



Bayesian optimal design accelerates discovery of soft material properties from bubble dynamics

Tianyi Chu¹ · Jonathan B. Estrada² · Spencer H. Bryngelson^{1,3,4}

Received: 9 September 2024 / Accepted: 27 January 2025

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2025

Abstract

An optimal sequential experimental design approach is developed to computationally characterize soft material properties at the high strain rates associated with bubble cavitation. The approach involves optimal design and model inference. The optimal design strategy maximizes the expected information gain in a Bayesian statistical setting to design experiments that provide the most informative cavitation data about unknown soft material properties. We infer constitutive models by characterizing the associated viscoelastic properties from measurements via a hybrid ensemble-based 4D-Var method (En4D-Var). The inertial microcavitation-based high strain-rate rheometry (IMR) method (Estrada et al. *J Mech Phys Solids* 112:291–317, 2018) simulates the bubble dynamics under laser-induced cavitation. We use experimental measurements to create synthetic data representing the viscoelastic behavior of stiff and soft polyacrylamide hydrogels under realistic uncertainties. The synthetic data are seeded with larger errors than state-of-the-art measurements yet matches known material properties, reaching 1% relative error within 10 sequential designs (experiments). We discern between two seemingly equally plausible constitutive models, Neo-Hookean Kelvin–Voigt and quadratic Kelvin–Voigt, with a probability of correctness larger than 99% in the same number of experiments. This strategy discovers soft material properties, including discriminating between constitutive models and discerning their parameters, using only a few experiments.

Keywords Viscoelastic material · Bayesian optimal experimental design · Data assimilation · High strain rate · Measurement

1 Introduction

Large and rapid deformations in compliant soft materials, such as those caused by shock waves or lasers, can lead to mechanical failure. Cavitation may occur when these materials are exposed to tensile waves, leading to high strain rates ($10^3 - 10^8$ 1/s). Energy-focused cavitation, when used appropriately, can benefit biologic, medical, and surgical applications, including tissue phantom studies, laser surgery, and DNA manipulation in target cells [1–5]. However, accu-

rate characterization of realistic soft materials and biotissues under such high strain rates and large deformations is challenging due to their common high compliance [6, 7], and mechanical behavior beyond the linear elastic regime [8, 9]. Therefore, to predict mechanical behavior at high strain rates, the constitutive response of the underlying tissue must be faithfully represented.

Inertial microcavitation-based high strain-rate rheometry (IMR) has been proposed by [10] for characterizing compliant materials at finite deformations and fast speeds. This high-strain rate rheometer combines laser-induced cavitation with physical bubble dynamics models to estimate the viscoelastic properties of hydrogels through observations of the bubble radius time history. The IMR method has been applied to characterize the mechanical behavior of commonly used biomimetic hydrogels, including polyacrylamide (PA) [10–12], agarose [13, 14], and gelatin [15]. The time efficiency of the cavitation experiments, however, is limited by factors such as the chemical, degassing, and swelling protocols necessary to create pristine samples for characterization [10, 16]. Therefore, an experimental design strategy

✉ Tianyi Chu
tchu72@gatech.edu

¹ School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

² Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48105, USA

³ Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

⁴ George W. Woodruff School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

is necessary to efficiently probe material responses to different physical mechanisms, such as deformation, pressure, and thermal effects, while preserving experimental or computational resources. This design approach is intended to be robust for characterizing soft materials under different sources of uncertainty, including variations in experimental configuration and observational noise. We use the computationally efficient IMR method to develop a simulation-based optimal experimental design (OED) approach for material parameter characterizations and the physical models and theory that underpin them.

The IMR-based OED seeks to optimize the design of cavitation experiments to yield the most informative data about the viscoelastic properties of the unknown material. Following the decision-theoretic approach by [17], the relative entropy, or Kullback–Leibler (KL) divergence, from the posterior to the prior within the Bayesian statistical setting is often used to measure the information provided by an experiment. Therefore, the design process focuses on optimizing the expectation of this utility function, also known as the expected information gain (EIG). However, the direct calculation of the EIG is hindered by the intractability of the inherent double-loop integral, which is due to the absence of closed forms and the inability of conventional Monte Carlo (MC) methods. Nonlinear models complicate the analytical integration of likelihood functions or posterior distributions, necessitating computational methods. Different approaches have been proposed to numerically evaluate the EIG, including nested Laplace approximations [18–21] and nested Monte Carlo (NMC) estimators [22–26]. The Laplace approach systematically introduces bias, though NMC provides accurate estimators using a finite number of Monte Carlo samples.

