
NEURAL MESSAGE PASSING FOR OBJECTIVE-BASED UNCERTAINTY QUANTIFICATION AND OPTIMAL EXPERIMENTAL DESIGN

Qihua Chen

University of Science and Technology of China
Hefei, Anhui, CN 230027
cqh@mail.ustc.edu.cn

Xuejin Chen

University of Science and technology of China
Hefei, Anhui, CN 230027
xjchen99@ustc.edu.cn

Hyun-Myung Woo*

Incheon National University
Incheon 22012, South Korea
hmwoo@inu.ac.kr

Byung-Jun Yoon*

Brookhaven National Laboratory
Upton, NY 11973, USA
Texas A&M University
College Station, TX 77843, USA
bjyoon@ece.tamu.edu

ABSTRACT

Various real-world scientific applications involve the mathematical modeling of complex uncertain systems with numerous unknown parameters. Accurate parameter estimation is often practically infeasible in such systems, as the available training data may be insufficient and the cost of acquiring additional data may be high. In such cases, based on a Bayesian paradigm, we can design robust operators retaining the best overall performance across all possible models and design optimal experiments that can effectively reduce uncertainty to enhance the performance of such operators maximally. While objective-based uncertainty quantification (objective-UQ) based on MOCU (mean objective cost of uncertainty) provides an effective means for quantifying uncertainty in complex systems, the high computational cost of estimating MOCU has been a challenge in applying it to real-world scientific/engineering problems. In this work, we propose a novel scheme to reduce the computational cost for objective-UQ via MOCU based on a data-driven approach. We adopt a neural message-passing model for surrogate modeling, incorporating a novel axiomatic constraint loss that penalizes an increase in the estimated system uncertainty. As an illustrative example, we consider the optimal experimental design (OED) problem for uncertain Kuramoto models, where the goal is to predict the experiments that can most effectively enhance robust synchronization performance through uncertainty reduction. We show that our proposed approach can accelerate MOCU-based OED by four to five orders of magnitude, without any visible performance loss compared to the state-of-the-art. The proposed approach applies to general OED tasks, beyond the Kuramoto model.

Keywords Kuramoto model · mean objective cost of uncertainty · message-passing neural network · uncertainty quantification · optimal experimental design (OED) · OED acceleration

1 Introduction

Real-world scientific and engineering applications often involve modeling complex uncertain systems with substantial uncertainties. For such systems, the amount of data available for learning the corresponding models may pale when compared to the model complexity, which may render accurate estimation of the model parameters practically infeasible. In the presence of significant model uncertainty, designing operators—such as predictors or controllers—based on the best

*Corresponding authors

(i.e., most likely) model may be unreliable. This issue may be alleviated by taking a Bayesian approach to consider model uncertainty to design a robust operator that can guarantee good average performance despite the uncertainty. Generally, robust operators have the advantage of preventing catastrophic failures caused by model mismatches and maintaining acceptable performances under model uncertainty. However, the overall performance will take a hit—at least to some extent, even if the expected degradation is acceptable—when compared to the optimal operator, whose design requires full knowledge of the true underlying model.

In this context, optimal experimental design (OED) aims to design, prioritize, or select experiments whose outcome is expected to effectively reduce the uncertainty in the current model such that the performance of the robust operator designed thereupon can be optimally improved. In general, an OED method first quantifies the uncertainty of the current model. Then, the expected efficacy of each potential experiment in the experimental design space is predicted, which can be used to intelligently select the best experiment that is expected to maximally reduce the model uncertainty that critically impacts the performance of the robust operator. Once the optimal experiment is designed (or selected), one may conduct the experiment at hand and use the actual experimental outcome to reduce model uncertainty. Within the available experimental budget, this OED cycle—which consists of *designing an experiment, conducting the designed experiment, and updating the model based on the experimental outcome* (and thereby reducing model uncertainty)—may be repeated. It should be noted that the experiment designed by an OED scheme may not necessarily be optimal, as the optimality of the experiment depends on the true model, which is *unknown*. However, *on average*, experiments designed via OED can reduce model uncertainty much more rapidly compared to experiments that are either randomly selected or designed in a heuristic manner. Consequently, OED can minimize the experimental cost and time required to reduce the model uncertainty—or equivalently, improve the performance of the robust operator designed based on an uncertain model—to the desired level.

In the context of OED, one fundamental question emerges naturally: how can one quantify the model uncertainty? In Bayesian OED, the model uncertainty has been quantified via variance of the posterior distribution of the parameters. To predict the potential efficacy of each experiment, the so-called *expected information gain (EIG)* has been widely used [1, 2, 3, 4] and has been shown to be effective, especially, when the objective is better identification or more accurate modeling of the system of interest. However, in real-world scientific applications where one may want to learn models for specific tasks or operations—e.g., control or classification—the aforementioned approaches may not always be effective as the reduced uncertainty may not directly help perform the tasks at hand or achieve the operational goals.

The mean objective cost of uncertainty (MOCU) [5, 6] addresses this shortcoming by quantifying the expected increase in the operational cost that is induced by model uncertainty, providing an effective means for objective-based uncertainty quantification (objective-UQ). As MOCU captures the model uncertainty that impacts the performance of the operator designed based on the model, it leads to an effective framework for optimal experimental design (OED) [7, 8], based on which the expected efficacy of potential experiments may be quantified. To date, MOCU-based OED has been widely applied to a number of applications [8, 9, 10, 11, 12, 13] [14, 15]. For example, MOCU-based OED was applied to pool-based Bayesian active-learning problems, where the objective is to improve the sample labeling efficiency [13, 14, 15]. MOCU-based acquisition functions—such as weighted MOCU [13] or soft MOCU [14, 15]—were defined, which were then used to identify and label the optimal sample expected to maximally reduce the classification error. In this setup, an experiment corresponds to selecting an unlabeled data point from a given pool and the experimental outcome is the true label of the selected data point. Furthermore, MOCU-based UQ and OED were also used to investigate the problem of designing a robust filter that can reliably predict the drug concentration level via uncertain stochastic differential equation (SDE)-based pharmacokinetic models [12]. In the study, the experimental design concerned selecting the best parameter to query for further measurement. The outcome of the optimal experiment—i.e., the additional measurement data of the selected parameter—is aimed at maximally reducing the prediction error of the robust filter.

Recently, the MOCU-based OED strategy has been developed for uncertain Kuramoto models [7], which consist of interconnected oscillators described by coupled ordinary differential equations (ODEs). Kuramoto models [16] have been extensively studied in various fields, thanks to their ability to simulate interesting collective behavior in complex networked systems such as the robustness of microgrid networks [17] or synchrony in brain networks [18, 19]. Given an uncertain Kuramoto model, the objective of the experimental design was to predict the optimal experiment that can most effectively reduce the control cost to achieve robust synchronization of the uncertain Kuramoto oscillators.

One practical challenge of MOCU-based OED is the high computational complexity that mainly originates from the complexity of predicting the expected efficacy (i.e., the expected remaining MOCU) of each experiment in the experimental space. In general, real-world scientific/engineering applications are designed based on highly complex operations, resulting in no closed-form expression for (expected remaining) MOCU. For example, in the original OED study for the robust control of uncertain Kuramoto models [7], for a given experiment, the expected remaining MOCU was estimated via Monte Carlo sampling [20], where each sample defining a Kuramoto model involves repetitively solving coupled differential equations to compute the operational cost of the best operator that is optimal for the

given Kuramoto model. In order to reduce the effective time complexity, parallelism was considered by computing the optimal operational cost of each Kuramoto model on different processing units. However, even with parallelism, MOCU estimation remained highly complex, limiting the applicability and scalability of the MOCU-based OED. In order to alleviate the computational bottleneck, a machine learning (ML)-based approach was proposed in [21]. In this work, a fully connected neural network model was trained to predict—without solving the ODE—whether or not a given Kuramoto model reaches frequency synchronization. This approach remarkably reduced the cost of predicting the optimal operator, thereby significantly accelerating the MOCU computation and expediting the MOCU-based OED by two orders of magnitude without any degradation in OED performance. However, the approach in [21] still resorted to Monte Carlo sampling for MOCU estimation, and it remained an open question whether it is possible to design accurate deep surrogate models for direct MOCU estimation by building on recent advances in deep learning (DL) techniques [22, 23, 24]. Such deep surrogate models would obviate the need for costly Monte Carlo sampling and further accelerate MOCU-based OED.

