

RESEARCH ARTICLE 

Optimal experiment design with adjoint-accelerated Bayesian inference

Matthew Yoko  and Matthew P. Juniper 

Department of Engineering, University of Cambridge, Cambridge, UK

Corresponding author: Matthew P. Juniper; Email: mpj1001@cam.ac.uk

Received: 09 January 2024; **Revised:** 15 March 2024; **Accepted:** 29 April 2024

Keywords: Bayesian data assimilation; information theory; optimal experiment design

Abstract


We develop and demonstrate a computationally cheap framework to identify optimal experiments for Bayesian inference of physics-based models. We develop the metrics (i) to identify optimal experiments to infer the unknown parameters of a physics-based model, (ii) to identify optimal sensor placements for parameter inference, and (iii) to identify optimal experiments to perform Bayesian model selection. We demonstrate the framework on thermoacoustic instability, which is an industrially relevant problem in aerospace propulsion, where experiments can be prohibitively expensive. By using an existing densely sampled dataset, we identify the most informative experiments and use them to train the physics-based model. The remaining data are used for validation. We show that, although approximate, the proposed framework can significantly reduce the number of experiments required to perform the three inference tasks we have studied. For example, we show that for task (i), we can achieve an acceptable model fit using just 2.5% of the data that were originally collected.

Impact Statement

Data-driven methods are becoming increasingly popular in many disciplines of engineering. In some applications, however, experimental data can be extremely expensive to collect, making data-driven methods costly to apply. We demonstrate a method that reduces the amount of data required for one data-driven method: Bayesian inference.

1. Introduction

Bayesian inference has proven to be a powerful data-driven approach for many problems in science and engineering. It allows noisy and sparse experimental data to be assimilated into qualitatively accurate physics-based models, rendering the model quantitatively accurate while quantifying the uncertainties in the model and data. In their review of machine learning in fluid dynamics, Brunton et al. (2020) argue that, for fluid mechanics problems, Bayesian inference may be superior to other machine learning techniques because of its robustness, but that it is hampered by the cost of the thousands of model evaluations required to evaluate the posterior distribution. This is true of typical sampling methods such as Markov Chain Monte Carlo, which can require hundreds of thousands of model evaluations to calculate the

 This research article was awarded Open Data badge for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



posterior. Recent work has, however, demonstrated sampling-free methods for variational data assimilation (Mons and Zaki, 2021; Buchta et al., 2022) and approximate Bayesian inference (MacKay, 2003; Isaac et al., 2015; Juniper and Yoko, 2022; Kontogiannis et al., 2022; Yoko and Juniper, 2024). These methods reduce the required model evaluations, making Bayesian inference feasible for computationally expensive models. Furthermore, the resulting model is interpretable, and tends to extrapolate better than physics-agnostic machine learning solutions.

In addition to parameter inference and uncertainty quantification, the Bayesian framework provides quantitative metrics to rank and compare candidate models. This is particularly useful when there are several plausible models, but no clear reason to prefer one over another. The Bayesian framework naturally enforces Occam's razor to select the simplest model that is capable of describing the data, for given measurement error and given priors (Jeffreys, 1973; Juniper and Yoko, 2022; Yoko and Juniper, 2024).

An outstanding problem is the cost of collecting data to infer the model parameters or discriminate between candidate models. In many cases, experimental data are expensive to collect, so we need to minimize the number of experiments required for parameter inference and model comparison. Similarly, sensors can be expensive, so we need to find the optimal sensor placement to obtain as much information as possible from the available sensors.

Approaches to optimal experimental design have been established by Fedorov (1972) in the frequentist framework, and Lindley (1956) in the Bayesian framework. Lindley's approach involves designing a utility function that describes the objective of the experiment, and then maximizing the utility over all possible experiments. This approach has been used to identify optimal experiments for parameter inference in a range of problems in physics and engineering (Ryan, 2003; Loredo, 2004; Huan and Marzouk, 2013). Many of these studies, however, have struggled with the computational cost of evaluating the utility function, which requires solving a high-dimensional double integral. The integral is typically evaluated using Monte Carlo integration, which requires many thousands of model evaluations. This makes the computational cost of performing Bayesian experimental design prohibitive.

Similarly, previous studies have attempted to apply Bayesian optimal sensor placement to problems in physics and engineering (Verma et al., 2019; Ercan and Papadimitriou, 2023). These studies have also relied on Monte Carlo integration to evaluate the utility function, which the authors note to be extremely computationally expensive. A method of approximating the utility function integral was introduced by Papadimitriou and Papadimitriou (2015), but it relies on an ad hoc decision about the correlation between sensors to avoid sensor clustering. Bidar et al. (2024) describe a methodology for optimal sensor placement that does not rely on calculating a utility function. Instead, Bidar et al. (2024) place sensors where the variance in a chosen predicted variable is largest. Like the work of Papadimitriou and Papadimitriou (2015), this requires an ad hoc method to avoid sensor clustering. Additionally, for multivariate problems, a second ad hoc decision must be made regarding which predicted variable should guide the sensor placement.

In this paper, we will demonstrate that the utility function can be evaluated cheaply with our adjoint-accelerated Bayesian inference framework. Similarly, we will show that the optimal sensor placement problem can be solved efficiently and without any ad hoc decisions. This framework is approximate, but we will demonstrate that it is still able to significantly reduce experimental costs. While the framework is general and has broad applicability, we will demonstrate it on examples from thermoacoustics. We choose thermoacoustics experiments because they are challenging to model, industrially relevant, and expensive to perform at large scale.

1.1. Thermoacoustic oscillations

Thermoacoustic oscillations occur in combustors when acoustic oscillations perturb the flame, causing heat release rate oscillations some time later. If these heat release rate oscillations coincide sufficiently with acoustic pressure oscillations, then the acoustic energy increases. This can lead to large amplitude oscillations which can degrade performance or cause structural damage. The phenomenon is challenging

to model because it is extremely sensitive to small changes (Juniper and Sujith, 2018). As a result, model predictions are similarly sensitive to small changes in the model parameters, the values of which are not known a priori (Juniper, 2018).

Thermoacoustic oscillations are often discovered late in the design process, when the first full-scale prototype is tested (Mongia et al., 2003). Once discovered, a solution is typically devised through trial-and-error. An extreme example of this occurred during the development of the F1, the main engine for the first stage of the Saturn V rocket. This required about 2000 full scale tests to remove thermoacoustic instability, costing around 2 billion U.S. dollars in the 1960s (Oefelein and Yang, 1993).

Thermoacoustics is therefore an ideal application for Bayesian inference and Bayesian optimal experiment design. We have previously used Bayesian inference to generate quantitatively accurate models of simple thermoacoustic systems (Juniper and Yoko, 2022; Yoko and Juniper, 2024) using a large amount of experimental data. In this paper, we revisit the dataset in Juniper and Yoko (2022) in order to assess how much data are strictly needed to achieve acceptable accuracy when performing several inference tasks.

2. Thermoacoustic problem description

In this paper, we demonstrate Bayesian optimal experiment design on a canonical thermoacoustic system, the hot wire Rijke tube. This is the simplest system that exhibits thermoacoustic oscillations, making it an ideal first test case for demonstrating this experiment design methodology. The system is cheap to operate, allowing us to collect a densely sampled dataset that can be used to validate the optimal experiment design methodology. It is also simple enough that we can gain valuable insight into how our experiment design methodology works. We now present the main features of the experimental rig and the physics-based model. Further details can be found in Juniper and Yoko (2022).

2.1. Experimental rig

The experimental rig is shown in Figure 1. It consists of a vertically mounted steel tube with a length of 1 m, inner diameter of 47.4 mm, and wall thickness of 1.7 mm. Both ends of the tube are open to ambient conditions. An electric heating element, shown in detail in Figure 2, is mounted on two support prongs and inserted into the bottom of the tube. The support prongs are attached to an electrically driven traverse so that the heater position can be controlled. The heater is connected to a programmable power supply that is set up to deliver a constant power output. We can therefore study how the thermoacoustic behavior changes with changes in (i) the heater position, and (ii) the heater power.

We record the acoustic pressure using eight probe microphones distributed along the length of the tube. The microphones are distributed with 0.1 m spacing, starting at 0.25 m from the bottom of the tube. For a

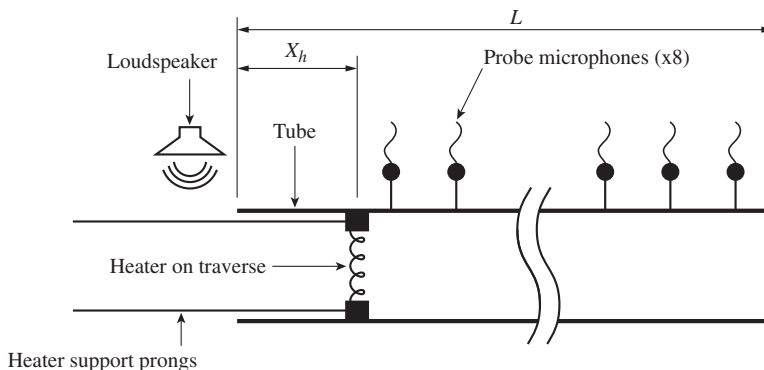


Figure 1. Diagram of the Rijke tube, rotated for convenience.

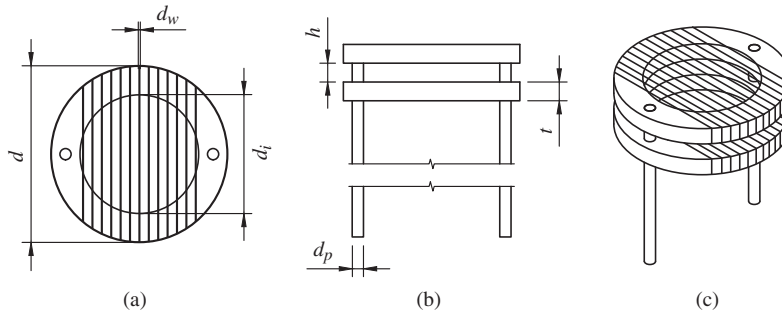


Figure 2. (a) Top view, (b) side view, and (c) isometric view of the heater, which consists of two identical concentric annular ceramic plates, each wound with nichrome wire. It is held in place by two threaded support prongs. The dimensions are $d = 47$ mm, $d_i = 31.6$ mm, $d_w = 0.6$ mm, $t = 5$ mm, $h = 5$ mm, and $d_p = 3$ mm. The power is supplied to the nichrome wire by two fabric-insulated copper wires (not shown), which each have diameter 4 mm

given combination of heater position and heater power, we use a loudspeaker to force the system near its fundamental frequency. We then terminate the forcing and record the pressure time series as the oscillations decay. From this signal, we extract (i) the decay rate, (ii) the natural frequency of oscillations, and (iii) the Fourier-decomposed pressure at each of the microphones, measured relative to a reference microphone. The reference microphone is chosen to be the microphone 0.75 m from the bottom of the tube.

2.2. Thermoacoustic network model

The thermoacoustic oscillations are modeled using a 1D thermoacoustic network model (Dowling and Stow, 2003; Juniper, 2018). The rig is divided into N acoustic elements in which forward traveling waves, $f(t - x/c)$, and backward traveling waves, $g(t - x/c)$, propagate. In a given element, the strength of these waves is considered to be constant. The acoustic pressure in the i th element is $p_i' = f_i + g_i$, and the acoustic velocity is $u_i' = (f_i - g_i)/(\bar{\rho}c)$, where $\bar{\rho}$ is the local mean density and c is the local sound speed.

The complex wave amplitudes in adjacent acoustic elements are related through jump conditions for the momentum and energy equations (Juniper, 2018). The jump conditions account for features of the rig that modify the wave amplitudes, such as the visco-thermal dissipation in the boundary layer, the drag and blockage of the heating element, and the fluctuating heat release rate of the heater. The wave reflection at either end of the tube is modeled with complex reflection coefficients. When modeling these jump conditions and boundary conditions, we introduce model parameters, the values of which are not known a priori. As a result, the models tend to be qualitatively accurate, but not quantitatively accurate. We therefore use Bayesian inference to (i) infer the most probable parameter values from data, and (ii) compare several candidate models for the jump conditions, and select the most likely model, given the data.