Variational methods have also been incorporated into the EIG estimators to improve the convergence rate and accuracy [27, 28]. Readers are referred to [29] and [30] for reviews on this topic. With appropriate EIG estimators, the remaining task of Bayesian OED (BOED) is to optimize the EIG within the domain of design variables. Multiple optimization methods have been considered, such as simulated annealing [31], interacting particle systems [32], stochastic optimization [24, 33–35], and Bayesian optimization (BO) [27, 36, 37]. The absence of an EIG gradient, the high computational cost of individual evaluations, and the prevalence of local optima in the search space often constrain the optimization of the EIG for complex nonlinear systems. Given these difficulties, BO is particularly effective as an optimizer due to its data efficiency, ability to balance exploration and exploitation, robustness to multi-modality, and capacity to handle noisy observations, enabling efficient global optimization of the design space with a manageable number of evaluations [38, 39]. We refer the reader to [40, 41] for a comprehensive review and practical implementation of BO.

Instead of using the same design throughout the experimental process, sequential or adaptive designs have gained popularity in Bayesian design literature due to their flexibility and efficiency [25, 42–44]. Unlike static designs, which use fixed experimental configurations, sequential designs aim to maximize the expected utility at each stage of experimentation based on the outcomes of previous experiments and possible predictions of future ones. For cavitation rheometry, inferring material parameters from bubble dynamics data is important for proceeding with the sequential design.

We use data assimilation (DA) techniques for bubble-dynamics-based rheometry to improve predictions in uncertainty-prone high-strain-rate regimes. We combine the IMR method with observational data such as bubble-radius trajectories. The information needed to describe complex systems comes from different sources and has different characteristics, such as modeling assumptions and measurement noise. Each source is unlikely to fully observe the system, leading to information discrepancies between the theoretical model and the data. DA rectifies this problem by addressing uncertainty in the model and the data. In particular, the ensemble Kalman filter (EnKF) is an often-used DA tool due to its simple conceptual formulation and relative ease of implementation [45]. It achieves relatively high accuracy for a small ensemble, approximating the state as a multivariate Gaussian. Applications of EnKF include oceanography [46, 47], atmospheric science [48–50], and engineering [51]. Other variants of EnKF, such as ensemble Kalman smoother (EnKS) [52], iterative EnKS [53, 54] and ensemble-based four-dimensional variational method (En4D-Var) [55, 56] have been explored. We refer the reader to [57] for a review of common DA methods.

[58] incorporated ensemble-based DA methods with the IMR solver to provide a scalable bubble-collapse rheometry framework. It reduces the number of simulations required for accurate characterizations from a large volume in the brute-force curve fitting strategy in [10] to 48 ensembles, offering computational advantages. The hybrid En4D-Var method also requires fewer measurements per data set to characterize the mechanical properties of hydrogels [13, 58]. DA methods require appropriate theoretical models as a prior, yet they do not provide information about how to select them. The most direct approach is to choose the model that minimizes the least-squares error. However, this approach does not consider the uncertainty from different sources, such as experimental setups or measurement errors. To address this, we use the Bayesian model selection framework [59, 60] to systematically determine the model probability in the presence of these uncertainties. We consider a library of potential constitutive models and calculate the likelihood of each model using the associated posterior distributions. The model parameters and probabilities establish the sample space and determine the optimal EIG for optimal design. This approach systemati-