In this work, we address this open problem by proposing the use of a deep learning technique to build an effective surrogate model for MOCU computation for the first time. We show that our proposed scheme is capable of accelerating MOCU-based OED by four to five orders of magnitude by eliminating the Monte Carlo sampling process and the need for repetitively solving ordinary differential equations for the uncertain Kuramoto models. Our main contributions are as follows. First, we propose a novel sampling-free approach that can effectively address the computational bottleneck of MOCU estimation by directly estimating MOCU for a given uncertainty class via neural message-passing [22]. Second, we introduce a novel axiomatic constraint loss that improves the estimation accuracy of the message-passing model to enhance the experimental design. Experiments show that our proposed approach accelerates OED up to five orders of magnitude while maintaining virtually identical performance compared to the original sampling-based approach [7]. To the best of our knowledge, this is the first work that demonstrates the possibility of designing data-driven surrogate models for efficient experimental design based on objective-UQ via MOCU, thereby laying the foundation for a wide range of real-world applications of MOCU-based OED.

The current paper is organized as follows. In Section 2, we provide the details of MOCU-based OED for uncertain Kuramoto models, where the operational goal is to design an effective control strategy with minimum cost to achieve robust synchronization, which was originally formulated in [7]. We also analyze the time complexity of MOCU estimation using the sampling-based method in [7]. In Section 3, we propose a DL approach for quantifying model uncertainty via MOCU. Specifically, we build a DL surrogate model using a message-passing neural network (MPNN) that directly quantifies MOCU, which then can be used for designing effective experiments that can enhance the performance and reduce the cost of control for uncertain Kuramoto models. We demonstrate the efficacy of the proposed approach in Section 4 based on comprehensive simulation results. We conclude the paper in Section 5 with important insights learned from this study, discussion of the limitations of the current work, and future research directions.

2 Optimal Experimental Design for Uncertain Kuramoto model

In this section, we present a brief review of the MOCU-based OED scheme for the robust synchronization of uncertain Kuramoto models, originally considered in [7]. First, we describe the uncertainty class of Kuramoto models. Based on this, we outline the OED setup, where the objective is the robust synchronization of an uncertain Kuramoto model. Next, we describe how one can prioritize the potential experiments based on the (expected remaining) MOCU. We then present how MOCU can be estimated via sampling, as proposed in a prior study [7] and analyze the computational complexity of the sampling-based MOCU estimation and experimental design. As the primary goal of the current study is to accelerate MOCU-based UQ and OED by leveraging a DL model, the latest sampling-based approach in [7] serves as the baseline to which we compare our proposed approach.

2.1 Robust synchronization of uncertain Kuramoto model

The Kuramoto model consists of N interacting oscillators described by N coupled ordinary differential equations:

$$\dot{\theta}_i(t) = \omega_i + \sum_{j=1}^N a_{i,j} \sin(\theta_j(t) - \theta_i(t)), \quad i = 1, 2, \dots, N, \quad (1)$$

where N is the number of oscillators, $\theta_i(t)$ is the instantaneous phase of the i th oscillator with the intrinsic natural frequency ω_i , and $a_{i,j}$ ($= a_{j,i}$) represents the coupling strength between the i th and j th oscillators. The model has been widely used to describe and investigate the collective behavior of constituents of a complex networked system, where examples include brain networks [25, 26, 27], power systems [28], microgrids [17], and synchronously flashing fireflies [29]. Of special interest has been the synchronization phenomena of Kuramoto models, which have been

shown to be related to emergent system-level properties—such as the stability of microgrid networks [17] or neurological disorders in brain networks [18, 19].

In a recent study [7], the optimal experimental design problem for uncertain Kuramoto models was investigated. The task of interest was to achieve global synchronization of a given Kuramoto model through external control in the presence of model uncertainty, where MOCU-based OED was used to select experiments that can effectively reduce the uncertainty that affects this goal. In our current study, we also consider this robust global synchronization problem, in order to demonstrate how a DL-based approach can significantly improve the computational efficiency of OED without degrading the overall efficacy.

As in [7], we assume that the coupling strength $a_{i,j}, i, j = 1, 2, \dots, N$, is uncertain and known only up to a range $a_{i,j} \in [a_{i,j}^{\ell}, a_{i,j}^u]$. This gives rise to an uncertainty class $\mathcal{A}$, which consists of all possible parameter vector, $\mathbf{a} = [a_{1,2}, a_{1,3}, \dots, a_{N-1,N}]^T \in \mathcal{A}$, distributed according to a prior distribution $P_{\mathcal{A}}(\mathbf{a})$ consistent with the given constraints. Given a Kuramoto model with uncertain coupling strength $\mathbf{a} \in \mathcal{A}$, our operational objective is to apply robust control to ensure that the oscillators are frequency synchronized such that

$$\lim_{t \rightarrow \infty} |\dot{\theta}_i(t) - \dot{\theta}_j(t)| = 0, \quad \forall i, j. \quad (2)$$

Specifically, we consider the control scheme in [7], where a control oscillator with natural frequency ω_{N+1} is added to the uncertain Kuramoto model to interact with the N oscillators with uniform interaction strength a_{N+1} (i.e., $a_{N+1} = a_{N+1,i}, \forall i$). The extended model with the control oscillator is given by:

$$\dot{\theta}_i(t) = \omega_i + \sum_{j=1}^N a_{i,j} \sin(\theta_j(t) - \theta_i(t)) + a_{N+1} \sin(\theta_{N+1}(t) - \theta_i(t)), \quad i = 1, 2, \dots, N+1. \quad (3)$$

For a Kuramoto model with a given interaction strength $\mathbf{a}$, we define $\xi(\mathbf{a})$ as the minimum (non-negative) interaction strength that ensures the global synchronization of the extended Kuramoto model when the interaction strength a_N of the control oscillator is set to $a_{N+1} = \xi(\mathbf{a})$. We also regard $\xi(\mathbf{a})$ as the cost of control since exerting stronger control would require higher power in practice.

In the presence of uncertainty $\mathcal{A}$, the cost of *robust* control that guarantees model synchronization would inevitably increase as a_{N+1} needs to be large enough to ensure synchronization for all possible models in the uncertainty class. For this reason, the optimal robust control (and the cost thereof) is defined as follows [7]:

$$\xi^*(\mathcal{A}) = \max_{\mathbf{a} \in \mathcal{A}} \xi(\mathbf{a}). \quad (4)$$

MOCU [5, 6] captures the expected differential cost induced by model uncertainty due to having to use the optimal *robust* control (that maintains good average control performance across the uncertainty class $\mathcal{A}$) instead of the optimal control that is model-specific (which cannot be used in practice as the true model is unknown) as follows:

$$M(\mathcal{A}) = \mathbb{E}_{\mathbf{a} \in \mathcal{A}} [\xi^*(\mathcal{A}) - \xi(\mathbf{a})]. \quad (5)$$

2.2 Experimental design space

In this study, we consider the same experimental design space as in [7]. Each experiment in this design space consists of the following steps. First, we select a pair of oscillators (i, j) in the original Kuramoto model that consists of N oscillators and isolate the pair from the $N - 2$ remaining oscillators. Next, we observe whether or not the isolated oscillator pair (i, j) gets spontaneously synchronized without external control. According to the binary experimental outcome (i.e., whether the oscillator pair gets synchronized or not), we can reduce the model uncertainty by utilizing Theorem 1 in [7] (shown below as a reference for readers):

Theorem 1 *Assume the Kuramoto model of two-oscillators:*

$$\begin{aligned} \dot{\theta}_1(t) &= \omega_1 + a \sin(\theta_2(t) - \theta_1(t)), \\ \dot{\theta}_2(t) &= \omega_2 + a \sin(\theta_1(t) - \theta_2(t)), \end{aligned} \quad (6)$$

*with initial phases $\theta_1(0), \theta_2(0) \in [0, 2\pi)$. For any solutions $\theta_1(t)$ and $\theta_2(t)$ to (6), there holds $|\dot{\theta}_1(t) - \dot{\theta}_2(t)| \rightarrow 0$ as $t \rightarrow \infty$ if and only if $(1/2) * |\omega_1 - \omega_2| \leq a$. ■*

Based on Theorem 1 [7], the model uncertainty can be reduced by increasing the lower bound of the uncertain interval $a_{i,j}^{\ell}$ to $\max((1/2) * |\omega_i - \omega_j|, a_{i,j}^{\ell})$ if the oscillator pair gets synchronized. Otherwise, the upper bound $a_{i,j}^u$ is reduced to $\min((1/2) * |\omega_i - \omega_j|, a_{i,j}^u)$.