3. Bayesian inference framework

Our Bayesian inference framework is described in detail for general problems in MacKay (2003) and with a focus on thermoacoustics in Yoko and Juniper (2024). Applying Bayesian inference to physical problems requires three components: (i) a physical system from which experimental data are collected, (ii) a physics-based model that can predict the outcome of the experiments, and (iii) an inference framework. Here, we describe each component in turn and introduce the terminology and notation required for the following sections.

3.1. Physical system

We consider the task of gathering data by performing experiments on a physical system. The system has a set of design parameters that we control to change the system behavior:

$$\mathbf{z}_i = \mathcal{M}(\mathbf{x}_i) + \mathcal{N}(0, \mathbf{C}_e), \quad (1)$$

where \mathbf{z} is the vector of observed variables, \mathcal{M} is the measurement operator, and \mathbf{x} is the vector of design parameters. The observations are corrupted by experimental error, which we assume to be normally distributed with zero mean and covariance \mathbf{C}_e . We therefore only consider the random component of experimental error. Systematic experimental error can be combined with structural error in the model, which can be quantified a posteriori (Yoko and Juniper, 2024). We perform a set of N experiments and use the subscript $i = \{1, 2, \dots, N\}$ to denote the index of the experiment.

For the Rijke tube described in Section 2.1, the vector of design parameters, \mathbf{x} , consists of (i) the position of the heater inside the tube, and (ii) the power delivered to the heater. The measurement vector, \mathbf{z} , contains eight complex numbers: the eigenvalue, whose real part is the decay rate and imaginary part is the natural frequency, and the Fourier-decomposed pressure of seven microphones, measured relative to the reference microphone.

3.2. Physics-based model

We develop several plausible physics-based models that predict the outcome of the experiments:

$$\mathbf{s}_{i,j} = \mathcal{H}_j(\mathbf{x}_i, \mathbf{a}), \quad (2)$$

where \mathbf{s} are the model predictions, \mathcal{H} is the candidate physics-based model, \mathbf{x} is the vector of design parameters, and \mathbf{a} is a vector of unknown model parameters. We consider M candidate models and use the subscript $j = \{1, 2, \dots, M\}$ to denote the index of the candidate model. In the inference framework, we begin by assuming that the model is correct, meaning that we do not explicitly add a term for model error. We expect that, for some set of parameters, \mathbf{a}_{MP} , the model predictions, \mathbf{s} , will match the experimental observations, \mathbf{z} , for all possible design parameters, \mathbf{x} , within the measurement uncertainty.

For the thermoacoustic network model of the Rijke tube described in Section 2.2, the unknown parameters arise from the sub-models for the visco-thermal dissipation from the boundary layer and the heating element, the reflection coefficients at the ends of the tube, and the fluctuating heat release rate of the heater. We do not know the values of these parameters a priori, and we must often select between several plausible sub-models for each of the physical mechanisms. We address the first problem using Bayesian parameter inference and uncertainty quantification, and the second with Bayesian model comparison.

3.3. Adjoint-accelerated Bayesian inference

3.3.1. Parameter inference

For each candidate model, \mathcal{H}_j , we use the experimental observations to infer the most probable parameters, \mathbf{a}_{MP} . We begin by assigning a prior probability distribution over the vector of unknown parameters, which we choose to be Gaussian distributed. This prior allows us to incorporate any knowledge we might have about the parameters, while also encoding our confidence in this prior knowledge. We then perform N experiments to collect the data, $\mathbf{D} = \{\mathbf{z}_1, \dots, \mathbf{z}_N\}$. We assimilate the data by performing a Bayesian update on the parameter values:

$$p(\mathbf{a}|\mathbf{D}, \mathcal{H}_j) = \frac{p(\mathbf{D}|\mathbf{a}, \mathcal{H}_j)p(\mathbf{a}|\mathcal{H}_j)}{p(\mathbf{D}|\mathcal{H}_j)}. \quad (3)$$

The left-hand side of equation (3), called the posterior, is the probability density of the parameters, given the experimental data and the model. In general, the posterior cannot be evaluated analytically, and numerical computation typically requires millions of model evaluations. This is computationally

expensive, even for simple models. At the parameter inference stage, however, we are only interested in finding the most probable parameters, which are those that maximize the posterior. We therefore avoid constructing the full posterior, and instead use an optimization algorithm to find its peak. This is often called the “maximum a posteriori” (MAP) estimate of the parameter values.

The denominator of the right-hand side of [equation \(3\)](#) does not depend on the parameters. We can therefore find the most probable parameters by maximizing the product $p(\mathbf{D}|\mathbf{a}, \mathcal{H}_j)p(\mathbf{a}|\mathcal{H}_j)$. Recalling that the experimental uncertainty is assumed to be Gaussian distributed, the distribution $p(\mathbf{D}|\mathbf{a}, \mathcal{H}_j)$ is a Gaussian distribution over the data, for a given set of parameters. We also choose the prior distribution, $p(\mathbf{a}|\mathcal{H}_j)$, to be Gaussian distributed over the parameters. Under these conditions, we can transform the optimization problem into a quadratic optimization problem by defining the cost function, \mathcal{J} , to be the negative log of the numerator of [equation \(3\)](#):

$$\mathcal{J} = \frac{1}{2} \sum_{i=1}^N \left\{ (\mathbf{s}_i(\mathbf{a}) - \mathbf{z}_i)^T \mathbf{C}_e^{-1} (\mathbf{s}_i(\mathbf{a}) - \mathbf{z}_i) + (\mathbf{a} - \mathbf{a}_p)^T \mathbf{C}_a^{-1} (\mathbf{a} - \mathbf{a}_p) + K_i \right\}, \tag{4}$$

where \mathbf{a}_p is the vector of prior parameter values, \mathbf{C}_a is the covariance matrix describing the uncertainty in the prior, and K is a constant from the Gaussian pre-exponential factors, which has no impact on the most probable parameters, \mathbf{a}_{MP} .

We can solve the optimization problem with the fewest model evaluations by using gradient-based optimization. This requires that we calculate $\partial \mathcal{J} / \partial \mathbf{a}$, which we see from [equation \(4\)](#) requires the model parameter sensitivities, $\partial \mathbf{s} / \partial \mathbf{a}$. We obtain these using first-order adjoint methods.¹ The adjoint method is a technique for calculating the gradient of a function with respect to many parameters, with a computational cost that is independent of the number of parameters (Giles and Pierce, 2000; Luchini and Bottaro, 2014). The optimization problem can, of course, be solved with any minimization algorithm. However, we will see in the subsequent sections that we will also need the parameter sensitivities for uncertainty quantification, model selection and optimal experiment design. Therefore, it makes sense to use the parameter sensitivities to solve the optimization problem with the fewest model evaluations.

The parameter inference process is illustrated in [Figure 3](#) for a simple system with a single parameter, a , and a single observable variable, z . In [Figure 3\(a\)](#), we show the marginal probability distributions of the prior, $p(a)$, and the data, $p(z)$. The prior and data are independent, so we construct the joint distribution, $p(a, z)$ by multiplying the two marginals. In [Figure 3\(b\)](#), we overlay the model predictions, s , for various values of a . Marginalizing along the model predictions yields the true posterior, $p(a|z)$. This is possible for a cheap model with a single parameter, but exact marginalization quickly becomes intractable as the number of parameters increases. In [Figure 3\(c\)](#), we plot the cost function, \mathcal{J} , which is the negative log of the unnormalized posterior. We show the three steps of gradient-based optimization that are required to find the local minimum, which corresponds to the most probable parameters, a_{MP} .

3.3.2. Uncertainty quantification

Once we have found the most probable parameter values by minimizing [equation \(4\)](#), we estimate the uncertainty in these parameter values using Laplace’s method (Jeffreys, 1973; MacKay, 2003; Juniper and Yoko, 2022). This method approximates the posterior as a Gaussian distribution with a mean of \mathbf{a}_{MP} , and a covariance given by the Hessian of the cost function:

$$\mathbf{C}_a^{MP^{-1}} \approx \frac{\partial^2 \mathcal{J}}{\partial a_i \partial a_m} = \sum_{i=1}^N \left\{ \mathbf{C}_a^{-1} + \mathbf{J}_i^T \mathbf{C}_e^{-1} \mathbf{J}_i + (\mathbf{s}(\mathbf{a}) - \mathbf{z}_i)^T \mathbf{C}_e^{-1} \mathbf{H} \right\}, \tag{5}$$

where the summation is over the N experimental configurations at which data were collected, which are labeled with the index i , $\mathbf{J}_i = \mathbf{J}(\mathbf{a}_{MP}, \mathbf{x}_i)$ is the Jacobian matrix containing the parameter sensitivities of the

¹ Other methods for obtaining the parameter sensitivities, such as automatic differentiation, are also suitable.

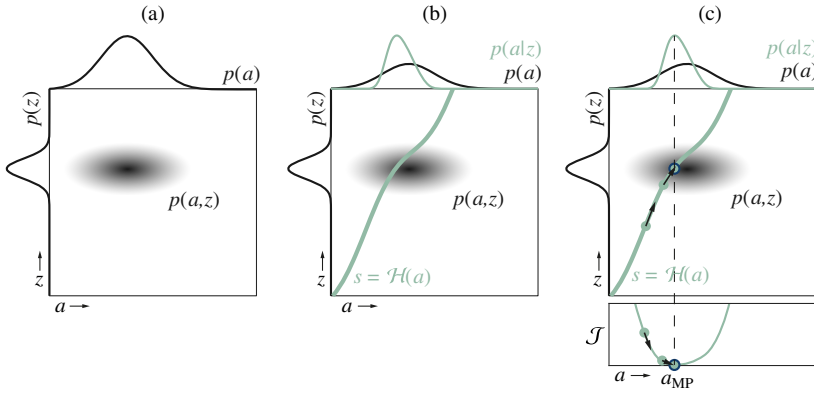


Figure 3. Illustration of parameter inference on a simple univariate system. (a) The marginal probability distributions of the prior and data, $p(a)$ and $p(z)$, as well as their joint distribution, $p(a,z)$ are plotted on axes of parameter value, a , vs. observation outcome, z . (b) The model, \mathcal{H} , imposes a functional relationship between the parameters, a , and the predictions, s . Marginalizing along the model predictions yields the true posterior, $p(a|z)$. This is computationally intractable for even moderately large parameter spaces. (c) Instead of evaluating the full posterior, we use gradient-based optimization to find its peak. This yields the most probable parameters, a_{MP} .

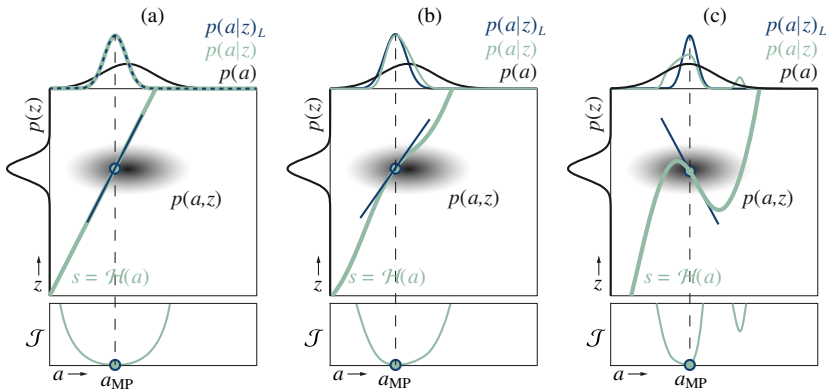


Figure 4. Illustration of uncertainty quantification for three univariate systems, comparing the true posterior; $p(a|z)$ to the approximate posterior from Laplace's method, $p(a|z)_L$. (a) The model is linear in the parameters, so the true posterior is Gaussian and Laplace's method is exact. (b) The model is weakly nonlinear in the parameters, the true posterior is slightly skewed, but Laplace's method yields a reasonable approximation. (c) The model is strongly nonlinear in the parameters, the posterior is multimodal and Laplace's method underestimates the uncertainty.

model predictions, $\partial s_k / \partial a_l$, and $\mathbf{H}_i = \mathbf{H}(\mathbf{a}_{MP}, \mathbf{x}_i)$ is the rank three tensor containing the second-order sensitivities, $\partial^2 s_k / \partial a_l \partial a_m$. We obtain \mathbf{J} using first-order adjoint methods, and \mathbf{H} using second-order adjoint methods.

