.

(2) Tomlin, A. S. The role of sensitivity and uncertainty analysis in combustion modelling. *Proceedings of the Combustion Institute* **2013**, *34*, 159−176.

(3) Hanson, R.; Davidson, D. Recent advances in laser absorption and shock tube methods for studies of combustion chemistry. *Prog. Energy Combust. Sci.* **2014**, *44*, 103−114.

(4) Goldsborough, S. S.; Hochgreb, S.; Vanhove, G.; Wooldridge, M. S.; Curran, H. J.; Sung, C.-J. Advances in rapid compression machine studies of low-and intermediate-temperature autoignition phenomena. *Prog. Energy Combust. Sci.* **2017**, *63*, 1−78.

(5) Yang, B. Towards predictive combustion kinetic models: Progress in model analysis and informative experiments. *Proceedings of the Combustion Institute* **2021**, *38*, 199−222.

(6) Rainforth, T.; Foster, A.; Ivanova, D. R.; Bickford Smith, F. Modern Bayesian experimental design. *Statistical Science* **2024**, *39*, 100−114.

(7) Ryan, E. G.; Drovandi, C. C.; McGree, J. M.; Pettitt, A. N. A review of modern computational algorithms for Bayesian optimal design. *International Statistical Review* **2016**, *84*, 128−154.

(8) Turányi, T. Sensitivity analysis of complex kinetic systems. Tools and applications. *J. Math. Chem.* **1990**, *5*, 203−248.

(9) Russi, T.; Packard, A.; Feeley, R.; Frenklach, M. Sensitivity analysis of uncertainty in model prediction. *J. Phys. Chem. A* **2008**, *112*, 2579−2588.

(10) Mosbach, S.; Braumann, A.; Man, P. L.; Kastner, C. A.; Brownbridge, G. P.; Kraft, M. Iterative improvement of Bayesian parameter estimates for an engine model by means of experimental design. *Combust. Flame* **2012**, *159*, 1303−1313.

(11) Huan, X.; Marzouk, Y. M. Simulation-based optimal Bayesian experimental design for nonlinear systems. *J. Comput. Phys.* **2013**, *232*, 288−317.

(12) Vom Lehn, F.; Cai, L.; Pitsch, H. Iterative model-based experimental design for efficient uncertainty minimization of chemical mechanisms. *Proceedings of the Combustion Institute* **2021**, *38*, 1033−1042.