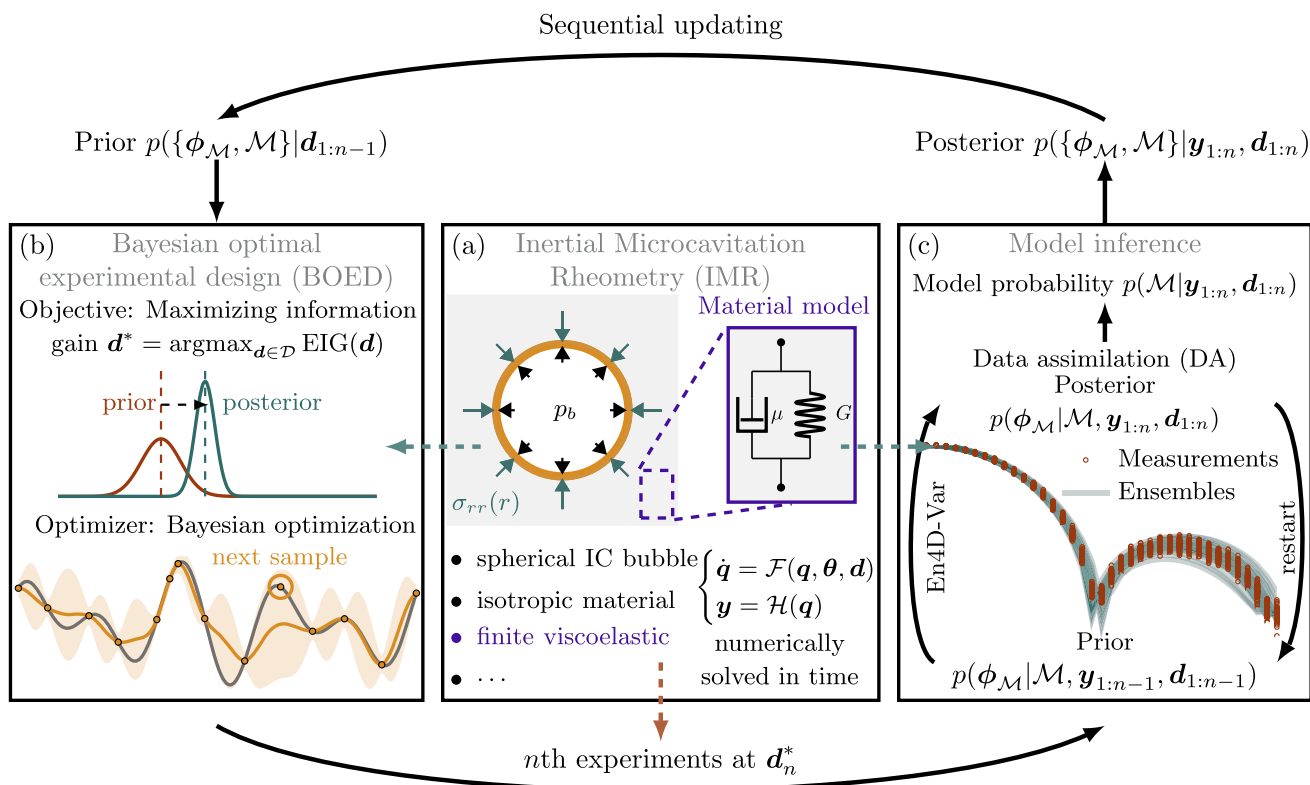


Fig. 1 Schematic of the IMR-based sequential BOED. Given a modeling parameter, $\theta = \{\mathcal{M}, \phi_{\mathcal{M}}\}$, which includes a constitutive model and its material properties, and a design \mathbf{d} that describes the experimental setup (for example, the equilibrium radius), the IMR approach

numerically solved the spherically symmetric motion of bubble dynamics. In computation, the complete flow states \mathbf{q} include bubble radius, bubble-wall velocity, temperature, and other variables, but they are only partially observable and are denoted as \mathbf{y}

cally and sequentially infers the model from measurements via different experimental designs.

Figure 1 shows a graphical overview of the sequential BOED procedure we use herein. In Sect. 2, we introduce the IMR method and use it to conduct efficient bubble cavitation simulations that provide the full flow states, \mathbf{q} . The optimal design process in Sect. 3 maximizes the EIG using BO to design the most informative cavitation experiments, denoted by \mathbf{d}^* . The model inference part in Sect. 4 characterizes the unknown material properties, $\phi_{\mathcal{M}}$, of each constitutive model, \mathcal{M} , by analyzing the bubble dynamics trajectories, \mathbf{y} , using En4D-Var. Taken together, they form the modeling parameter, $\theta = \{\mathcal{M}, \phi_{\mathcal{M}}\}$, that describes the constitutive behavior of the soft material. Subsequently, the marginal likelihood is used to calibrate the model probability. When the prior is updated using the posterior, one iteration of the sequential design is completed. Soft material properties are shown to be accurately and efficiently characterized by iterating optimal design and model inference processes. The performance of the sequential approach is demonstrated in Sect. 5 using two synthetic data sets for stiff and soft PA hydrogels. Sections 6 and 7 summarizes the main contributions and limitations.