The accuracy of this approximation depends on the functional dependence between the model predictions and the parameters. This is shown graphically in Figure 4 for three univariate systems. In Figure 4(a), the model is linear in the parameters. Marginalizing a Gaussian joint distribution along any intersecting line produces a Gaussian posterior, so Laplace's method is exact. In Figure 4(b), the model is weakly nonlinear in the parameters. The true posterior is skewed, but the Gaussian approximation is still

reasonable. This panel also shows a geometric interpretation of Laplace’s method: the approximate posterior is given by linearizing the model around \mathbf{a}_{MP} , and marginalizing the joint distribution along the linearized model. In Figure 4(c), the model is strongly nonlinear in the parameters, so the true posterior is multi-modal and the main peak is highly skewed. In this case, the gradient-based optimization algorithm will only find a single local minimum, which will depend on the choice of initial condition for the optimization. This can be avoided by reducing the extent of the nonlinearity captured by the joint distribution by (i) shrinking the joint distribution by providing more precise prior information or more precise experimental data, or (ii) re-parameterizing the model to reduce the strength of the nonlinearity (MacKay, 2003, Chapter 27).

In many cases, this approximation will be justifiable, given the substantial reduction in computational cost compared to sampling methods, which are the only viable alternative for constructing the posterior. In previous work, we compared the computational cost of our framework to two sampling approaches (Yoko and Juniper, 2024). The comparison was done on a computationally cheap thermoacoustic network model, similar to the one used in this study. Applying our framework to this model, we can compute the posterior probability of five unknown parameters in under 5 s on a single core on a laptop. The same inference problem takes 35 CPU hours running on a workstation when solved with Markov Chain Monte Carlo, and 22 CPU hours when solved with Hamiltonian Monte Carlo.

3.3.3. Uncertainty propagation

We have quantified the uncertainty in the parameter values, but we are also interested in how the parametric uncertainty affects the model predictions. We quantify the prediction uncertainty by propagating the parameter uncertainty through the model. This is done cheaply by linearizing the model around \mathbf{a}_{MP} and propagating the uncertainties through the linear model:

$$\mathbf{C}_{s_i} = \mathbf{J}_i^T \mathbf{C}_a \mathbf{J}_i, \tag{6}$$

where \mathbf{C}_{s_i} is the covariance matrix describing the uncertainty in the model predictions at $\mathbf{s}_i = \mathcal{H}_j(\mathbf{x}_i, \mathbf{a}_{MP})$. The marginal uncertainty of each predicted variable is given by the diagonal elements of \mathbf{C}_{s_i} , because the prediction uncertainties are Gaussian.

3.3.4. Model comparison

In many cases, we are faced with a set of candidate models and are required to identify which model is the best. In the Bayesian framework, we can compare candidate models by calculating the posterior probability of each model, given the data:

$$p(\mathcal{H}_j | \mathbf{D}) \propto p(\mathbf{D} | \mathcal{H}_j) p(\mathcal{H}_j). \tag{7}$$

The first factor on the right-hand side of equation (7) is the denominator of equation (3), which is referred to as the marginal likelihood (ML). The second factor is the prior probability that we assign to each model. If we have no reason to prefer one model over another, we assign equal prior probabilities to all models and rank them according to their ML. The ML is calculated by integrating the numerator of equation (3) over parameter space. For a Gaussian posterior, this integral is approximated as

$$p(\mathbf{D} | \mathcal{H}_i) \approx p(\mathbf{D} | \mathbf{a}_{MP}, \mathcal{H}_i) \times p(\mathbf{a}_{MP} | \mathcal{H}_i) |\mathbf{C}_a^{MP}|^{1/2}. \tag{8}$$

The ML can be broken down into two components, which provide insight into the model comparison. The first factor on the right-hand side of equation (8) is referred to as the best fit likelihood (BFL), which is a measure of how well the model fits the data. The second factor, called the Occam factor (OF), penalizes the model based on its complexity, where the complexity is measured by how precisely the parameter values must be tuned for the model to fit the data to within the experimental uncertainty. A model with a large BFL fits the data well, and vice-versa. A model with a large OF has low parametric complexity, and is unlikely to over-fit the data. Conversely, a model with a small OF is likely to over-fit the data. Therefore, the model with the largest ML is the simplest model that is capable of describing the data, for given

measurement error and given priors. The Bayesian model comparison framework therefore naturally enforces Occam's razor to select the best model.

4. Bayesian optimal experiment design

The framework described in Section 3 allows the modeler to benefit from the work of the experimentalist by using experimental observations to improve the accuracy of physics-based models. We now extend this framework to enable the modeler to inform the experimental design process in order to minimize the experimental cost. In doing so, we show that, when applying a Bayesian framework to physics-based problems, there is a strong mutually beneficial relationship between experimentation and modeling.

There are a number of ways in which the modeler could inform the experiment design. In this study, we answer three specific questions an experimentalist may face: (i) which experimental design and operating point would provide the maximum information about the unknown parameters, (ii) where should the sensors be placed to provide the maximum information about the unknown parameters, and (iii) which experimental design would maximize the discrimination between candidate models?

Each of these questions can be answered by using metrics from information theory to quantify the information content of a candidate experiment. We follow the general approach proposed by Lindley (1956), which has been applied to many other optimal experiment design studies (Ryan, 2003; Loredo, 2004; Huan and Marzouk, 2013; Verma et al., 2019). For each of the experimental questions listed above, we define a suitable information-based utility function, $u(\mathbf{x}, \mathbf{z}, \mathbf{a})$. We then calculate the *expected* utility by integrating over all possible realizations of the parameters and data:

$$U(\mathbf{x}) = \int_{\mathbf{Z}} \int_{\mathbf{A}} u(\mathbf{x}, \mathbf{z}, \mathbf{a}) p(\mathbf{a}|\mathbf{x}, \mathbf{z}) p(\mathbf{z}|\mathbf{x}) d\mathbf{a} d\mathbf{z}, \quad (9)$$

where $U(\mathbf{x})$ is the expected utility, $u(\mathbf{x}, \mathbf{z}, \mathbf{a})$ is the utility function, and \mathbf{Z} and \mathbf{A} are the support of $p(\mathbf{z}|\mathbf{x})$ and $p(\mathbf{a}|\mathbf{x}, \mathbf{z})$, respectively, that is, the set of \mathbf{z} and \mathbf{a} that can occur with nonzero probability.

Previous studies have faced difficulties with the cost of computing the expected utility within their inference frameworks, which typically requires Monte Carlo integration over high-dimensional spaces (Ryan, 2003; Loredo, 2004; Huan and Marzouk, 2013; Verma et al., 2019). We will show that in the adjoint-accelerated Bayesian inference framework, the expected utility can be computed cheaply.

4.1. Optimal design for parameter inference

We consider the situation where we have collected and assimilated a set of data $\mathbf{D}_i = \{\mathbf{z}_1, \dots, \mathbf{z}_i\}$, where i could be zero.² We want to know which experiment to perform next in order to gain maximal information about the unknown parameters. In this case, a suitable utility function is the information content of the next experiment, which we can calculate as the change in Shannon entropy between the prior and posterior parameter distributions (Lindley, 1956). The Shannon entropy of a probability distribution is defined as

$$S_i = - \int_{\mathbf{A}} P(\mathbf{a}|\mathbf{D}_i) \log_2(p(\mathbf{a}|\mathbf{D}_i)) d\mathbf{a}, \quad (10)$$

where S_i is the Shannon entropy of the parameter probability distribution after the i th experiment has been assimilated.³ The utility function, u , which we have chosen to be the information content of experiment $i + 1$, is therefore,

² When $i = 0$, we have not yet collected any data, so \mathbf{D} is an empty set and $p(\mathbf{a}|\mathbf{D})$ is the prior, $p(\mathbf{a})$.

³ For the purpose of optimal experiment design, the choice of the logarithm base is arbitrary. We have chosen to work with base 2 logarithms, which provide information in units of bits, where one bit of information halves the uncertainty in the parameters.

$$\begin{aligned}
 u(\mathbf{x}, \mathbf{z}, \mathbf{a}) &= \Delta S_{i+1} = S_i - S_{i+1} \\
 &= - \int_{\mathbf{A}} P(\mathbf{a}|\mathbf{D}_i) \log_2(p(\mathbf{a}|\mathbf{D}_i)) d\mathbf{a} + \int_{\mathbf{A}} P(\mathbf{a}|\{\mathbf{D}_i, \mathbf{z}_{i+1}\}) \log_2(p(\mathbf{a}|\{\mathbf{D}_i, \mathbf{z}_{i+1}\})) d\mathbf{a}.
 \end{aligned}
 \tag{11}$$

We note that the utility function involves integration over \mathbf{a} , so it is no longer a function of the unknown parameters, that is, for this choice of utility function, $u(\mathbf{a}, \mathbf{x}, \mathbf{z}) = u(\mathbf{x}, \mathbf{z})$. The expected utility therefore simplifies to

$$\begin{aligned}
 U &= \int_{\mathbf{Z}} \int_{\mathbf{A}} u(\mathbf{x}, \mathbf{z}) p(\mathbf{a}|\mathbf{x}, \mathbf{z}) p(\mathbf{z}|\mathbf{x}) d\mathbf{a} d\mathbf{z} \\
 &= \int_{\mathbf{Z}} u(\mathbf{x}, \mathbf{z}) p(\mathbf{z}|\mathbf{x}) d\mathbf{z}.
 \end{aligned}
 \tag{12}$$

To evaluate the remaining integral, we once again exploit Laplace’s method, recalling that the probability distribution over the parameters is always Gaussian in our framework. The difference in Shannon entropy between two Gaussians is given by

$$u(\mathbf{x}, \mathbf{z}) = \Delta S_{i+1} = \frac{1}{2} \log_2 \left(\frac{|\mathbf{C}_{\mathbf{a}}^{-1}|_{i+1}}{|\mathbf{C}_{\mathbf{a}}^{-1}|_i} \right)
 \tag{13}$$

from which we see that maximizing the information gained about the parameters is equivalent to finding the maximum reduction in parameter uncertainty. Recall that, from Laplace’s approximation in [equation \(5\)](#), assimilating the measurement \mathbf{z}_{i+1} updates the parameter uncertainty according to

$$(\mathbf{C}_{\mathbf{a}}^{-1})_{i+1} \approx (\mathbf{C}_{\mathbf{a}}^{-1})_i + \mathbf{J}_{i+1}^T \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{J}_{i+1} + (\mathbf{s}(\mathbf{a})_{i+1} - \mathbf{z}_{i+1})^T \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{H}_{i+1},
 \tag{14}$$

where the model predictions, $\mathbf{s}(\mathbf{a})_{i+1}$, Jacobian, \mathbf{J}_{i+1} , and Hessian, \mathbf{H}_{i+1} , are evaluated with the previous most probable parameters, \mathbf{a}_i , and with the candidate experiment design parameters, \mathbf{x}_{i+1} .

From [equations \(13\) and \(14\)](#), we can see that the choice of which experiment to perform next is dependent on the outcome of that experiment, \mathbf{z}_{i+1} , as one might expect. However, the data-dependence only occurs in the second-order term in [equation \(14\)](#), which is exactly zero for models that are linear in the parameters, and is often small compared to the first-order term for nonlinear models.⁴ We therefore proceed with a first-order approximation, in which the posterior covariance is independent of the experimental outcome:

$$(\mathbf{C}_{\mathbf{a}}^{-1})_{i+1} \approx (\mathbf{C}_{\mathbf{a}}^{-1})_i + \mathbf{J}_{i+1}^T \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{J}_{i+1}.
 \tag{15}$$

If the second-order sensitivities are found to be non-negligible for a model of interest, we may still be able to neglect the data-dependent term on the grounds that the discrepancy, $(\mathbf{s}(\mathbf{a})_{i+1} - \mathbf{z}_{i+1})$, will become small as the model is updated with more data. In this case, we expect that the initial experiments may not be optimal, but that the chosen experiments will become optimal as more experiments are assimilated and the discrepancy decreases. In many cases, this may be an acceptable sacrifice for the computational cost savings of the proposed approach.

By neglecting the data-dependent term, the utility function becomes independent of \mathbf{z} , and we are able to plan the subsequent experiment using only the model and its adjoint:

$$U = \Delta S_{i+1} = \frac{1}{2} \log_2 \left(\frac{(|(\mathbf{C}_{\mathbf{a}}^{-1})_i + \mathbf{J}_{i+1}^T \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{J}_{i+1}|)}{|\mathbf{C}_{\mathbf{a}}^{-1}|_i} \right).
 \tag{16}$$

We follow a greedy sequential experiment design process, where we iteratively (i) use [equation \(16\)](#) to identify the most informative experiment, (ii) perform that experiment to collect the data, \mathbf{z}_{i+1} , (iii) minimize [equation \(4\)](#) to find \mathbf{a}_{i+1} , and (iv) solve [equation \(5\)](#) to find $(\mathbf{C}_{\mathbf{a}})_{i+1}$.