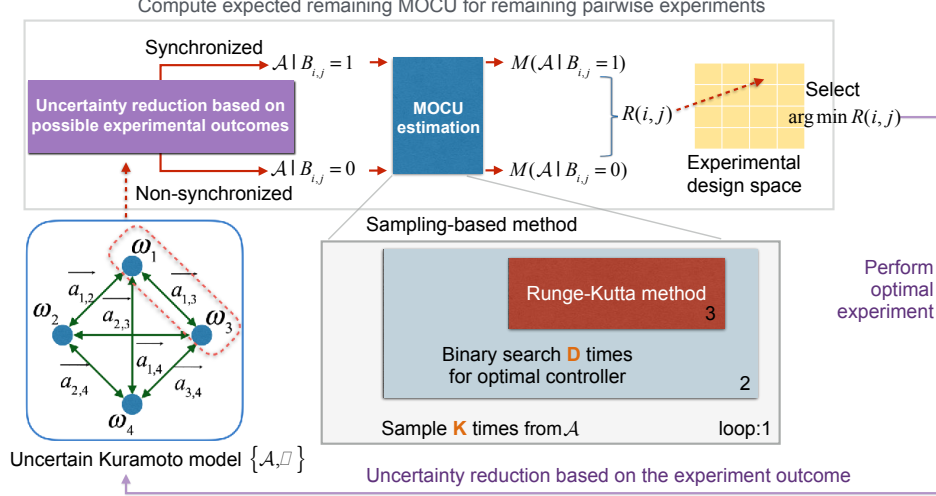


Figure 1: Overview of the MOCU-based optimal experimental design for robust synchronization of uncertain Kuramoto models, where MOCU is estimated based on the sampling-based approach [7].

2.3 MOCU-based optimal experimental design

Figure 1 provides an overview of the MOCU-based OED scheme, where MOCU is computed using a Monte Carlo estimator [7]. The MOCU-based OED provides a systematic means to identify the optimal experiment that is expected to most sharply reduce the uncertainty affecting the cost of the operation of interest—namely, the cost of achieving robust synchronization through control. To prioritize experiments, we first compute the *expected remaining MOCU* for every pairwise synchronization experiment, which is given by:

$$R(i, j) = \mathbb{E}_{B_{i,j}} [M(\mathcal{A} | B_{i,j})] = \sum_{b \in \{0,1\}} P(B_{i,j} = b) M(\mathcal{A} | B_{i,j} = b), \quad (7)$$

where $\mathcal{A} | B_{i,j}$ is the remaining uncertainty conditioned on the experimental outcome $B_{i,j}$. The probability $P(B_{i,j} = b)$ is given by:

$$P(B_{i,j} = 1) = \frac{a_{i,j}^u - \tilde{a}_{i,j}}{a_{i,j}^u - a_{i,j}^\ell}, \quad P(B_{i,j} = 0) = \frac{\tilde{a}_{i,j} - a_{i,j}^\ell}{a_{i,j}^u - a_{i,j}^\ell}, \quad (8)$$

where $\tilde{a}_{i,j} = \min(\max((1/2) * |\omega_i - \omega_j|, a_{i,j}^\ell), a_{i,j}^u)$. Here we assume that the uncertain interaction strength $a_{i,j}$ is uniformly distributed between the bounds $a_{i,j}^\ell$ and $a_{i,j}^u$ and that the experimental outcome $B_{i,j}$ is used to update either the lower or upper bound according to Theorem 1 as described in Sec. 2.2. The pairwise synchronization experiment for the oscillator pair (i, j) yielding the minimum expected remaining MOCU $R(i, j)$ is selected as the optimal experiment (i^*, j^*) :

$$(i^*, j^*) = \arg \min_{1 \leq i < j \leq N} R(i, j). \quad (9)$$

Ideally, as described in Algorithm 1, the expected remaining MOCU should be re-estimated after each experiment, based on the reduced uncertainty class updated according to the outcome of the previous experiment. However, since this increases the overall computational cost of the experimental design procedure, one may take a simpler strategy if desired, where all experiments are prioritized based on their respective expected remaining MOCU values estimated based on the initial uncertainty class. Although this approach is clearly sub-optimal from a theoretical point of view, the strategy has been shown to lead to virtually identical performance compared to the optimal strategy in a number of examples considered in previous studies [7, 21].

Algorithm 1 MOCU-based optimal experimental design (OED) for robust synchronization of uncertain Kuramoto model with N oscillators

Input: Uncertain Kuramoto model $\{N, \mathcal{A}, \omega\}$
 experimental design space $\mathcal{E} = \{(1, 2), \dots, (N-1, N)\}$

Output: Reduced uncertainty class $\mathcal{A}$

```

while  $\mathcal{E}$  is not empty do
  /* Compute expected remaining MOCU */
  for  $(i, j)$  in  $\mathcal{E}$  do
     $\tilde{a}_{i,j} \leftarrow \min \left( \max \left( \frac{1}{2} |\omega_i - \omega_j|, a_{i,j}^\ell \right), a_{i,j}^u \right)$ 
     $P(B_{i,j} = 1) \leftarrow \frac{a_{i,j}^u - \tilde{a}_{i,j}}{a_{i,j}^u - a_{i,j}^\ell}$ 
     $P(B_{i,j} = 0) \leftarrow \frac{\tilde{a}_{i,j} - a_{i,j}^\ell}{a_{i,j}^u - a_{i,j}^\ell}$ 
     $R(i, j) \leftarrow \sum_{b \in \{0,1\}} P(B_{i,j} = b) M(\mathcal{A} | B_{ij} = b)$ 
  end for
  /* Select and conduct the optimal experiment */
   $(i^*, j^*) \leftarrow \arg \min_{(i,j) \in \mathcal{E}} R(i, j)$ 
  Conduct the optimal experiment  $(i^*, j^*)$ 
  if  $(i^*, j^*)$  gets synchronized then
     $\mathcal{A} \ni a_{i^*,j^*}^\ell \leftarrow \tilde{a}_{i^*,j^*}$ 
  else
     $\mathcal{A} \ni a_{i^*,j^*}^u \leftarrow \tilde{a}_{i^*,j^*}$ 
  end if
  remove  $(i^*, j^*)$  from  $\mathcal{E}$ 
end while
return  $\mathcal{A}$ 

```

2.4 Computational complexity

One major obstacle in applying MOCU-based OED to real-world applications has been the high computational cost of computing the expected remaining MOCU for the experiments under consideration. In many cases, there may be no closed-form expression of MOCU, and it may have to be numerically estimated as in [7]. The sampling-based approach [7] for MOCU estimation involves repeated sampling from the uncertainty class (to obtain a sample estimate of the expected cost increase) as well as a binary search to find the optimal operational cost.

The sampling-based MOCU estimation process is summarized in Algorithm 2. Given an uncertainty class $\mathcal{A}$, K sample points are randomly drawn from $P(\mathcal{A})$, where each point corresponds to one possible Kuramoto model within the given uncertainty class. Generally, as the sample size K increases, the accuracy of the estimated MOCU improves. Finding the optimal control (and its cost) for each sample point (i.e., a specific Kuramoto model in $\mathcal{A}$) requires numerical optimization. In [7], the optimal control was predicted via binary search. As a consequence, the overall computational complexity for sampling-based MOCU estimation is given by $\mathcal{O}(KDMN^2)$, where D is the average complexity of the binary search, and M is the time duration for solving the ordinary differential equations (e.g., a fourth-order Runge-Kutta method was used in [7]), and N is the number of oscillators in the original Kuramoto model. Furthermore, to estimate the expected remaining MOCU for a given experiment, we need to estimate the MOCU conditioned on every possible experimental outcome (which is binary in this case). Finally, this process needs to be repeated for all $\binom{N}{2}$ pairwise experiments to be able to identify the optimal experiment with the smallest $R(i, j)$.

As described in the previous section, this entire process—estimating the expected remaining MOCU for all (remaining) experiments in the experimental design space, identifying the optimal experiment, conducting the selected experiment, and reducing the uncertainty class according to the experimental outcome—may be repeated until we have performed all possible experiments or until we have exhausted the experimental budget. This will increase the computational complexity of OED even further.