(13) Sheen, D. A.; Wang, H. The method of uncertainty quantification and minimization using polynomial chaos expansions. *Combust. Flame* **2011**, *158*, 2358−2374.
(14) Wang, S.; Ding, Y.; Miao, J.; Hanson, R. K. Shock tube and multi-species laser absorption measurements of rate constants for methanol pyrolysis. *Proceedings of the Combustion Institute* **2023**, *39*, 755−763.
(15) Sheen, D. A.; Manion, J. A. Kinetics of the reactions of H and CH3 radicals with n-butane: an experimental design study using reaction network analysis. *J. Phys. Chem. A* **2014**, *118*, 4929−4941.
(16) Li, S.; Tao, T.; Wang, J.; Yang, B.; Law, C. K.; Qi, F. Using sensitivity entropy in experimental design for uncertainty minimization of combustion kinetic models. *Proceedings of the Combustion Institute* **2017**, *36*, 709−716.
(17) Zhang, Y.; Dong, W.; Vandewalle, L. A.; Xu, R.; Smith, G. P.; Wang, H. Neural network approach to response surface development for reaction model optimization and uncertainty minimization. *Combust. Flame* **2023**, *251*, 112679.
(18) Feng, C.; Marzouk, Y. M. A layered multiple importance sampling scheme for focused optimal Bayesian experimental design, 2019. https://arxiv.org/abs/1903.11187 (accessed 2024−07−11).
(19) Attia, A.; Alexanderian, A.; Saibaba, A. K. Goal-oriented optimal design of experiments for large-scale Bayesian linear inverse problems. *Inverse Problems* **2018**, *34*, 095009.
(20) Smith, F. B.; Kirsch, A.; Farquhar, S.; Gal, Y.; Foster, A.; Rainforth, T. Prediction-Oriented Bayesian Active Learning. *International Conference on Artificial Intelligence and Statistics* **2023**, 7331−7348.
(21) Wang, J.; Li, S.; Yang, B. Combustion kinetic model development using surrogate model similarity method. *Combustion Theory and Modelling* **2018**, *22*, 777−794.
(22) Lin, K.; Zhou, Z.; Wang, Y.; Law, C. K.; Yang, B. Using active subspace-based similarity analysis for design of combustion experiments. *Proceedings of the Combustion Institute* **2023**, *39*, 5177−5186.
(23) Huan, X.; Marzouk, Y. Gradient-based stochastic optimization methods in Bayesian experimental design. *International Journal for Uncertainty Quantification* **2014**, *4*, 479−510.
(24) Foster, A.; Jankowiak, M.; O'Meara, M.; Teh, Y. W.; Rainforth, T. A unified stochastic gradient approach to designing bayesian-optimal experiments. *International Conference on Artificial Intelligence and Statistics* **2020**, 2959−2969.
(25) Goda, T.; Hironaka, T.; Kitade, W.; Foster, A. Unbiased MLMC stochastic gradient-based optimization of Bayesian experimental designs. *SIAM Journal on Scientific Computing* **2022**, *44*, A286−A311.
(26) Carlon, A. G.; Dia, B. M.; Espath, L.; Lopez, R. H.; Tempone, R. Nesterov-aided stochastic gradient methods using Laplace approximation for Bayesian design optimization. *Computer Methods in Applied Mechanics and Engineering* **2020**, *363*, 112909.
(27) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems* **2019**, *32*, 8026−8037.
(28) Oh, J.-H.; Wiersema, P.; Kim, K.; Mayhew, E.; Temme, J.; Kweon, C.-B.; Lee, T. Fast uncertainty reduction of chemical kinetic models with complex spaces using hybrid response-surface networks. *Combust. Flame* **2023**, *253*, 112772.
(29) Foster, A. E.Variational, Monte Carlo and policy-based approaches to Bayesian experimental design. *Ph.D. thesis*, University of Oxford, 2021.
(30) Rainforth, T.; Cornish, R.; Yang, H.; Warrington, A.; Wood, F. On nesting monte carlo estimators. *International Conference on Machine Learning* **2018**, 4267−4276.
(31) Bishop, C. M.; Nasrabadi, N. M. *Pattern Recognition and Machine Learning*; Springer, 2006.
(32) Chaloner, K.; Verdinelli, I. Bayesian experimental design: A review. *Statistical Science* **1995**, *10*, 273−304.
(33) Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization, 2014. https://arxiv.org/abs/1412.6980 (accessed 2024−07−11).
(34) Chaloner, K. Optimal Bayesian Experimental Design for Linear Models. *Annals of Statistics* **1984**, *12*, 283−300.
(35) Tao, Y.; Smith, G. P.; Wang, H. Critical kinetic uncertainties in modeling hydrogen/carbon monoxide, methane, methanol, formaldehyde, and ethylene combustion. *Combust. Flame* **2018**, *195*, 18−29.
(36) Zhang, X.; Wang, G.; Zou, J.; Li, Y.; Li, W.; Li, T.; Jin, H.; Zhou, Z.; Lee, Y.-Y. Investigation on the oxidation chemistry of methanol in laminar premixed flames. *Combust. Flame* **2017**, *180*, 20−31.
(37) Zhou, Z.; Lin, K.; Wang, Y.; Wang, J.; Law, C. K.; Yang, B. OptEx: An integrated framework for experimental design and combustion kinetic model optimization. *Combust. Flame* **2022**, *245*, 112298.
(38) Smith, G. P.; Tao, Y.; Wang, H. Foundational fuel chemistry model version 1.0 (FFCM-1), 2016. https://web.stanford.edu/group/haiwanglab/FFCM1/pages/FFCM1.html (accessed 2024−07−11).
(39) McKay, M. D.; Beckman, R. J.; Conover, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **2000**, *42*, 55−61.
(40) Goodwin, D. G., Speth, R. L., Moffat, H. K., Weber, B. W. Cantera: An Object-oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes. Version 2.5.1, 2021. https://www.cantera.org (accessed 2024−07−11).
(41) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2016**, 770−778.
(42) Elfwing, S.; Uchibe, E.; Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural networks* **2018**, *107*, 3−11.
(43) Shen, W.; Huan, X. Bayesian sequential optimal experimental design for nonlinear models using policy gradient reinforcement learning. *Computer Methods in Applied Mechanics and Engineering* **2023**, *416*, 116304.
(44) Chen, H. Bayesian Inference and Experimental Design of Combustion Kinetic Models. *Master Thesis*, Massachusetts Institute of Technology, 2024.

https://doi.org/10.1021/acs.energyfuels.4c02299
Energy Fuels 2024, 38, 15630−15641