⁴We have used second-order adjoint methods to confirm that this assumption is valid for the thermoacoustic models of interest in this study.

To assist with interpretation of [equation \(16\)](#), we consider the special case of a system with univariate design and observable variables (multiple unknown parameters are still permitted). Under these conditions, the Jacobian, \mathbf{J} , reduces to a column vector, \mathbf{j} , and the measurement covariance, \mathbf{C}_e , reduces to the variance, V_e . The quantity $\mathbf{J}^T \mathbf{C}_e^{-1} \mathbf{J}$ in the numerator of [equation \(16\)](#) reduces to $V_e^{-1} \mathbf{j}^T \mathbf{j}$, which is a rank-one perturbation of the inverse covariance matrix \mathbf{C}_a^{-1} . Using the identities for rank-one perturbations, [equation \(16\)](#) reduces to

$$U = \frac{1}{2} \log_2 (1 + V_e^{-1} \mathbf{j}_{i+1}^T \mathbf{C}_{ai} \mathbf{j}_{i+1}). \quad (17)$$

This recovers the result that MacKay (1992) arrived at for scalar interpolation problems: to maximize information gain we must (i) maximize V_e^{-1} , which is to say that we learn the most when we make precise observations, or (ii) maximize $\mathbf{j}^T \mathbf{C}_a \mathbf{j}$, which is the posterior variance of the model predictions, as shown in [equation \(6\)](#). This produces the intuitive result that we learn the most about the unknown parameters when we perform experiments with the design parameters at which our model is most uncertain. This is similar to the optimal sampling approach commonly used in training Gaussian processes, where the acquisition function is chosen to be the variance in the Gaussian process prediction (Huang et al., 2006; Sengupta and Juniper, 2022).

While this aids with interpretation, MacKay's result is not directly applicable to multivariate systems, unless an ad hoc decision is made about which variable's uncertainty should drive the decision of which experiment to perform next. This is avoided in our framework by maximizing the change in Shannon entropy, which automatically balances the information gained from each observed variable based on how sensitive it is to the unknown parameters.

4.2. Optimal sensor placement for parameter inference

We now consider the case where one or more of the observed variables, \mathbf{z} , are spatially varying and are observed with point measurement sensors. We would like to know where to place the sensors in order to gain as much information as possible about the unknown parameters. We may have an existing rig with a fixed number of sensors, and we would like to know where to place the sensors to gain maximal information. Alternatively, we may be designing a new rig, for which we already have a qualitative model, and we would like to know (i) how many sensors we need to buy, and (ii) where we should make provision for instrument access.

To answer these questions, we add the sensor locations to the vector of design parameters, \mathbf{x} , and find the design parameters that maximize the change in Shannon entropy in the same way as in [Section 4.1](#). We place the sensors sequentially, with each sensor location selected to provide maximum information based on (i) the model and (ii) the information from the existing sensor layout.

This process naturally accounts for the local reduction in information in the vicinity of existing sensors, with the correlation length determined by the model sensitivity. This removes the need to define ad hoc methods to avoid sensor clustering, as done in previous studies on sparse sensor placement (Papadimitriou and Papadimitriou, 2015; Bidar et al., 2024).

4.3. Optimal design for model comparison

Finally, we consider the case where we are trying to identify the best model from a set of candidate models $\mathcal{H}_{j,j} = \{1, 2, \dots, M\}$. We assume that we have already performed the optimal experiments to learn the unknown parameters of each model. We now want to identify the optimal experiment design, \mathbf{x}_{i+1} , which maximizes the discrimination between the candidate models.

At the (as yet undetermined) experiment design \mathbf{x}_{i+1} , each candidate model will make a slightly different prediction, $\mathbf{s}_{i+1,j}$, with slightly different uncertainty, $(\mathbf{C}_s)_{i+1,j}$. Before we make the next observation, \mathbf{z}_{i+1} , each model encodes a belief that the data will occur with a probability distribution given by a Gaussian centered around the model prediction:

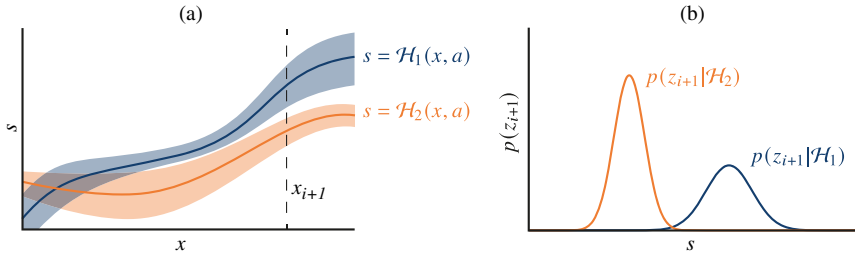


Figure 5. (a) At each candidate experiment design, x_{i+1} , the two candidate models make slightly different predictions, with different uncertainties. (b) Each model encodes a belief that the next data point will fall within the distribution $p(z_{i+1}|\mathcal{H}_j)$.

$$\begin{aligned}
 P_j &= p(\mathbf{z}_{i+1}|\mathcal{H}_j) = \mathcal{N}(\mu_j, \mathbf{C}_j), \\
 \mu_j &= \mathbf{s}_{i+1,j} = \mathcal{H}_j(\mathbf{x}_{i+1}, \mathbf{a}_{i,j}), \\
 \mathbf{C}_j &= (\mathbf{C}_s)_j + \mathbf{C}_e,
 \end{aligned}
 \tag{18}$$

where \mathbf{C}_s is the prediction uncertainty, which is calculated using equation (6), and we have introduced the shorthand $P_j = p(\mathbf{z}_{i+1}|\mathcal{H}_j)$ to simplify the subsequent notation. This is illustrated for two candidate models in Figure 5.

In order to maximally discriminate between the models, we want to choose \mathbf{x}_{i+1} such that the distributions P_j , where $j = \{1, 2, \dots, M\}$, are as dissimilar as possible. In this way, the models that make poor predictions receive the maximum penalty when the data arrive. We therefore choose our utility function to be a measure of dissimilarity between the distributions. A common measure for the dissimilarity between two distributions is the directed Kullback–Leibler (KL) divergence (Kullback and Leibler, 1951):

$$D_{KL}(P_j \| P_k) = \int_{\mathbf{z}} P_j \log_2 \left(\frac{P_j}{P_k} \right) d\mathbf{z},
 \tag{19}$$

which measures how unexpected the measurement, \mathbf{z}_{i+1} , would be if we used \mathcal{H}_k as our model when \mathcal{H}_j was the better model. This is a suitable measure for our experimental goals, but it only compares two distributions, and we would like to compare multiple models at once. In order to measure the dissimilarity between M distributions, we use the average divergence (Sgarro, 1981) as our utility function:

$$u = \frac{1}{M(M-1)} \sum_{j=1}^M \sum_{k=1}^M D_{KL}(P_j \| P_k),
 \tag{20}$$

which computes the mean directed KL divergence between all pairs of models. The inner summation measures the average rate at which our confidence in each of the models goes to zero if \mathcal{H}_j is the best model. The outer summation averages these rates while proposing each model as the best model. By maximizing this quantity, we find the experiment design for which the measurement would surprise us the most, thereby reducing our confidence in the largest number of bad models at once.

As before, evaluating the expected utility requires the solution of multiple high-dimensional integrals. Substituting Gaussian distributions for P_j and P_k , and following a similar analysis to Section 4.1, we find that the expected utility reduces to

$$U = \frac{1}{2M(M-1)} \sum_{j=1}^M \sum_{k=1}^M \left(\log_2 \left(\frac{|\mathbf{C}_j|}{|\mathbf{C}_k|} \right) - d + \text{tr}(\mathbf{C}_k^{-1} \mathbf{C}_j) + (\mu_j - \mu_k)^T \mathbf{C}_k^{-1} (\mu_j - \mu_k) \right),
 \tag{21}$$

where d is the number of observed variables. We once again find that the integrals, which typically make optimal experiment design expensive, become trivial within our framework. We are therefore able to

identify optimal experiments using only the model and its adjoint, and follow a greedy selection process as in the previous sections.

To gain further insight into the expected utility, we consider the case of comparing two candidate models with univariate design and observable variables. Under these conditions, [equation \(21\)](#) reduces to the symmetric KL divergence between a pair of univariate Gaussians:

$$u = \frac{1}{2} \left\{ \left(\frac{1}{V_1} + \frac{1}{V_2} \right) (\mu_1 - \mu_2)^2 + \frac{(V_1 - V_2)^2}{V_1 V_2} \right\}, \quad (22)$$

where μ_j and V_j are the expected value and variance of the distributions $p(\mathbf{z}_{i+1} | \mathcal{H}_j)$. As in [Section 4.1](#), we recover the result [MacKay \(1992\)](#) arrived at for scalar interpolation problems: to maximize the discrimination between models we must (i) gather data where the model predictions maximally disagree, measured relative to the confidence in their predictions, and (ii) gather data where the confidence in the models is maximally different. The first result serves to maximize the best-fit-likelihood reward on the model which fits the new data point better. The second result, which is perhaps less intuitive than the first, serves to maximize the Occam penalty on the model with more parametric flexibility.

5. Results

We demonstrate the three applications of Bayesian experiment design described in [Sections 4.1–4.3](#) by applying them to the electrically heated Rijke tube. We have already collected a densely sampled dataset from this system, which we have previously used for Bayesian parameter inference and model comparison ([Juniper and Yoko, 2022](#)).

For each of the experimental questions posed in [Section 4](#), we demonstrate Bayesian experiment design by recursively selecting the optimal experiment from the densely sampled set. This allows us to compare the model predictions to the full dataset of both observed and unobserved data. The unobserved data act as a validation set, which is hidden from the model and the assimilation process. This allows us to demonstrate that in each case we can achieve our experimental goals using only a few well-chosen experimental observations.

In the original dataset ([Juniper and Yoko, 2022](#)), we perform experiments with the rig in four conditions. First, we study the empty tube so that we can infer the model parameters for (i) the end reflection coefficients, and (ii) the strength of the visco-thermal dissipation of the boundary layer on the tube walls. Second, we traverse the heater support prongs through the tube with the heater removed. From this data, we infer the model parameters for the strength of the visco-thermal dissipation in the boundary layer on the heater prongs. Third, we traverse the cold heater through the tube to infer the model parameters for the strength of the visco-thermal dissipation of the heater. Finally, we traverse the heater through the tube and sweep through several heater powers to infer the model parameters for (i) the fluctuating heat release rate models, and (ii) how the visco-thermal dissipation changes with heater power. At each of these four stages, we assume that the parameters inferred at the previous stages remain constant, and so we fix them in the model.

5.1. Optimal design for parameter inference

We first demonstrate the optimal experiment design framework on the simple sub-problem of inferring the visco-thermal dissipation of the cold heating element using as few experimental observations as possible. We then apply the method to the more complex problem of inferring the model parameters for the case when the heater is switched on.

5.1.1. Assimilating heater dissipation from cold experiments

When traversing the cold heater through the tube, the vector of design parameters, \mathbf{x} , contains only the position of the heater within the tube. The measurement vector, \mathbf{z} , contains the complex eigenvalue whose

real part is the growth rate and imaginary part is the natural frequency of oscillations. We assimilate this data into a model with two complex parameters, \mathbf{a} , which are used to model the strength of (i) the viscous dissipation, and (ii) the thermal dissipation of the heating element. In this section, we assume that we have already performed and assimilated the optimal experiments to infer the reflection coefficients, and the strength of the visco-thermal dissipation on the tube walls and heater support prongs.

In Figure 6(a.i), (a.ii), we compare the predictions of the prior model to the experimental data, which has not yet been assimilated. As expected, the prior model performs poorly, because the prior parameter values are inaccurate. However, we have specified correspondingly large uncertainties in the prior parameters, so the model predictions have appropriately large error bars, which extend beyond the limits of the y-axis.