Algorithm 2 Sampling-based MOCU computation**Input:** Uncertain Kuramoto model $\{N, \mathcal{A}, \omega\}$, natural frequency of controlling oscillator $\bar{\omega}$ **Output:** $M(\mathcal{A})$

```

for  $i = 1 \rightarrow K$ 
  draw a sample  $\mathbf{a}_i$  from  $P(\mathcal{A})$ 
  /* Initialize binary search space */
   $a_{N+1}^u \leftarrow 2$ 
  while  $\{\mathbf{a}_i, a_{N+1}^u, \omega, \bar{\omega}\}$  is not synchronized do
     $a_{N+1}^u \leftarrow a_{N+1}^u + 2$ 
  end while
   $a_{N+1}^l \leftarrow 0$ 
  /* Binary search for  $\xi(a_i)$  */
  while  $a_{N+1}^u - a_{N+1}^l < 2.5e^{-4}$  do
     $c \leftarrow \frac{a_{N+1}^u + a_{N+1}^l}{2}$ 
     $state \leftarrow RungeKuttaSolver(\mathbf{a}_i, c, \omega, \bar{\omega})$ 
    if  $state$  is synchronize then
       $a_{N+1}^u \leftarrow c$ 
    else
       $a_{N+1}^l \leftarrow c$ 
    end if
  end while
   $\xi(\mathbf{a}_i) \leftarrow c$ 
end for
return  $\sum_{i=1}^K \frac{1}{K} (\max_{1 \leq j \leq K} \xi(\mathbf{a}_j) - \xi(\mathbf{a}_i))$ 

```

3 Message-Passing Neural Network for Uncertainty Quantification

In this section, we propose a data-driven approach for accelerating objective-UQ, thereby significantly expediting MOCU-based OED. At the heart of the proposed method lies a message-passing neural network (MPNN) [22] that incorporates a novel axiomatic constraint. MPNN is a graph neural network-based regression model that can predict the output by repeatedly exchanging messages between neighboring nodes that are connected in the input undirected weighted graph. Similar architectures have been shown to achieve superior performance over other existing schemes in molecular analysis [22, 30], recommender systems [31], and visual scene understanding [32, 33]. In what follows, we first describe the MPNN architecture used to directly quantify (expected remaining) MOCU for an uncertain Kuramoto model. Next, we propose a novel constraint function that penalizes the deviation from an axiom regarding MOCU. Finally, we present the training procedure of the proposed MPNN model.

3.1 Model architecture

Figure 2 illustrates the overall architecture of the MPNN model that takes an uncertain Kuramoto model, represented as $\{\{\omega_i\}_{i=1}^N, \mathcal{A}\}$, as its *input* and predicts the MOCU for the given uncertain model as the *output*. Specifically, we set the initial node features x_v and edge features e_{vw} of the MPNN model to the natural frequency of the oscillators and the lower and upper bounds of the interaction strength of the uncertain Kuramoto model, respectively. The model operates in two phases: a *message-passing* phase and a *readout* phase. The message-passing phase recurrently interchanges information across the graph to build a node-level neural representation of the input uncertain Kuramoto model. The readout phase aggregates the node-level representation of the Kuramoto model to a graph-level representation and makes a prediction. The message-passing phase runs for T repeated steps with message function M_t and vertex update function U_t , where $T = 3$ was used in this study. During the message-passing phase, for each vertex v , a message m_v^t is aggregated from the neighbors through the edges, and the vertex hidden states h_v^t is then updated based on m_v^t as follows:

$$m_v^{t+1} = \sum_{w \in N(v)} M_t(h_v^t, h_w^t, e_{vw}) \quad (10)$$

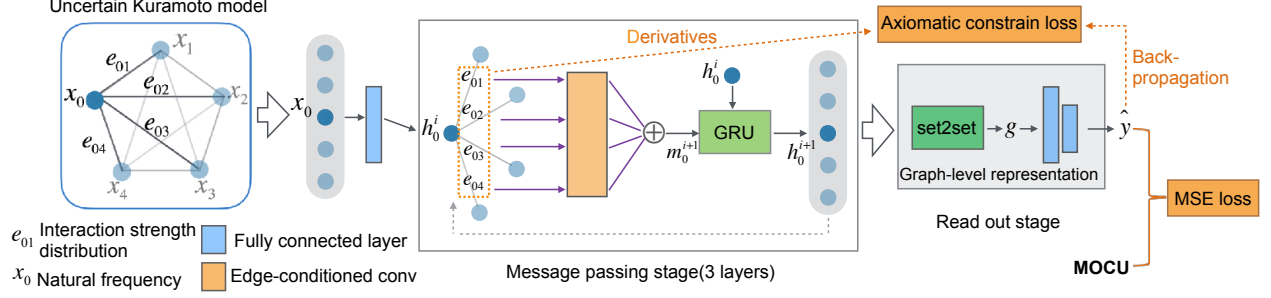


Figure 2: Neural message-passing architecture for uncertainty quantification of the uncertain Kuramoto model.

$$h_v^{t+1} = U_t(h_v^t, m_v^{t+1}), \quad (11)$$

where $N(v)$ is the set of neighbors of v in the graph. h_v^0 is initialized as a learnable linear mapping of the initial node features x_v followed by the ReLU (Rectified Linear Unit) activation. Here, the message function M_t is an Edge-Conditioned Convolution [34], a learnable network that dynamically generates filters based on the edge features. We employ a Gated Recurrent Unit (GRU) [35] as the update function U_t which applies gating mechanisms to manage the flow of information.

In the readout phase, the final prediction is made by applying the readout function R based on the final hidden states h_v^T of all the vertices according to

$$\hat{y} = R(\{h_v^T | v \in G\}), \quad (12)$$

where R consists of a set2set model [36] which generates a graph-level representation g that is invariant to the permutation of input node features. Then, the vector g goes through two fully connected layers to obtain the final prediction $\hat{y}$.

3.2 Learning objective

In the context of MOCU-based OED, the trained MPNN model serves as a surrogate model that predicts the expected remaining MOCU of each experiment. The predicted MOCU values are then used to prioritize the experiments in the experimental design space based on their expected efficacy and select the optimal experiment as described in (9). For this reason, from the perspective of OED, it is necessary for the trained model not only to predict the expected remaining MOCU accurately but also to maintain the relative order of the possible experiments when they are prioritized according to their respective expected remaining MOCU values. In order to enhance the capability of the trained model in preserving this order, we define the overall loss function by combining the mean squared error (MSE) loss ℓ_{MSE} with a novel axiomatic constraint loss ℓ_{AC} as follows:

$$L(\mathcal{A}, y, \hat{y}) = \ell_{MSE} + \lambda \ell_{AC}, \quad (13)$$

where $\mathcal{A}$ is the uncertainty class, y is the true MOCU value, $\hat{y}$ is the estimate by the model, and $\lambda \in [0, 1]$ is a weight parameter. In practice, y is numerically estimated via the Monte Carlo sampling approach [7]. The MSE loss is defined as a squared difference $\ell_{MSE} = (y - \hat{y})^2$ between the numerical MOCU estimate y via the sampling-based approach and the estimate $\hat{y}$ by the proposed model.

In addition to this conventional MSE loss, in this study, we introduce a novel constraint loss that penalizes the deviation from an axiom of MOCU. Our axiomatic constraint loss is motivated by the characteristics of MOCU and the MOCU-based OED scheme that are elaborated in Sec. 2.3. As described before, given the outcome of a pairwise synchronization experiment, we either increase the lower bound $a_{i,j}^l$ or decrease the upper bound $a_{i,j}^u$, as a result of which the objective uncertainty quantified by MOCU is reduced. However, due to the stochastic nature of the numerical estimates obtained by the sampling-based approach [7], the random fluctuations in the estimated MOCU values may overshadow the actual change in MOCU that results from the updated uncertainty class. We introduce a novel constraint loss to alleviate such fluctuation artifacts. Formally, according to the definition of MOCU in Eq. (5) and the uncertainty class $\mathcal{A}$ of the Kuramoto model defined in Sec. 2.1, the MOCU is an increasing (decreasing) function of the upper bound $a_{i,j}^u$ (lower bound $a_{i,j}^l$), which can be represented as follows:

$$\frac{\partial \text{MOCU}}{\partial a_{i,j}^l} \leq 0 \text{ and } \frac{\partial \text{MOCU}}{\partial a_{i,j}^u} \geq 0 \text{ for } 1 \leq i < j \leq N. \quad (14)$$

Based on the axiom above, we define the axiomatic constraint loss function as follows:

$$\ell_{AC} = \sum_{1 \leq i < j \leq N} [\max(\frac{\partial \hat{y}}{\partial a_{i,j}^\ell}, 0)^2 + \max(-\frac{\partial \hat{y}}{\partial a_{i,j}^u}, 0)^2]. \quad (15)$$

This loss serves as a soft constraint that stabilizes the numerical accuracy of the learned model.

3.3 Training details

We trained our MPNN based on a two-phase training strategy. First, the MPNN model was trained solely based on a training dataset that contains samples generated by the Kuramoto model with $N = 5$ oscillators. In the second phase, we created a training dataset where 50% of the data points were obtained by randomly sampling the data points from the five-oscillator Kuramoto model dataset and the remaining 50% was generated based on a seven-oscillator model. This strategy can save the amount of computation needed for creating the training data, as generating the data points for the seven-oscillator Kuramoto model is much more costly compared to generating the data points for the five-oscillator model. The trained MPNN model is then used to perform MOCU evaluation and OED for both uncertain Kuramoto models with five oscillators as well as Kuramoto models with seven oscillators. In each phase, we trained the model for 400 epochs on the training set with a batch size of 128, a learning rate of 0.001, and Adam optimizer [37]. The training takes approximately 1 GPU hour for each phase. It should be noted that this does not include the time for generating the training data, which is computationally costly as it involves solving a large number of ordinary differential equations for Kuramoto models with different parameter values. The strength of the axiomatic constraint loss λ was set to 0.0001 in order to give the model larger freedom at the beginning of the training process and stronger regularization when ℓ_{MSE} approaches the same magnitude as ℓ_{AC} . In each epoch, we evaluated the model on a validation set. The best model with the lowest validation loss was then used for performance evaluation and OED simulations in Sec. 4.