Figure 6(a.iii) shows the potential information gain from an experiment performed at each heater position, as estimated using equation (16). This reveals that the most information can be gained by collecting data at the ends of the tube. This is an intuitive result because it places the heater near the point

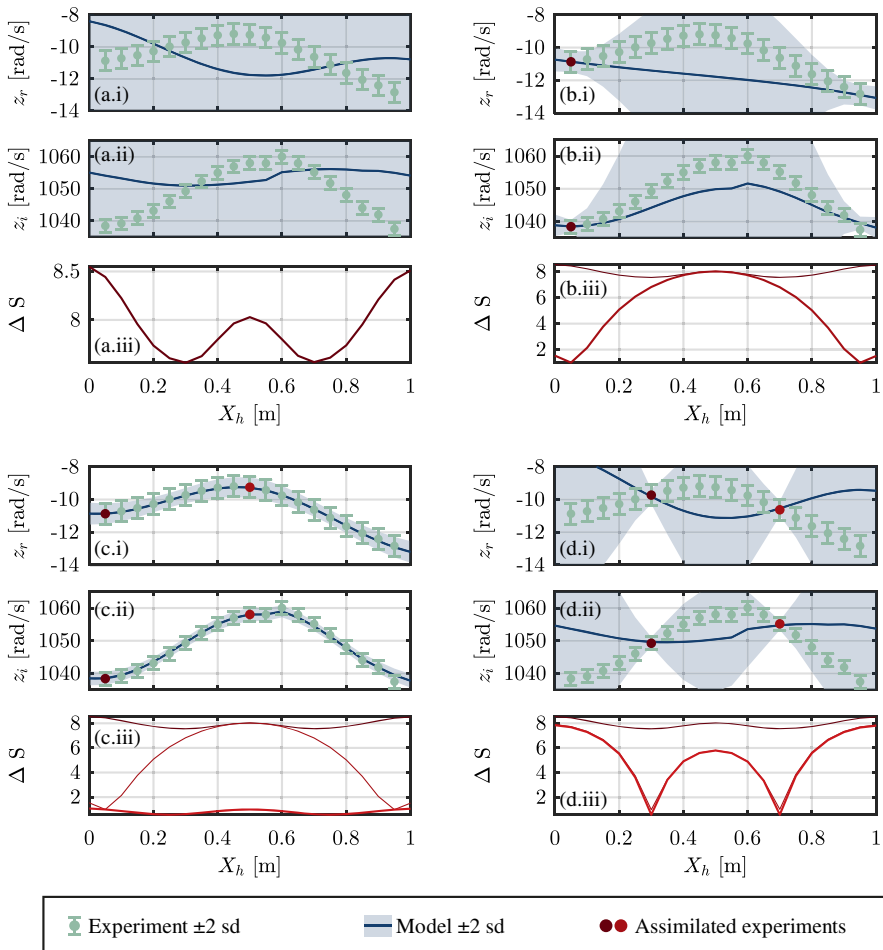


Figure 6. Three steps of active data selection comparing experimental data to model predictions after assimilating: (a) no data, (b) the first datapoint with maximum information content, and (c) the second datapoint with maximum information content. For each step, we show (i) growth rate, z_r , (ii) angular frequency, z_i , and (iii) information content, ΔS , plotted against heater position, X_h . For comparison, we also show (d) the result from assimilating the two experiments with minimum information content.

of maximum acoustic velocity fluctuations, and therefore maximum viscous dissipation. Furthermore, the acoustic pressure (and therefore temperature) fluctuations approach zero at the boundaries, so we expect the thermal dissipation to approach zero there. With this experiment, we can therefore easily infer the parameters controlling viscous dissipation without being influenced by effects of thermal dissipation.

Because the model is symmetric in X_h with respect to the unknown parameters,⁵ the information gain at $X_h = 0.05$ and $X_h = 0.95$ is equal, so either experiment is a viable candidate experiment. We arbitrarily choose to assimilate the measurement taken at $X_h = 0.05$, which updates the parameter probability distribution. The new model predictions are plotted in Figure 6(b.i), (b.ii). We see that, near the assimilated data point, the model is now accurate and the uncertainty has collapsed to the uncertainty in the data. Away from the assimilated data point, however, the model predictions deviate from the data and the uncertainty increases. We also see that, because the symmetry of the problem is encoded in our physics-based model, we have also learned about the system response when the heater is placed at the opposite end of the tube.

Looking at the expected information gain in Figure 6(b.iii), we see that we do not expect to learn much more by conducting another experiment near the ends of the tube. However, there is still information to be gained from an experiment conducted at the center of the tube. This experiment places the heater in the acoustic pressure (and therefore temperature) anti-node and velocity node, making thermal dissipation maximum and viscous dissipation zero. With this experiment, we easily infer the parameters controlling thermal dissipation, in the absence of viscous dissipation.

We assimilate the measurement taken at $X_h = 0.5$ and plot the new model predictions in Figure 6(c.i), (c.ii). This shows that, after assimilating just these two data points, the model matches the experimental data for all heater positions, and the model uncertainty is now similar in magnitude to the measurement uncertainty at all heater positions. The expected information gain, the bold red line in Figure 6(c.iii), is now relatively flat and close to zero at all heater positions. This shows that no experiment will be substantially more informative than any other, and there is therefore little information to be gained by assimilating more data. We have therefore gained all meaningful information about the unknown parameters using just two carefully chosen experimental observations. Although a skilled experimentalist may be able to identify the most informative experiments from physical insight alone, this becomes more difficult when assimilating data into more complex models containing more physical mechanisms and more model parameters.

An expected value of two parameters can, of course, be obtained from any two observations. Their posterior variances, however, depend strongly on which two observations are selected. To demonstrate this, we assimilate the two data points with the lowest information content, as plotted in Figure 6(d.i)–(d.iii). The results clearly show that assimilating these two points only achieves local accuracy and local certainty in the model predictions.

As a final demonstration of the effectiveness of this approach, we sequentially assimilate all the available data, and record the information gained as each experiment is assimilated. Figure 7(a) shows the Shannon entropy plotted against the number of experiments assimilated, while Figure 7(b) shows the change in Shannon entropy against the number of experiments assimilated. We compare the rate of information gain achieved by 1,000 random experiment designs to the limit cases in which we recursively select the most informative and least informative experiments.

We see from Figure 7(a) that assimilating the best available experiment at each step causes the Shannon entropy to reduce rapidly with the first two observations, that is, we quickly become confident in the most probable value of the unknown parameters. After the second experiment is assimilated, the model uncertainty reduces below the experimental uncertainty, and further experiments are relatively uninformative. By contrast, assimilating the worst experiments results in a gradual reduction in uncertainty. These limit cases bound the rate of information gain achieved by the 1,000 random experiment designs, indicating that our methodology is likely optimal in this case. If experiments are designed randomly, at

⁵ The heater dissipation has equal effect on either side of the tube centerline. The model predictions themselves are not symmetric, because the heater prongs break the symmetry (as X_h increases, the length of prong inserted into the tube increases). The model parameters for the visco-thermal dissipation of the prongs are inferred prior to conducting the heater traverse experiments.

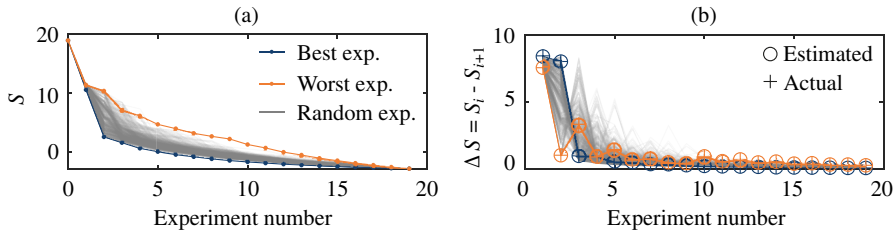


Figure 7. Comparison of learning rate for three experiment design strategies: (blue) sequentially performing the most informative experiments, (orange) sequentially performing the least informative experiments, (gray) 1,000 instances of sequentially performing random experiments. Plot (a) shows how the Shannon entropy of the parameter probability distribution decreases as additional experiments are assimilated. Plot (b) shows the information gained from each experiment, which is given by the change in Shannon entropy. We show the information gain estimated before the data are assimilated using equation (16) (+), as well as the actual achieved information gain, calculated after the experiment is assimilated (\circ).

least seven experiments are needed to be 95% confident that you will gain the same amount of information as the two optimal experiments.

This is reinforced by Figure 7(b), which shows that by assimilating the best experiments we gain significant information from the first two observations, and negligible information thereafter. By comparison, when we assimilate the worst experiments we gain significant information from the first observation, and then a small but not negligible amount of information throughout the remainder of the dataset. We see that the information gained by assimilating the worst experiments tends to oscillate between a small information gain, followed by a negligible information gain. This is because once an experiment is assimilated, its symmetric pair becomes the next least informative experiment, and carries almost no new information. If experiments are designed randomly, it is likely that most of the information will be gained within the first seven experiments, following which it is unlikely that further information will be gained.

In Figure 7(b) we also compare the information gain estimated using equation (16) against the actual information gain, which is calculated from the Shannon entropy after the experiment is assimilated. We see that the values are almost indistinguishable, so in this case there is no penalty from neglecting the data-dependent term in equation (5), because the second-order parameter sensitivities are indeed small.

5.1.2. Assimilating a heat release rate model from hot experiments

We apply the same approach to experiments with the heater turned on. In this case, the design parameter vector, \mathbf{x} , includes the heater position and heater power, while the measurement vector, \mathbf{z} , contains the complex eigenvalue. The model includes three complex parameters, \mathbf{a} , one of which models the fluctuating heat release rate at the heater, and the remaining two model how the heater's visco-thermal dissipation changes with heater power.

In Figure 8(a.i), (a.ii), we compare the predictions of the prior model to the experimental data, which have not been assimilated yet. Once again, we see that the prior model performs poorly, because the prior parameter values are inaccurate, but that the model predictions have appropriately large error bars.

Figure 8(a.iii) shows the potential information gain for each experiment, as estimated using equation (16). We see that we can gain the most information by performing experiments at higher heater powers. This is to be expected, because the three unknown parameters model effects that are linear in heater power, so they are most easily observed at high heater powers. We also see that the information gain peaks at heater positions that produce the strongest thermoacoustic effect (where the magnitude of the product of acoustic pressure and velocity is maximum). By contrast, placing the heater at the acoustic velocity node (near the center of the tube) yields very little information. This is because both the thermoacoustic effect

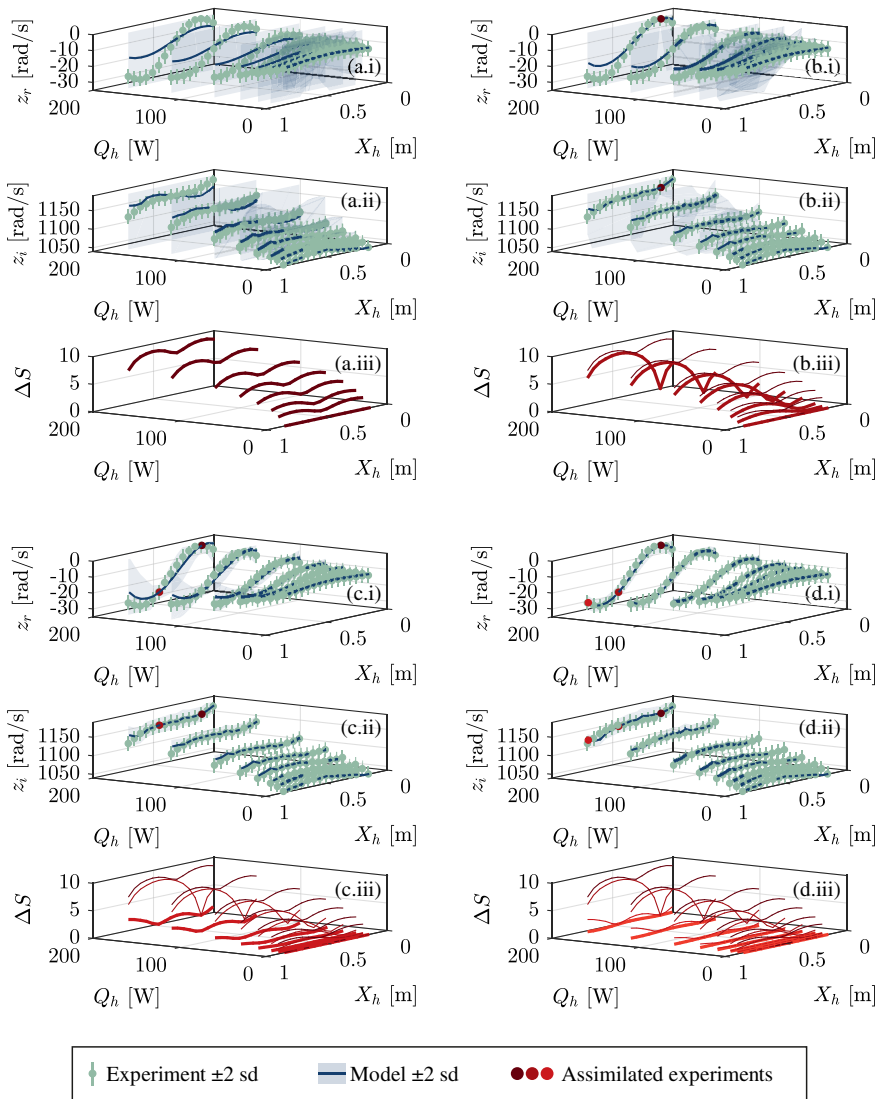


Figure 8. Four steps of active data selection for assimilating data from experiments with the heater active. We compare experimental data to model predictions after assimilating: (a) no data, and (b–d) the first, second, and third observation with maximum information content. For each step, we show (i) growth rate, z_r , (ii) angular frequency, z_i , and (iii) information content, ΔS , plotted against heater position, X_h , and heater power, Q_h .

and the viscous dissipation are zero at the velocity node, so an experiment performed here only provides information about the variation of thermal dissipation with heater power.