4 Simulation Results

We comprehensively validated our proposed approach based on two different types of Kuramoto models: one type that consists of five oscillators with unknown interaction strength and the other with seven oscillators with unknown interaction strength. First, we assessed the performance of the proposed MPNN model for predicting the MOCU value for a given uncertainty class and compared it against the latest Monte Carlo sampling approach in a recent study [7]. Second, we validated the efficacy of the proposed axiomatic constraint loss used for training the MPNN model by evaluating its impact on preserving the relative ranking based on the actual MOCU values. Finally, we evaluated the performance of the proposed approach in the context of OED for robust synchronization of uncertain Kuramoto models.

4.1 Generation of the training dataset

Table 1: Equations and distributions of the parameters used for generating the data points in the training dataset.

Equations	Parameter distribution
$F_{ij} = \frac{ w_i - w_j }{2}$	$w_i \sim U[-C, C]$
$\frac{a_{i,j}^u + a_{i,j}^\ell}{2} = [b_{ij}d_{i,j}^{strong} + (1 - b_{ij})d_{i,j}^{weak}]F_{ij}$	$d_{i,j}^{strong} \sim U[0, D_1]$
$a_{i,j}^u - a_{i,j}^\ell = 2 * d_{i,j}^{uncertain} F_{ij}$	$d_{i,j}^{weak} \sim U[0, D_2]$
	$d_{i,j}^{uncertain} \sim U[0, D_3]$
	$b_{i,j} \sim \text{Ber}(p = 0.5)$

Table 1 shows the equations and the parameter distributions based on which the training samples were generated. $U[x, y]$ is the uniform distribution ranging from x to y , and $\text{Ber}(p)$ is a Bernoulli random variable determining whether the oscillator pair have strong interaction strength factor $d_{i,j}^{strong}$ or weak interaction strength factor $d_{i,j}^{weak}$. For 67% of the training dataset, we set $b_{i,j} = b_{i,k}$ ($i \leq j < k \leq N, \forall i$), separating oscillators into two partitions (weak or strong). For the rest of the dataset, we sampled $b_{i,j}$ independently without such restriction. We generated a dataset that consists of 70,000 uncertain Kuramoto models with $N = 5$ oscillators and 28,000 uncertain Kuramoto models with $N = 7$ oscillators. For Kuramoto models with $N = 5$, we use $C = 6$, $D_1 = 1.1$, $D_2 = 0.6$, and $D_3 = 0.3$, respectively. For Kuramoto models with $N = 7$, we have $C = 10$, $D_1 = 1.2$, $D_2 = 0.25$, and $D_3 = 0.6$, respectively. We used the sampling-based approach [7] with a sample size of $K = 20,480$ as a ground truth approach to compute the MOCU for uncertain Kuramoto models in the training dataset. The resulting dataset had an average MOCU of 0.2378 (with a standard deviation of 0.21) for $N = 5$, and an average MOCU of 0.7310 (with a standard deviation of 0.6296) for

$N = 7$. The ground truth MOCU was then normalized to zero mean and unit variance for training. Based on this dataset, 96% of the data points were used for training, and the rest of the data points were used for validation. All experiments were performed on a workstation equipped with *Intel i9-9900K*, 32GB RAM, and *GeForce RTX 2080 Ti*. The Python code of our MPNN model for accelerated MOCU estimation and MOCU-based OED is available at <https://github.com/Levishery/AccelerateOED>.

4.2 Performance evaluation for MOCU prediction

We evaluated the performance of the proposed model on a test set consisting of 3,000 uncertain Kuramoto models with $N = 5$ oscillators and 2,000 uncertain Kuramoto models with $N = 7$ oscillators, sampled from the same distribution described in Table 1. Figure 3 shows the comparison results of the proposed models with benchmark models in terms of time complexity and the MSE on the test dataset. MP denotes the MPNN model trained with only the MSE loss, whereas MP+ corresponds to the MPNN model trained with the loss function in (13) that incorporates the axiomatic constraint. Ensemble denotes the prediction from an ensemble of three MPNN models, each of which is trained with different random parameter initialization. The ensemble prediction is obtained by taking the average of the predictions made by the different MPNN models in the ensemble. As a reference, the variance of the sampling-based approach [7] on the test set is also shown. For comparison, we trained two other machine learning (ML) models, a convolutional neural network (CNN) [38] and a multilayer perceptron (MLP). The input of the CNN model is a three-channel $N \times N$ image, representing natural frequency, and interaction strength lower/upper bound, respectively. MLP accepts the input vector of length N^2 , which is the concatenation of the natural frequency, the upper triangle of the lower/upper bound matrix. Note that, unlike the input for the MPNN, the dimension of the input for the CNN and MLP is fixed, which requires training separate ML models for the MOCU estimation of uncertain Kuramoto models with different numbers of oscillators.

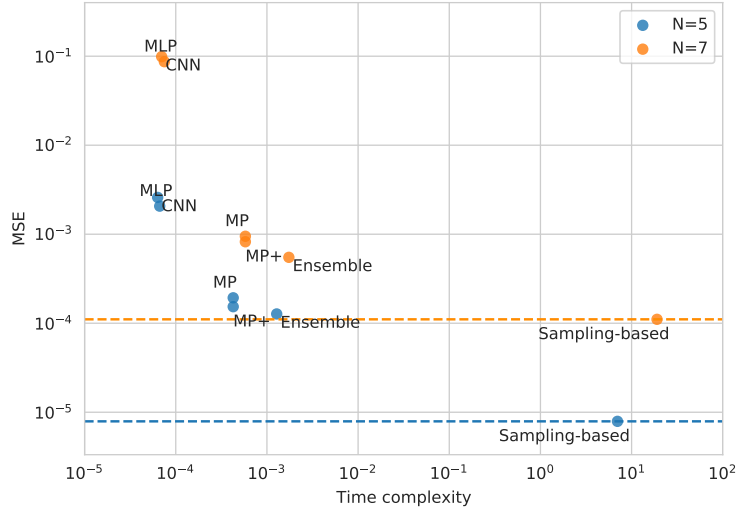


Figure 3: Comparison of the time complexity and accuracy among different methods for predicting MOCU.

As shown in Fig. 3, the proposed models (MP, MP+, and Ensemble MP+) outperformed the baseline neural networks (i.e., CNN and MLP) by significant margins in terms of prediction accuracy. This was likely due to the effective inductive bias [39] introduced by the proposed models. In fact, the inductive bias of the MPNN limits the search space of the models to be invariant to node and edge permutations, which is beneficial for graphical data. During training, we observed that all the deep learning models converge to similar training MSE loss, while the test accuracy significantly differs across different models, as shown in Fig. 3. The results in Fig. 3 clearly indicate the generalizability of the MPNN. Our proposed message-passing models tended to yield accurate predictions that were very close to those obtained by the original sampling-based approach. However, the proposed models were significantly faster than the sampling-based approach, resulting in four to five orders of magnitude speed-up. Especially, the speed gains were more prominent for the larger system. For example, for Kuramoto models with $N = 7$ oscillators, the sampling-based approach took 19 seconds on average to estimate the MOCU for a given uncertain Kuramoto model. On the contrary, it took only an average of 1.160 seconds for the MPNN model to estimate MOCU for the entire test set (2,000 data points). This amounts to a 32,758 speed-up. We expect that the relative computational efficiency of the proposed method with respect to the sampling-based approach will further increase for more complex Kuramoto models with

a larger number of oscillators. Additionally, results shown in Fig. 3 show that the constraint loss ℓ_{AC} improves the prediction accuracy. It is worth noting that the main advantage of incorporating this loss function into the learning objective is its ability to maintain the relative rank of the MOCU of each experiment, which will be discussed further in Sec. 4.3. The results also showed that the ensemble scheme is beneficial for enhancing accuracy at an increased computational cost.

Besides, we conducted additional experiments to further compare the performance of the trained MPNN model against other popular models designed for graph-based regression tasks [40, 41]. Specifically, we trained and tested Principal Neighbourhood Aggregation (PNA) [40] that replaces the sum aggregator in the MPNN model with a degree-scaled aggregator and Graphormer [41] that incorporates structural encoding schemes (such as centrality encoding and spatial encoding) that leverage structural information of a graph reflected on nodes and their neighbors. For Graphormer, we tested two representative architectures *Graphormer_{SLIM}* and *Graphormer_{BASE}*. The simulation results indicated that there is no noteworthy difference among the considered models in predicting MOCU. We also investigated the efficacy of the proposed axiomatic constraint loss by incorporating it into the training of the PNA model along with the MSE loss. As expected, the proposed axiomatic constraint loss was shown to be effective in improving the predictive performance of the PNA model. Detailed performance comparison results can be found in the supplementary information.

4.3 Effect of axiomatic constraint loss on regression accuracy

As discussed in Section 3.2, for the surrogate models to be effective in the context of MOCU-based OED, they need to be capable of maintaining the order of experiments in terms of their true expected remaining MOCU. If the (expected remaining) MOCU predicted by the surrogate models significantly changes the priority of the available experiments—compared to the case when the experiments are prioritized based on the true (expected remaining) MOCU—adopting such surrogate models would result in the degradation of the efficacy of the selected experiments. In the MOCU-based OED loop, the uncertainty class is updated based on the experimental outcome of the selected experiment. In our application, we reduce the range of uncertain interaction strength $a_{i,j} \in [a_{i,j}^\ell, a_{i,j}^u]$ based on the experimental outcome (i.e., whether the selected oscillator pair is synchronized or not) of the selected oscillator pair (i, j) : either by increasing the lower bound $a_{i,j}^\ell$ if (i, j) gets synchronized or by decreasing the upper bound $a_{i,j}^u$, otherwise. While the extent of uncertainty reduction depends on various factors, MOCU should be a non-increasing function of model uncertainty. In other words, when the uncertain range decreases for one of the interaction strengths $a_{i,j}$, MOCU should decrease or stay the same (if the reduced range for $a_{i,j}$ does not affect the task—i.e., robust synchronization), but never increase. In that regard, we assessed the performance of the proposed MPNN model trained using the loss function in (13) in terms of its ability to preserve the order of MOCU values as we change $a_{i,j}^\ell$ or $a_{i,j}^u$.

Table 2: Evaluation of different methods for their respective ability to preserve ranking.

Number of oscillators	$N = 5$		$N = 7$	
Method	$a_{i,j}^\ell \uparrow$	$a_{i,j}^u \downarrow$	$a_{i,j}^\ell \uparrow$	$a_{i,j}^u \downarrow$
Sampling-based	0.6599	0.6543	0.5720	0.5660
MP	0.8553	0.8466	0.7540	0.7600
MP+	0.9254	0.9178	0.8660	0.8475

Table. 2 summarizes the performance of the different MOCU estimation schemes in terms of their ability to preserve the ordering of the MOCU values—and ultimately, the ranking of the potential experiments according to their expected efficacy in reducing MOCU. We used the same test data described in 4.2 except for reducing the dataset size to one-fifth for the sampling-based approach due to its high computational cost. Analogous to updates in OED, for each uncertain Kuramoto model in the test set, we randomly selected oscillator pair (i, j) and then increased $a_{i,j}^\ell$ or decreased $a_{i,j}^u$ to $(1/2) * (a_{i,j}^\ell + a_{i,j}^u)$, which are denoted as $a_{i,j}^\ell \uparrow$ and $a_{i,j}^u \downarrow$ in Table. 2, respectively. As the performance metric, we measured the percentage of the test data for which a given model yields lower MOCU for the reduced uncertainty class as expected.

As we can see in Table 2, the MPNN models (both MP and MP+) were very effective in preserving the ordering of the estimated MOCU values for different lower and upper bounds. This is likely due to the stochastic nature of the numerical estimates obtained by the sampling-based approach, where the statistical deviation between the estimates may be relatively larger compared to the actual difference between the MOCU values for small changes in the bounds $a_{i,j}^\ell$ and $a_{i,j}^u$. The results in Table 2 show that the message-passing network model tends to effectively regularize

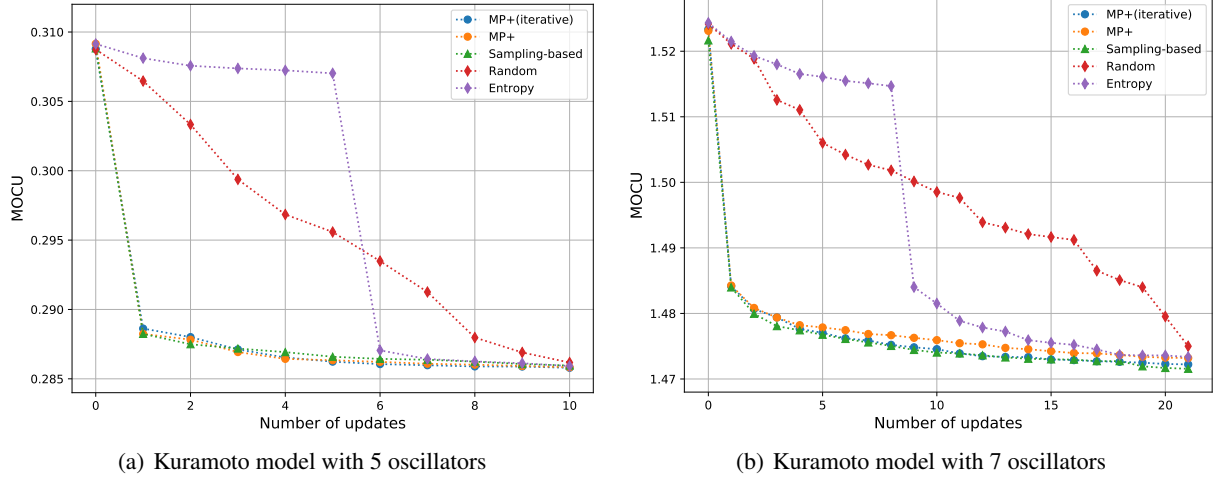


Figure 4: Objective model uncertainty (quantified by MOCU) decreases with sequential experimental updates. Various experimental design strategies are compared to assess their respective efficacy in prioritizing advantageous experiments in the design space.

the statistical fluctuations in the MOCU estimates in the training data generated by the sampling-based scheme [7]. Furthermore, as we would expect, incorporating the axiomatic constraint loss ℓ_{AC} in Eq. (15) into the learning objective clearly improves the preservation of the ordering of the MOCU estimates. Our results demonstrate the solid ability of our proposed MPNN model to preserve the MOCU-based ordering of different uncertainty classes. This is a critical property as a surrogate model for MOCU-based experimental design since it allows us to reliably prioritize potential experiments in the experimental design space by accurately quantifying the expected impact of a given experiment on reducing model uncertainty.

4.4 Evaluation of OED performance

Finally, we verified the efficacy of the proposed approach by applying the MPNN model to the OED loop. Our goal was to assess how effective the experiments selected using the MPNN-based scheme were in reducing the model uncertainty and how its performance compares to other OED schemes. For simulations, we used the uncertainty classes considered in [21]. Specifically, we assumed that the $N = 5$ oscillator model has natural frequencies $\omega = [-2.50, -0.6667, 1.1667, 2.0, 5.8333]^T$. The uncertainty class is defined as:

$$\mathbf{a}^u = [1.0541 \quad 0.6325 \quad 0.7762 \quad 1.4375 \quad 1.0542 \quad 0.6900 \quad 1.6819 \quad 0.4791 \quad 2.6833 \quad 2.2041]^T, \quad (16)$$

$$\mathbf{a}^\ell = [0.7791 \quad 0.4675 \quad 0.5737 \quad 1.0625 \quad 0.7792 \quad 0.5100 \quad 1.2431 \quad 0.3541 \quad 1.9833 \quad 1.6291]^T. \quad (17)$$

For the $N = 7$ oscillator model, we assumed that the natural frequencies of the oscillators were defined as $\omega = [-3.46, -1.96, -0.68, -0.38, -0.37, 6.12, 8.3287]^T$. The uncertainty class was defined as:

$$\mathbf{a}^u = \begin{bmatrix} 0.848 & 0.988 & 1.446 & 1.607 & 3.820 & 0.915 & 0.400 & 0.850 & 0.419 & 4.162 \\ & 1.090 & 0.122 & 0.039 & 2.124 & 0.872 & 0.007 & 2.737 & 1.804 & 1.360 & 0.744 & 1.174 \end{bmatrix}^T, \quad (18)$$

$$\mathbf{a}^\ell = \begin{bmatrix} 0.073 & 0.172 & 0.153 & 0.054 & 0.501 & 0.463 & 0.043 & 0.015 & 0.096 & 0.501 \\ & 0.103 & 0.007 & 0.009 & 0.139 & 0.408 & 0.000 & 0.131 & 0.119 & 0.300 & 0.286 & 0.131 \end{bmatrix}^T. \quad (19)$$

Note that these data points are covered by the distribution of the training data used to train our MPNN models. The natural frequency of the external control oscillator is set to the average frequency of the oscillators in the model.