We assimilate the best experiment and plot the results in Figure 8(b.i), (b.ii). We see, once again, that the model becomes locally accurate and certain, and that the accuracy and certainty reduce with distance from the assimilated experiment. Unlike in Section 5.1.1, the symmetry of the underlying physics is broken by the steady and fluctuating heat release from the heater. As a result, making an observation with the heater placed at one end of the tube no longer tells us the behavior of the system when the heater is placed at the other end of the tube. However, we see that making an observation at the highest heater power has also provided information about the system’s behavior at lower heater powers. The model has become

confident in its predictions about the system behavior when the heater is in the same position, but operating at lower power outputs.

We see from [Figure 8\(b.iii\)](#) that the first observation reduces the information we expect to gain from subsequent experiments performed at nearby heater positions, at all heater powers. An experiment performed near the downstream end of the tube now carries slightly less information than it did before the first observation was made. This implies that, while the physics is not fully symmetrical, some of the phenomena may be, and so we have gained information about the behavior of the system on one end of the tube through an observation at the other end.

The model predictions after assimilating the second most informative experiment are shown in [Figure 8\(c.i\)](#), [\(c.ii\)](#). The second observation causes a more widespread collapse in uncertainty, although the model is still significantly more certain at heater positions where observations have been made. The uncertainty is not below experimental uncertainty for all candidate design parameters, so we expect we could still learn more through additional observations.

This is confirmed through [Figure 8\(c.iii\)](#), which shows that the expected information gain has reduced for all heater positions and heater powers. However, there is still more information to be gained from a further experiment performed at the downstream end of the tube (near $X_h = 1$), at the highest heater power.

Assimilating the third most informative experiment yields the model predictions shown in [Figure 8\(d.i\)](#), [\(d.ii\)](#). The model is now accurate at all pairs of design parameters, regardless of whether the data were assimilated. Furthermore, the model uncertainty is below experimental uncertainty everywhere.

We see from [Figure 8\(d.iii\)](#) that the final plot of expected information gain is relatively flat, so there is no clear candidate for further experiments. Furthermore, the magnitude of expected information gain is small for all further experiments, so there is little value in performing further experiments for the purpose of parameter inference. In the initial dataset, we collected data at 120 operating points, but we now see that three carefully selected operating points are sufficient, which is just 2.5% of the full dataset.

While the chosen experiments are somewhat intuitive, even a skilled experimentalist might struggle to select the three optimal experiments in this case. Even though the model remains relatively simple, the task of selecting optimal experiments for learning the values of multiple parameters with coupled effects becomes increasingly challenging as the number of parameters increases.

We now contrast the learning rate achieved by 1,000 random experiment designs against recursively selecting the best and worst experiments, shown in [Figure 9](#). [Figure 9\(a\)](#) shows that selecting the best experiments leads to a rapid reduction of Shannon entropy, corresponding to a rapid learning of the parameter values. By comparison, selecting the worst experiments leads to a very gradual reduction in Shannon entropy, corresponding to gradual learning of the parameter values. Most strikingly, we see that selecting the worst experiments begins by using a set of experiments which carry no information at all. These correspond to the experiments performed with zero heater power, which naturally provide no information about (i) the thermoacoustic effect, or (ii) how the heater visco-thermal dissipation changes with heater power.

If the experiments are chosen randomly, the learning rate falls between that achieved by selecting the best or worst experiments. In contrast to [Figure 7](#), we see that the random designs are not tightly bounded by the best and worst designs.⁶ This is because the larger experimental design space makes it considerably less likely that very good or very bad experiments will be selected at random. This is still a relatively small design space, with only two design variables. For more complex problems, it will become increasingly difficult to choose good experiments through either physical insight or random chance.

[Figure 9\(b\)](#) shows that the three best experiments provide significant information gain, following which we gain little information from the subsequent experiments. Selecting the worst experiments yields a small information gain throughout the dataset, apart from the first few cold experiments, which provide

⁶ There are 9.7×10^{34} unique experiment designs containing 60 of the 120 experiments. Our sample of 1,000 random designs only explores a fraction of these, so the less probable outcomes are not seen.

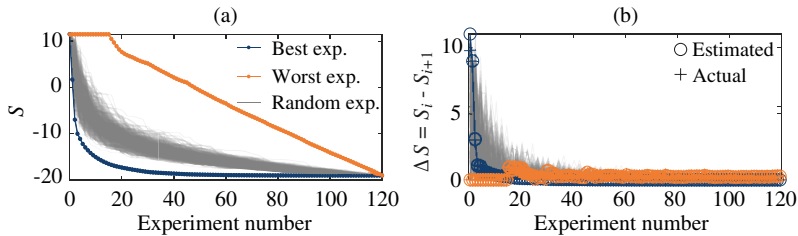


Figure 9. Comparison of learning rate for three experiment design strategies: (blue) sequentially performing the most informative experiments, (orange) sequentially performing the least informative experiments, (gray) 1,000 instances of sequentially performing random experiments. Plot (a) shows how the Shannon entropy of the parameter probability distribution decreases as additional data are assimilated. Plot (b) shows the information gained from each experiment, which is quantified from the change in Shannon entropy before and after the data were assimilated. We show the information gain estimated before the data are assimilated, using equation (16) (+), as well as the actual achieved information gain, calculated after the data are assimilated (○).

zero information. If the experiments are chosen randomly, at least 36 experiments are needed to be 95% confident that you will gain the same amount of information as the three optimal experiments.

Comparing the estimated and achieved information gains, we see that the first best observation provided more information than expected. The under-prediction is due to the neglected data-dependent term in equation (5), which is small but not negligible in this case. After assimilating just one observation, however, the data-discrepancy reduces such that the data-dependent term becomes negligible for subsequent observations.

5.2. Optimal sensor placement for parameter inference

We now demonstrate the application of optimal sensor placement to our thermoacoustic rig. Specifically, we search for the optimal sensor placement to determine the characteristics of the empty tube. In this case, the design parameter vector, \mathbf{x} , contains the sensor locations, while the measurement vector, \mathbf{z} , contains the eigenvalue and the complex pressure at the sensor locations. There are three unknown parameters, \mathbf{a} : two complex parameters corresponding to the upstream and downstream reflection coefficients, and a third real parameter corresponding to the strength of the visco-thermal dissipation.

In Section 5.1, we assimilated only growth rate and natural frequency data. In this case, however, the growth rate and natural frequency data does not provide enough information to infer unique values for all five parameters. Instead, we learn a range of parameter sets that can describe the data acceptably well. In some cases, this may reduce the accuracy of the model when extrapolating to unseen configurations. To disentangle these parameters, we need to add more information, which comes from our observations of the complex pressure at multiple points along the length of the tube. The data are collected with a redundant array of microphones, allowing us to demonstrate optimal sensor placement by assimilating only data from the most informative microphones, and identifying which microphones are redundant.

As will be seen shortly, for this demonstration we need precise measurements of the phase of the pressure eigenmode. The microphone array used to collect the data could not, however, be phase-calibrated to the accuracy that would be required for this demonstration. We must therefore artificially phase-calibrate the microphones using the model output as ground truth. This effectively renders this demonstration a partially synthetic experiment.

Figure 10(a.i), (a.ii) shows the model predictions of the pressure eigenmode, after the growth rate and natural frequency data have been assimilated. We see that assimilating the growth rate and frequency allow us to estimate the unknown parameters well enough to produce reasonable predictions about the pressure eigenmode. However, the uncertainty in the pressure predictions is large because we have not yet gained enough information about the parameter values. The posterior joint distributions between each pair

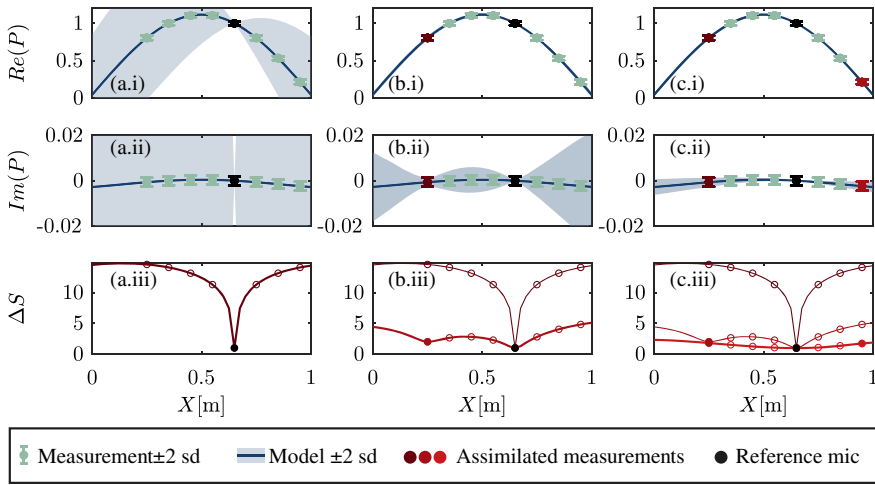


Figure 10. Three stages of optimal sensor placement: (a) reference mic only, (b) one additional mic, and (c) two additional mics. Figures show (i) the real component of the pressure, $Re(P)$, vs. axial position in the tube, X , (ii) the imaginary component of pressure, $Im(P)$, and (iii) the expected information gain, ΔS , from a microphone placed at any axial position. Predictions are plotted as solid lines, with uncertainties indicated with shaded regions. Available microphone data are plotted in teal in (i, ii), and as open circles in (iii). Assimilated microphone data are colored with the appropriate shade of red.

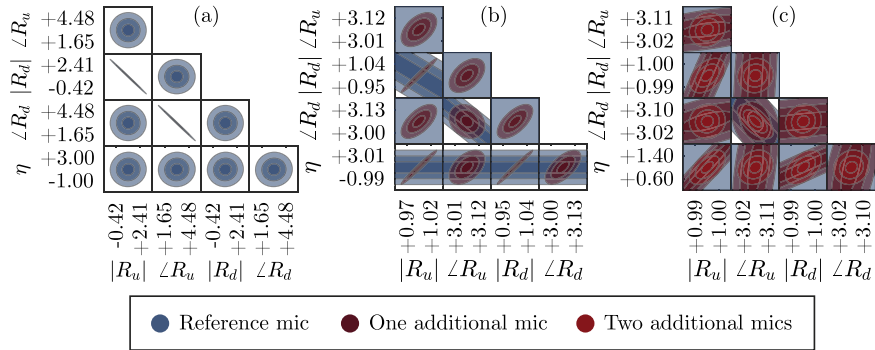


Figure 11. Posterior joint probability distributions after three stages of optimal sensor placement: (a) reference mic only, (b) one additional mic, and (c) two additional mics. The joint distribution between pairs of parameters is indicated in each frame using contours of one, two, and three standard deviations from the mean. The parameters are the absolute value and angle of the upstream, R_u and downstream reflection coefficients, R_d , and the boundary layer dissipation strength, η . The axes are labeled with the ± 3 standard deviation bounds.

of parameters are shown in Figure 11(a), from which we see that there is a strong correlation between the values of $|R_u|$ and $|R_d|$, and between $\angle R_u$ and $\angle R_d$.