For each uncertainty class, we determined a series of experiments based on a specific experimental design scheme and assessed how each experimental update reduces model uncertainty quantified by MOCU. For comparison, we evaluated the performance of the proposed MPNN-based scheme, the original sampling-based scheme in [7], the entropy-based scheme, and the random selection approach. In the MPNN-based scheme, we estimated the expected remaining MOCU for all $\binom{N}{2}$ experiments in the experimental design space using the MPNN model, which was then used to prioritize the experiments. Starting from the experiment with the smallest expected remaining MOCU, we

performed the selected experiment, updated the uncertainty class based on the outcome, and continued this process with the next best experiment until all experiments in the experimental space were exhausted. In the sampling-based scheme, MOCU was numerically estimated via sampling as described in [7], where the experimental sequence was determined in a similar fashion based on the remaining expected MOCU. We also considered an iterative scheme based on our MPNN model, where the expected remaining MOCU values of the available experiments were re-estimated after each experimental update, based on which the next experiment was selected in an iterative manner. Due to the formidably high computational cost, the sampling-based scheme was not considered for iterative experimental design. In the entropy-based approach, we prioritized the experiments based on the uncertain range $a_{i,j}^u - a_{i,j}^l$, from the experiment for the pair with the largest uncertain range to the smallest. Finally, in the random selection approach, we performed experimental updates in a randomized order. For reliable estimation of the MOCU at each experimental update, we numerically computed the MOCU values 10 times using the sampling-based scheme and used the average. In each MOCU computation, a sample size of $K = 20,480$ was used. The overall experiments were repeated 70 times by randomly sampling the true (unknown) Kuramoto model from the uncertainty class.

Figure 4 shows the performance evaluation results of the various experimental design schemes. For each scheme, the corresponding curve shows how the model uncertainty (quantified by MOCU) decreases as the number of experimental updates increases. Results are shown for uncertain Kuramoto models that consist of 5 oscillators (see Fig. 4(a)) and uncertain Kuramoto models with 7 oscillators (see Fig. 4(b)). As shown in Fig. 4, the sampling-based approach achieved the best OED performance in reducing the model uncertainty. The proposed MPNN-based scheme (MP+) yielded virtually identical OED performance compared to the sampling-based scheme in both simulations. However, the computational gain of adopting the MPNN model for MOCU-based OED is huge, as it can accelerate the computational speed of OED by four to five orders of magnitude for the Kuramoto models considered in this study. For example, for $N = 7$, it took an average of 2,995.8 seconds (with a standard deviation of 0.4988) for the original sampling-based OED scheme to compute the expected remaining MOCU values for all possible experiments for a single experimental update. In comparison, it took only an average of 0.0340 seconds (with a standard deviation of 0.0030) for our proposed MPNN model. Similarly, for $N = 5$, the computation of the expected remaining MOCU values for all possible experiments for a single experimental update took an average of 1,345.4 seconds (with a standard deviation of 0.5262) when the sampling-based OED scheme was used. On the other hand, the MP model needed only an average of 0.0321 seconds (with a standard deviation of 0.0019) for the same task. The random and entropy-based experimental design approaches, in contrast, failed to effectively reduce the model uncertainty within the first few experiments. On average, the random selection scheme resulted in a linear decrease of the model uncertainty with respect to the number of updates, as one would expect. The entropy-based scheme, which prioritizes the experiment that can reduce the uncertainty of the interaction strength $a_{i,j}$ with the largest uncertain range $a_{i,j}^u - a_{i,j}^l$, was not able to initially reduce the model uncertainty by a significant amount even after several experimental updates. This clearly illustrates why it is important for OED techniques to focus on reducing the model uncertainty that practically matters, which can only be done by quantifying how the uncertainty impacts the operational objective (e.g., achieving robust global frequency synchronization of an uncertain Kuramoto model with minimum control cost).

Interestingly, the iterative MPNN-based OED scheme did not result in performance improvement in our simulations. The potential underlying reason for this observation may be found in the MOCU reduction trends shown in Fig. 4. For both $N = 5$ and $N = 7$, a single experiment—if properly selected by a MOCU-based OED scheme—is able to effectively reduce the objective model uncertainty. As a result, the subsequent experiments have only a relatively small impact on uncertainty reduction. However, we expect that an iterative OED scheme would be beneficial in more complicated scenarios, where multiple experiments would be needed to effectively reduce the overall model uncertainty.

5 Concluding Remarks

In this work, we proposed a novel approach for accelerating MOCU-based objective-UQ and OED based on neural message-passing. To the best of our knowledge, this is the first study that demonstrates the possibility of designing accurate data-driven surrogate models for efficient objective-UQ and OED based on the concept of MOCU. As demonstrated, based on uncertain Kuramoto models that consist of interacting oscillators whose coupling strengths are not known with certainty, the proposed approach significantly accelerates MOCU estimation and MOCU-based OED. Compared to the original sampling-based approach [7], the proposed model was shown to improve the speed by four to five orders of magnitude for Kuramoto models with $N = 5$ or $N = 7$ oscillators. We expect that the overall gain in terms of computational efficiency will be even higher for a more complex Kuramoto model with a larger number of oscillators.

It is worth noting that training the neural network requires generating sufficient training data, where the original sampling-based approach [7] is used for MOCU computation for various uncertain Kuramoto models. However, once the training data are generated, actual training of the MPNN requires a relatively short time and the trained MPNN can

be used for extremely fast and accurate MOCU estimation for various uncertain Kuramoto models. Considering that the OED process typically requires a large number of MOCU computations, our proposed scheme has the added benefit of moving the heavy computational cost of MOCU evaluations from online to offline. Nevertheless, it would be an interesting research direction to apply the proposed axiomatic constraint loss in a *self-supervised* manner to reduce the model’s dependency on the amount of labeled training data. Similar to the idea of constructive learning in [42], we may pre-train the deep learning model by randomly generating paired uncertainty classes and encouraging the model to predict the correct ranking, without using any ground truth labels. We expect that such a pre-training procedure may provide a good starting point for the model that can ultimately reduce the amount of labeled data needed to train a high-quality surrogate model.

In this study, we have mainly focused on applying the proposed MPNN to the objective-UQ and OED for uncertain Kuramoto models in order to demonstrate its advantages over the state-of-the-art, recently presented in [7]. However, it should be noted that the proposed model architecture and the learning scheme can be easily adapted to accelerate MOCU estimation and MOCU-based OED for different types of *systems* (i.e., beyond Kuramoto oscillator models) and other operational *objectives* (i.e., beyond robust synchronization or control), and a wide range of *experimental design* problems. For example, it would be interesting to apply the proposed MPNN-based approach to accelerate MOCU-based Bayesian active learning (BAL). As proposed in [13, 14], soft-MOCU (SMOCU) or weighted-MOCU (WMOCU)—which are variants of MOCU—can be used to guide the pool-based active learning process in order to increase the labeling efficiency and to effectively improve the accuracy of optimal Bayesian classifiers with a relatively small number of label acquisitions. By incorporating a DL-based surrogate model for MOCU estimation (or the evaluation of SMOCU/WMOCU-based acquisition functions), the overall computational efficiency of the BAL process could be significantly accelerated without degrading the learning outcomes.

Considering the enormous OED speed improvement that can be attained by our proposed MPNN-based method, we can envision the application of MOCU-based OED for experimental design in real-world complex uncertain systems, which was previously intractable due to the high computational cost. For example, Kuramoto oscillators are known to be able to model the brain network synchronization phenomena [19, 22, 43, 44, 45, 46]. Many cognitive and behavioral states, including perception, memory, and action, have been described as the emergent properties of coherent or phase-locked oscillations in transient neuronal ensembles. As shown in [45], non-invasive transcranial alternating-current stimulation (tACS) at the beta frequency (20 Hz) can increase local activity in the beta band and modulate the connectivity pattern of the stimulated motor cortex. In the aforementioned study [45], they model tACS as a controlling oscillator in the Kuramoto model similar to our work. The model was able to faithfully reproduce key features of the experimental data. Recently, the Kuramoto model was used to study the relationship between a brain network’s global synchronization and strong connections over a long distance, where the connectivity matrix was constructed from viral tracing experiments [46]. This research area provides promising venues for applying our MPNN-based scheme proposed in this work. It is our plan to explore the potential application of the accelerated objective-UQ and OED scheme via message-passing networks to the optimal therapeutic stimulation of brain networks under uncertainty.

Acknowledgment

The work of H.-M. Woo and B.-J. Yoon was supported in part by the NSF Award 1835690 and the DOE Award DE-SC0019303.