Figure 10(a.iii) shows the information that is expected to be gained from a microphone placed at any position along the length of the tube. The locations where microphone data are available are marked with circles. This reveals that there is no microphone installed in the optimal location. Fortunately, however, the best available microphone carries only slightly less information than a microphone placed at the optimal location.

We assimilate the data from the best available microphone and plot the new model predictions in Figure 10(b.i), (b.ii). We see from Figure 10(b.i) that the real component of pressure is now well

defined, with error bars that aren't distinguishable on the scale of the plot. Adding a second microphone allows us to identify a unique pair of values for $\angle R_u$ and $\angle R_d$, collapsing the uncertainty in the prediction of $Re(P)$. This is reflected in Figure 11(b), where we see that the correlation between $\angle R_u$ and $\angle R_d$ has been removed. The uncertainty in the model's predictions of the imaginary part of pressure, shown in Figure 10(b.ii), has also reduced significantly. However, there is still a three-way correlation between $|R_u|$, $|R_d|$, and η , seen in Figure 11(b), which requires more information to further reduce the uncertainty. We see from Figure 10(b.iii) that we expect to gain further information from a sensor placed near the downstream end of the tube.

After assimilating the data from the third microphone, the uncertainty in the pressure eigenmode prediction reduces below experimental uncertainty throughout the domain, as seen in Figure 10(c.i), (c.ii). We therefore see that, out of the array of eight microphones installed in the rig, five were redundant. Furthermore, the installed microphone locations were not optimal, and more information could have been gained if the upstream microphone had been installed closer to the upstream boundary.

We see from Figure 11(c) that using the information from the three best microphones results in a posterior parameter distribution with predominantly uncorrelated parameter values. In other words, we have inferred unique parameter values with small error bars using the minimum number of sensors.

In Figure 12, we show how the prediction uncertainty reduces as microphones are sequentially placed. Figure 12(a) shows the pressure predictions in Figure 12(a.i), (a.ii), and expected information gain in Figure 12(a.iii) as the data from each of the microphones are assimilated in best-to-worst order. Figure 12(b) shows the same process when the microphone data are assimilated in worst-to-best order. We see that assimilating the best microphone data first causes a rapid collapse in uncertainty as the first two additional microphones are added, following which the remaining microphones add negligible information. In comparison, adding the microphones in reverse order leads to a gradual reduction in uncertainty.

This is reinforced by Figure 13 which shows (a) the Shannon entropy and (b) the change in Shannon entropy as data from each subsequent microphone are assimilated. Along with selecting the best and worst

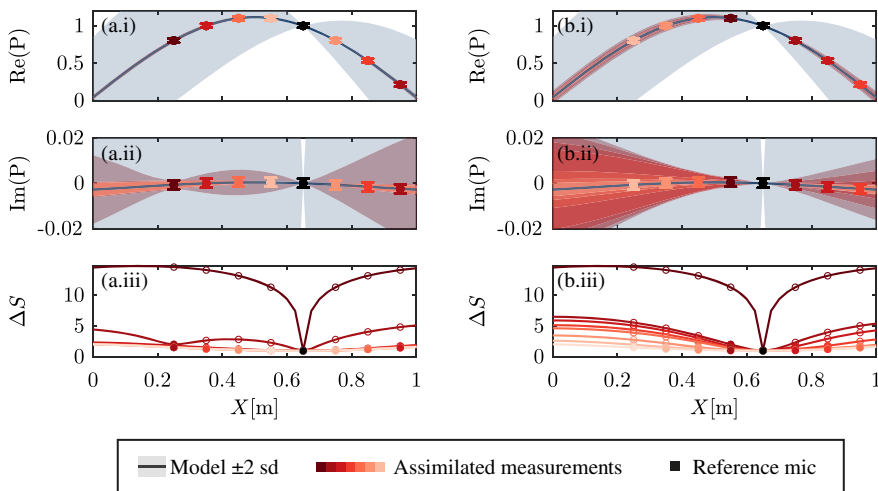


Figure 12. Comparison between sequentially assimilating all available mic data. At each step, we select the (a) best and (b) worst microphones. Figures show (i) the real component of the pressure, $Re(P)$, vs. axial position in the tube, X , (ii) the imaginary component of pressure, $Im(P)$, and (iii) the expected information gain, ΔS , from a microphone placed at any axial position. Predictions are plotted as solid lines, with uncertainties indicated with shaded regions. Uncertainties after assimilating only the reference mic are shaded in blue. Uncertainties after assimilating additional mics are colored with shades of red from dark (fewest additional measurements) to light (most additional measurements).

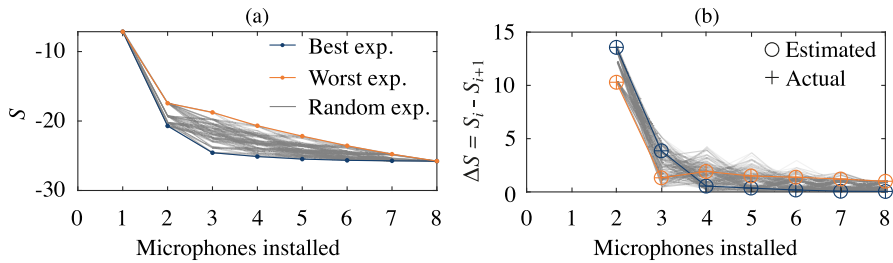


Figure 13. Comparison of learning rate for three sensor placement strategies: (blue) sequentially placing the sensors in the best locations, (orange) sequentially placing the sensors in the worst locations, (gray) 1,000 instances of sequentially placing the sensors in random locations. Plot (a) shows how the Shannon entropy of the parameter probability distribution decreases as additional mic data are assimilated. Plot (b) shows the information gained from each microphone, which is quantified from the change in Shannon entropy before and after the mic data were assimilated. We show the information gain estimated before the data are assimilated, using equation (16) (+), as well as the actual achieved information gain, calculated after the data are assimilated (\circ).

microphone positions, we also show the results from 1,000 random sequential sensor placements. Figure 13(a) shows that assimilating the best microphone data first reduces the Shannon entropy more rapidly than assimilating the worst microphone data first, and placing the sensors randomly falls between these two cases. Figure 13(b) confirms that the two best additional microphones provide significant amount of information, following which the subsequent microphones provide negligible information. In comparison, when assimilating the worst microphones first, the first additional microphone provides a significant amount of information, following which each subsequent microphone contributes a small, but not negligible, amount of information.

5.3. Optimal design for model comparison

We now consider the third task of identifying the experiment designs that allow us to most easily discriminate between multiple candidate models. We demonstrate this using the hot data previously described in Section 5.1.2. We introduce two additional models for the purpose of comparison with the baseline model, which was trained in Section 5.1.2. The first additional model, which we label model A, is a physics-based model with additional parameters that allow the heater feedback to vary with heater position. There is no physical basis for this, so we expect the model comparison to favor the baseline model. The second additional model, which we label model B, is a polynomial surface with the minimum degree capable of describing the data. This model disregards the physics entirely, so we once again expect the baseline model to be favored. The baseline model has 6 unknown parameters, model A has 14 unknown parameters, and model B has 16 unknown parameters.

We begin by assimilating a small subset of the experiments to estimate the parameters of each of the models so that the models make similar predictions, making model comparison nontrivial. We select the optimal experiments for learning each model's parameters using the process described in Section 4.1, then combine these optimal experiments and train all three models on the same set of initial experiments. We see in Figure 14 that the predictions of both model A and model B are comparable to the baseline model's predictions. This makes it difficult to identify the best model by inspection. We would like to use Bayesian model comparison to identify the best model, so we sequentially identify and assimilate additional experiments that maximize the discrimination between the candidate models. For clarity, we demonstrate this by separately comparing model A to the baseline model, and B to the baseline model. The utility function defined in Section 4.3 can, however, be used to identify the optimal experiments for distinguishing between arbitrarily many models at once.

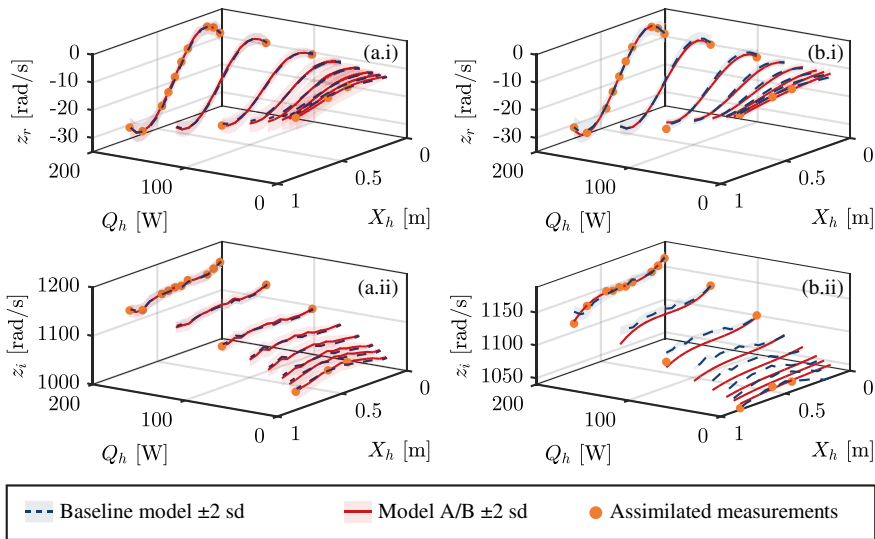


Figure 14. Predictions produced by three candidate models. We compare (i) growth rate, z_r , and (ii) angular frequency, z_i predictions produced by the baseline model (blue) against those produced by (a) model A (red) and (b) model B (red). The initial experiments selected to train the models are shown with orange markers.

We sequentially assimilate each experiment in the dataset, selecting both the best and worst experiments for model comparison. The results are shown in Figure 15(a) for comparing model A against the baseline model, and in Figure 15(b) for comparing model B against the baseline model. We see from Figure 15(a.i) that model A is only slightly less probable than the baseline, while Figure 15(b.i) shows that model B is substantially less probable than the baseline.

Comparing the BFLs between model A and the baseline model, shown in Figure 15(a.ii), we see that they are essentially identical. This tells us that both models fit the data equally well, which is evident in Figure 14(a). This should be expected because model A is identical to the baseline model but has additional unnecessary parameters. The penalty of these additional parameters is seen by comparing the OFs in Figure 15(a.iii), which shows that model A has a very small OF (very negative $\log(\text{OF})$). The smaller OF penalizes model A, resulting in a smaller ML.

As mentioned in Section 4.3, when choosing optimal experiments to discriminate between two models, we choose experiments where (i) the model predictions maximally disagree, and (ii) the model uncertainties maximally disagree. Choice (i) results in a penalty in the BFL of the model with the poorer fit of the new data point, and choice (ii) results in a penalty in the OF on the more flexible model. With the predictions of model A being almost identical to those of the baseline model, we cannot rely on the first criteria. Instead, each experiment is selected based on how much it would increase the Occam penalty on the more complex model. We see this in Figure 15(a.iii), which shows that the $\log(\text{OF})$ of model A decreases rapidly at first, and gradually plateaus as the experiments with high discriminatory information are used up. As a result, we see from Figure 15(a.iv) that, when selecting the best experiments, the ML ratio increases rapidly at first and then plateaus. By comparison, selecting the worst experiments leads to a more gradual increase in model discrimination. For the most part, random experiment designs fall between the best and worst designs. In the region of 30–60 experiments, however, some random experiment designs increase the discrimination more than our optimal design methodology. We note, however, that this is a particularly challenging case because the models are very similar, and the majority of the discriminatory evidence is provided by initial set of experiments that are identified for parameter inference.

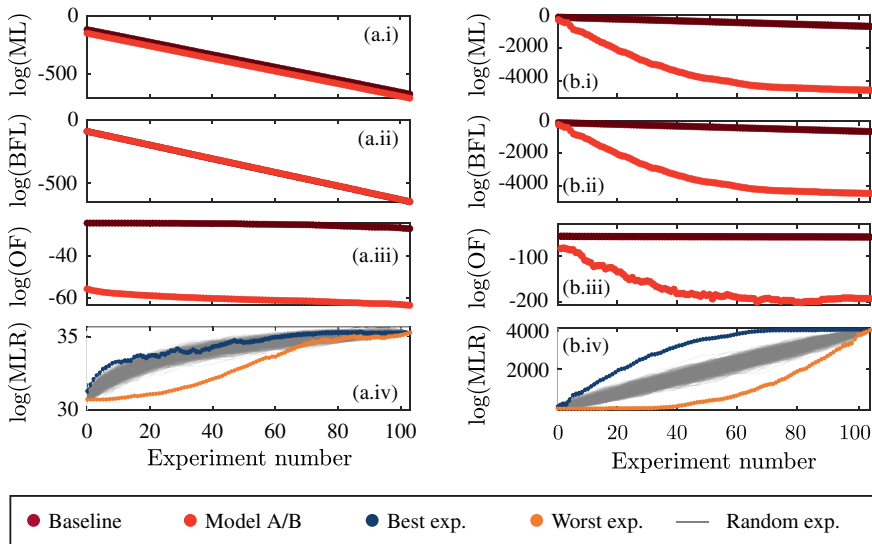


Figure 15. Four model comparison metrics are plotted against the number of experiments assimilated. The metrics are (i) the log marginal likelihood, $\log(\text{ML})$, (ii) the log best fit likelihood, $\log(\text{BFL})$, (iii) the log Occam factor, $\log(\text{OF})$, and (iv) the log of the ratio of marginal likelihoods between two models, $\log(\text{MLR})$. In panel (a), we compare the baseline model (dark red) to model A (light red), and in panel (b), we compare the baseline model (dark red) to model B (light red). In (i–iii), we only show the results produced by selecting the best experiment at each step. In (iv), we show the results produced by recursively selecting (blue) the best experiments, (orange) the worst experiments, and (gray) random experiments. A positive $\log(\text{MLR})$ means that the baseline model is preferred.

Comparing the BFLs between model B and the baseline model, shown in Figure 15(b.ii), we see that model B is increasingly penalized as more data are added. This is because the polynomial model cannot describe the data as well as the physics-based baseline model. Similarly, Figure 15(b.iii) shows that the OF penalty of model B becomes increasingly negative as more data are added, because model B is the more complex model. In this case, the algorithm must select between penalizing the model based on its fit to the data, or penalizing the model based on its complexity. Figure 15(b.ii), (b.iii) shows that both of these penalties increase rapidly at first and then plateau, indicating that the algorithm is able to select the experiments that maximize the discrimination from both penalties. As a result, we see from Figure 15 (b.iv) that, when selecting the best experiments, the ML ratio increases rapidly at first and then plateaus. By comparison, selecting the worst experiments leads to a more gradual increase in model discrimination, with most of the discriminatory information added by the later experiments. In this case, all 1,000 random experiment designs fall between the best and worst designs. Similarly to Figure 9, we see that the random designs are not tightly bounded by the best and worst designs, indicating that it is unlikely that very good (or bad) experiments will be selected at random.

6. Note on application

In the current study, we have demonstrated experiment designs that have strictly adhered to the optimal design approach. We note, however, that this approach tends to select experiment designs that are concentrated at the extremes of the parameter space (see Figures 8 and 14). Our experience has shown that training models on a narrow portion of the parameter space can lead to systematic errors going undiscovered. As a result, a strict implementation of this approach may not be appropriate for models that have not been tested on densely sampled datasets to eliminate systematic or structural error in the data or

model. One potential approach to resolve this issue is to add an extra term to the utility function that encourages the selection of experiments in regions of the parameter space that have not yet been explored. With this modification, the algorithm will initially select experiments with high information content, and then gradually switch to exploring the parameter space. This is conceptually similar to optimal sampling strategies that are commonly used in Bayesian optimization with Gaussian processes (Sengupta and Juniper, 2022).

The current study has also demonstrated optimal experiment design for parameter inference by selecting the single best experiment at each iteration. In the greedy selection process, we iteratively (i) use equation (16) to identify the most informative experiment, (ii) perform that experiment to collect the data, \mathbf{z}_{i+1} , (iii) minimize equation (4) to find \mathbf{a}_{i+1} , and (iv) solve equation (5) to calculate $(\mathbf{C}_a)_{i+1}$. In practical situations, it may be better to identify and conduct a batch of promising experiments before assimilating the data and identifying the next batch. This is also possible using the approach described in this paper if we iteratively (i) use equation (16) to identify the most informative experiment, (ii) use equation (15) to approximate $(\mathbf{C}_a)_{i+1}$, (iii) repeat steps (i) and (ii) until a batch of n approximately optimal experiments have been identified, (iv) perform the n experiments to collect the data $\mathbf{D} = \{\mathbf{z}_{i+1}, \dots, \mathbf{z}_{i+n}\}$, (v) minimize equation (4) to find \mathbf{a}_{i+n} , and (vi) solve equation (5) to calculate $(\mathbf{C}_a)_{i+n}$. In the early stages of parameter inference, not all experiments in the batches will be optimal because (i) the first-order approximation errors will accumulate, and (ii) all experiments in the batch are selected based on the current knowledge of the parameters, which may be inaccurate. However, as more data are collected, the second-order term will vanish, the parameter values will converge to the posterior, and the batches will become closer to optimal.

Finally, this paper has not addressed the task of selecting a stopping criterion, which is a case-dependent problem that cannot be prescribed in general. In some cases, it may be sensible to continue collecting data or placing sensors until a budget is exhausted, while in others, it may be sensible to continue until the model uncertainty or expected information gain falls below a threshold. One could also monitor the convergence of the parameters, and continue collecting data until the change in the parameter values drops below a threshold.

7. Conclusion

In this paper, we have developed a computationally efficient framework for Bayesian optimal experiment design. We have considered three specific questions an experimentalist may face. These are (i) which experiments would provide the maximum information about the unknown parameters of a model, (ii) where should the sensors be placed to provide maximum information about the unknown parameters, and (iii) which experiments would maximize the discrimination between candidate models?

The framework relies on approximate Bayesian inference, accelerated by adjoint methods. This greatly reduces the computational cost when compared with sampling methods such as Markov Chain Monte Carlo, but assumes that the posterior probability distributions are approximately Gaussian. This assumption mainly impacts the estimation of parameter uncertainty, and can be verified a posteriori.

We have demonstrated this framework on thermoacoustic oscillations, an industrially relevant problem which affects the development of aerospace engines. The testing of these engines is expensive, so there is a large financial incentive to minimize the number of experiments conducted. We have shown that, in a lab rig which displays thermoacoustic oscillations, we can significantly reduce the number of experiments required to perform Bayesian inference by (i) performing fewer experiments and (ii) using fewer sensors.

While other studies have found that Bayesian experiment design is too computationally expensive, we have shown that this can be overcome within the adjoint-accelerated Bayesian inference framework, as long as the posterior probability distributions are approximately Gaussian. We have also shown that the framework can be applied to a real-world problem with real experimental data, whereas most other studies have demonstrated their methodologies using synthetic data which is free of systematic error.

The framework is general and can be applied to a wide range of problems, provided there is (i) a source of data, (ii) a physics-based model, and (iii) an adjoint of the model. Adjoint or automatically

differentiated codes are readily available in the field of fluid mechanics because they have been widely used for shape optimization. This work is a further example of the utility of adjoint codes or automatically differentiated codes.

In the future, we aim to apply this framework to more complex problems in thermoacoustics, such as industrial gas turbine test rigs. There is scope to apply this methodology to a wide range of other problems in fluid mechanics, structural mechanics, control, and other fields.

Acknowledgments. M.Y. acknowledges funding for his PhD from The Cambridge Trust, The Skye Foundation, and The Oppenheimer Memorial Trust.

Author contribution. Conceptualization: M.Y.; Data curation: M.Y.; Formal analysis: M.Y.; Investigation: M.Y.; Methodology: M.Y.; Software: M.Y., M.P.J.; Supervision: M.P.J.; Visualization: M.Y.; Writing—original draft: M.Y.; Writing—review and editing: M.Y., M.P.J. Both authors approved the final submitted draft.

Data availability statement. The experimental data used in this study are available at <https://doi.org/10.17863/CAM.108251>.

Competing interest. The authors declare none.

References

- Bidar O, Anderson S, Qin N and Bank W** (2024) Sensor placement for data assimilation of turbulence models using eigenspace perturbations. *Physics of Fluids* 36, 27–29.
- Brunton SL, Noack BR and Koumoutsakos P** (2020) Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics* 52, 477–508.
- Buchta DA, Laurence SJ and Zaki TA** (2022) Assimilation of wall-pressure measurements in high-speed flow over a cone. *Journal of Fluid Mechanics* 947, 1–13.
- Dowling AP and Stow SR** (2003) Acoustic analysis of gas turbine combustors. *Journal of Propulsion and Power* 19(5), 751–764.
- Ercan T and Papadimitriou C** (2023) Bayesian optimal sensor placement for parameter estimation under modeling and input uncertainties. *Journal of Sound and Vibration* 563, 117844.
- Fedorov VV** (1972) *Theory of Optimal Experiments*. New York: Academic Press.
- Giles MB and Pierce NA** (2000) An introduction to the adjoint approach to design. *Flow, Turbulence and Combustion* 65(3–4), 393–415.
- Huan X and Marzouk YM** (2013) Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics* 232(1), 288–317.
- Huang D, Allen TT, Notz WI and Zeng N** (2006) Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of Global Optimization* 34(3), 441–466.
- Isaac T, Petra N, Stadler G and Ghattas O** (2015) Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet. *Journal of Computational Physics* 296, 348–368.
- Jeffreys H** (1973) *Scientific Inference*, 3rd Edn. Cambridge: Cambridge University Press.
- Juniper MP** (2018) Sensitivity analysis of thermoacoustic instability with adjoint Helmholtz solvers. *Physical Review Fluids* 3(11), 110509.
- Juniper MP and Sujith R** (2018) Sensitivity and nonlinearity of thermoacoustic oscillations. *Annual Review of Fluid Mechanics* 50, 661–689.
- Juniper MP and Yoko M** (2022) Generating a physics-based quantitatively-accurate model of an electrically-heated Rijke tube with Bayesian inference. *Journal of Sound and Vibration* 535, 117096.
- Kontogiannis A, Elgersma SV, Sederman AJ and Juniper MP** (2022) Joint reconstruction and segmentation of noisy velocity images as an inverse Navier–Stokes problem. *Journal of Fluid Mechanics* 944, 1–36.
- Kullback S and Leibler R** (1951) On information and sufficiency. *Annals of Mathematical Statistics* 22(1), 79–86.
- Lindley DV** (1956) On a measure of the information provided by an experiment. *Annals of Mathematical Statistics* 27(4), 986–1005.
- Loredo TJ** (2004) Bayesian adaptive exploration. *AIP Conference Proceedings* 707, 330–346.
- Luchini P and Bottaro A** (2014) Adjoint equations in stability analysis. *Annual Review of Fluid Mechanics* 46, 493–517.
- MacKay DJC** (1992) Information-based objective functions for active data selection. *Neural Computation* 4(4), 590–604.
- MacKay DJC** (2003) *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.
- Mongia HC, Held TJ, Hsiao GC and Pandalai RP** (2003) Challenges and progress in controlling dynamics in gas turbine combustors. *Journal of Propulsion and Power* 19(5), 822–829.
- Mons V and Zaki TA** (2021) Ensemble-variational assimilation of statistical data in large-eddy simulation. *Physical Review Fluids* 6(10), 1–35.

- Oefelein JC and Yang V** (1993) Comprehensive review of liquid-propellant combustion instabilities in F-1 engines. *Journal of Propulsion and Power* 9(5), 657–677.
- Papadimitriou DI and Papadimitriou C** (2015) Optimal sensor placement for the estimation of turbulence model parameters in CFD. *International Journal for Uncertainty Quantification* 5(6), 545–568.
- Ryan KJ** (2003) 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.
- Sengupta U and Juniper MP** (2022) Thermoacoustic stabilization of combustors with gradient-augmented Bayesian optimization and adjoint models. *International Journal of Spray and Combustion Dynamics* 14(3–4), 266–272.
- Sgarro A** (1981) Informational divergence and the dissimilarity of probability distributions. *Calcolo* 18, 293–302.
- Verma S, Papadimitriou C, Lüthen N, Arampatzis G and Koumoutsakos P** (2019) Optimal sensor placement for artificial swimmers. *Journal of Fluid Mechanics* 884, A24.
- Yoko M and Juniper MP** (2024) Adjoint-accelerated Bayesian inference applied to the thermoacoustic behaviour of a ducted conical flame. *Journal of Fluid Mechanics* 985, A38.