References

- [1] Kenneth J Ryan. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *Journal of Computational and Graphical Statistics*, 12(3):585–603, 2003.
- [2] Quan Long, Marco Scavino, Raúl Tempone, and Suojin Wang. Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Computer Methods in Applied Mechanics and Engineering*, 259:24–39, 2013.
- [3] Panagiotis Tsilifis, Roger G Ghanem, and Paris Hajali. Efficient Bayesian experimentation using an expected information gain lower bound. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):30–62, 2017.
- [4] Joakim Beck, Ben Mansour Dia, Luis FR Espath, Quan Long, and Raul Tempone. Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain. *Computer Methods in Applied Mechanics and Engineering*, 334:523–553, 2018.
- [5] Byung-Jun Yoon, Xiaoning Qian, and Edward R Dougherty. Quantifying the objective cost of uncertainty in complex dynamical systems. *IEEE Transactions on Signal Processing*, 61(9):2256–2266, 2013.

- [6] Byung-Jun Yoon, Xiaoning Qian, and Edward R Dougherty. Quantifying the multi-objective cost of uncertainty. *IEEE Access*, 9:80351–80359, 2021.
- [7] Youngjoon Hong, Bongsuk Kwon, and Byung-Jun Yoon. Optimal experimental design for uncertain systems based on coupled differential equations. *IEEE Access*, 9:53804–53810, 2021.
- [8] Roozbeh Dehghannasiri, Byung-Jun Yoon, and Edward R Dougherty. Optimal experimental design for gene regulatory networks in the presence of uncertainty. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 12(4):938–950, 2014.
- [9] Ariana Broumand, Mohammad Shahrokh Esfahani, Byung-Jun Yoon, and Edward R Dougherty. Discrete optimal Bayesian classification with error-conditioned sequential sampling. *Pattern Recognition*, 48(11):3766–3782, 2015.
- [10] Roozbeh Dehghannasiri, Byung-Jun Yoon, and Edward R Dougherty. Efficient experimental design for uncertainty reduction in gene regulatory networks. In *BMC Bioinformatics*, volume 16, pages 1–18. Springer, 2015.
- [11] Anjana Talapatra, Shahin Boluki, Thien Duong, Xiaoning Qian, Edward Dougherty, and Raymundo Arróyave. Autonomously efficient experiment design for materials discovery with Bayesian model averaging. *Physical Review Materials*, 2(11):113803, 2018.
- [12] Guang Zhao, Xiaoning Qian, Byung-Jun Yoon, Francis J Alexander, and Edward R Dougherty. Model-based robust filtering and experimental design for stochastic differential equation systems. *IEEE Transactions on Signal Processing*, 68:3849–3859, 2020.
- [13] Guang Zhao, Edward Dougherty, Byung-Jun Yoon, Francis Alexander, and Xiaoning Qian. Uncertainty-aware active learning for optimal Bayesian classifier. In *International Conference on Learning Representations*, 2021.
- [14] Guang Zhao, Edward Dougherty, Byung-Jun Yoon, Francis J Alexander, and Xiaoning Qian. Bayesian active learning by soft mean objective cost of uncertainty. In *International Conference on Artificial Intelligence and Statistics*, pages 3970–3978. PMLR, 2021.
- [15] Guang Zhao, Edward Dougherty, Byung-Jun Yoon, Francis J Alexander, and Xiaoning Qian. Efficient active learning for Gaussian process classification by error reduction. In *Thirty-fifth Conference on Neural Information Processing Systems*, 2021.
- [16] Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In *International symposium on mathematical problems in theoretical physics*, pages 420–422. Springer, 1975.
- [17] Per Sebastian Skardal and Alex Arenas. Control of coupled oscillator networks with application to microgrid technologies. *Science advances*, 1(7):e1500339, 2015.
- [18] Klaus Lehnertz, Stephan Bialonski, Marie-Therese Horstmann, Dieter Krug, Alexander Rothkegel, Matthias Staniek, and Tobias Wagner. Synchronization phenomena in human epileptic brain networks. *Journal of neuroscience methods*, 183(1):42–48, 2009.
- [19] Ali Mohseni, Shahriar Gharibzadeh, and Fatemeh Bakouie. The role of driver nodes in managing epileptic seizures: Application of Kuramoto model. *Journal of theoretical biology*, 419:108–115, 2017.
- [20] Nicholas Metropolis and Stanislaw Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.
- [21] Hyun-Myung Woo, Youngjoon Hong, Bongsuk Kwon, and Byung-Jun Yoon. Accelerating optimal experimental design for robust synchronization of uncertain Kuramoto oscillator model using machine learning. *IEEE Transactions on Signal Processing*, 69:6473–6487, 2021.
- [22] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.
- [23] Huanlai Xing, Zhiwen Xiao, Rong Qu, Zonghai Zhu, and Bowen Zhao. An efficient federated distillation learning system for multitask time series classification. *IEEE Transactions on Instrumentation and Measurement*, 71:1–12, 2022.
- [24] Zhiwen Xiao, Xin Xu, Huanlai Xing, Fuhong Song, Xinhan Wang, and Bowen Zhao. A federated learning system with enhanced feature extraction for human activity recognition. *Know.-Based Syst.*, 229(C), oct 2021.
- [25] Constance Hammond, Hagai Bergman, and Peter Brown. Pathological synchronization in Parkinson’s disease: networks, models and treatments. *Trends in neurosciences*, 30(7):357–364, 2007.
- [26] Michael Breakspear, Stewart Heitmann, and Andreas Daffertshofer. Generative models of cortical oscillations: Neurobiological implications of the Kuramoto model. *Frontiers in Human Neuroscience*, 4:190, 2010.

- [27] Ruben Schmidt, Karl J. R. LaFleur, Marcel A. de Reus, Leonard H. van den Berg, and Martijn P. van den Heuvel. Kuramoto model simulation of neural hubs and dynamic synchrony in the human cerebral connectome. *BMC Neuroscience*, 16(1):54, 2015.
- [28] Yufeng Guo, Dongrui Zhang, Zhuchun Li, Qi Wang, and Daren Yu. Overviews on the applications of the Kuramoto model in modern power system analysis. *International Journal of Electrical Power & Energy Systems*, 129:106804, 2021.
- [29] J. Buck and E. Buck. Synchronous fireflies. *Scientific American*, 234(5):74–79, 82, 1976.
- [30] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. *Advances in neural information processing systems*, 28, 2015.
- [31] Rianne van den Berg, Thomas N Kipf, and Max Welling. Graph convolutional matrix completion. *arXiv preprint arXiv:1706.02263*, 2017.
- [32] Danfei Xu, Yuke Zhu, Christopher B Choy, and Li Fei-Fei. Scene graph generation by iterative message passing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5410–5419, 2017.
- [33] Cheng Zhang, Zhaopeng Cui, Yinda Zhang, Bing Zeng, Marc Pollefeys, and Shuaicheng Liu. Holistic 3d scene understanding from a single image with implicit representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8833–8842, 2021.
- [34] Martin Simonovsky and Nikos Komodakis. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3693–3702, 2017.
- [35] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.
- [36] Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order Matters: Sequence to sequence for sets. *International Conference on Learning Representations*, 2016.
- [37] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [38] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
- [39] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.
- [40] Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, and Petar Veličković. Principal neighbourhood aggregation for graph nets. *Advances in Neural Information Processing Systems*, 33:13260–13271, 2020.
- [41] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do Transformers really perform bad for graph representation? In *NeurIPS*, 2021.
- [42] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.
- [43] Manfred G Kitzbichler, Marie L Smith, Søren R Christensen, and Ed Bullmore. Broadband criticality of human brain network synchronization. *PLoS computational biology*, 5(3):e1000314, 2009.
- [44] Fabiano AS Ferrari, Ricardo L Viana, Sergio Roberto Lopes, and Ruedi Stoop. Phase synchronization of coupled bursting neurons and the generalized Kuramoto model. *Neural Networks*, 66:107–118, 2015.
- [45] Christiane A Weinrich, John-Stuart Brittain, Magdalena Nowak, Reza Salimi-Khorshidi, Peter Brown, and Charlotte J Stagg. Modulation of long-range connectivity patterns via frequency-specific stimulation of human cortex. *Current Biology*, 27(19):3061–3068, 2017.
- [46] Hannah Choi and Stefan Mihalas. Synchronization dependent on spatial structures of a mesoscopic whole-brain network. *PLOS Computational Biology*, 15:1–21, 04 2019.

Supplemental Material for “Neural message passing for objective-based uncertainty quantification and optimal experimental design”

Performance comparison of various deep learning models for MOCU prediction

Table 3: Performance comparison of various deep learning models for predicting MOCU.

Methods	#params	test MSE($\times 10^{-4}$)	
		N=5	N=7
MP [22]	154, 593	1.930	9.483
MP+	154, 593	1.534	8.226
PNA [40]	675, 902	2.226	9.836
PNA+	675, 902	1.925	7.966
Graphormer _{SLIM} (batchsize= 128) [41]	578, 241	5.196	15.53
Graphormer _{SLIM} (batchsize= 1024) [41]	578, 241	1.663	9.640