2 Methods

2.1 Theoretical bubble dynamics model

Different spherical bubble dynamics models have been explored in the pursuit of characterizing the viscoelastic properties of surrounding materials; cavitation in soft materials is one prominent example [10, 61–64]. In these models, the Keller–Miksis equation [65] is applied to govern the spherically symmetric motion of bubble dynamics in a viscoelastic material assumed to be nearly incompressible. Upon nondimensionalization using the maximum bubble radius, R_{\max} , the far-field pressure, p_{∞} , the surrounding material density ρ , and the far-field temperature T_{∞} , the dimensionless Keller–Miksis equation is

$$\begin{aligned} & \left(1 - \frac{\dot{R}^*}{c^*}\right) R^* \ddot{R}^* + \frac{3}{2} \left(1 - \frac{\dot{R}^*}{3c^*}\right) \dot{R}^{*2} \\ & = \left(1 + \frac{\dot{R}^*}{c^*} + \frac{R^*}{c^*} \frac{d}{dt}\right) \left(p_b^* - \frac{1}{We} \dot{R}^* + S^* - 1\right). \end{aligned} \quad (1)$$

The details of dimensionless parameters are summarized in Table 1. The bubble contents are assumed to consist of

two components: water vapor and gas considered to be non-condensable, characterized by gas constants R_v and R_g , on the time scales of inertial cavitation [66, 67]. This mixture is assumed to be homobaric and follow the ideal gas law and the time-dependent pressure inside the bubble, $p_b^*(t)$, is coupled to the energy equation [10, 64]. We assume that the mass and heat transfer of the gases within the bubble obeys Fick’s law and Fourier’s law. By neglecting the initial bubble growth phase, the laser-induced cavitation model begins when the bubble reaches its maximum radius and thermodynamic equilibrium, $R^*(0) = 1$.

While the Keller–Miksis equation accurately describes spherical bubble dynamics to first order in the Mach number, appropriate constitutive relations are necessary to model the dynamic behavior of the surrounding media in terms of the time-dependent stress integral, $S^*(t)$. Combinations of springs and dashpots, such as the Kelvin–Voigt and Maxwell models, are often used to account for the change in strain rate throughout the bubble expansion-collapse life cycles during an inertial microcavitation event. We aim to develop a systematic method for selecting appropriate viscoelastic constitutive models for different gel specimens. To this end, we examine a range of constitutive models for the surrounding media, as described in Table 2. Specifically, the Kelvin–Voigt model, incorporating either just a Neo-Hookean elastic term [61] or an additional second-order strain-stiffening term [11], often better represents the nonlinear viscoelastic behavior at high strain rates [14]. These two models are different orders of Taylor expansion of the more general Fung model [68]. More models are available, but they are beyond the scope of this work. The stress integral associated with the quadratic law Kelvin–Voigt (qKV) model is

$$S^* = \overbrace{-\frac{4U^*}{\text{Re}R^*} - \frac{1}{2\text{Ca}_\infty} \left[5 - \frac{4}{\lambda} - \frac{1}{\lambda^4} \right]}^{\text{Neo-Hookean Kelvin-Voigt}} + \frac{\alpha}{\text{Ca}_\infty} \left[\frac{177}{20} + \frac{1}{4\lambda^8} + \frac{2}{5\lambda^5} - \frac{3}{2\lambda^4} + \frac{2}{\lambda^2} - \frac{6}{\lambda} - 4\lambda \right], \tag{2}$$

where α represents strain stiffening when positive and strain softening when negative [69]. A lower bound of α to maintain positive strain energy is $\alpha \geq -2/(2\lambda^2 + 1/\lambda^4 - 3)$. When $\alpha = 0$, 2 reduces to the same form of the stress integral for the Neo-Hookean model, in which dynamic shear moduli are used instead of the quasistatic moduli to account for the strain stiffening effect during cavitation. For more details, readers are referred to [10]. Recently, [13] proposed a generalized variant of qKV (Gen. qKV) that extends the capability to accommodate variations in the ground-state shear modulus, G_∞ , traditionally considered constant in the qKV model. Later, we will adopt this Gen. qKV model to account for the measurement error in the quasistatic shear modulus.

We use the modeling parameter

$$\theta \equiv \{\mathcal{M}, \phi_{\mathcal{M}}\}, \tag{3}$$

to represent a candidate mathematical constitutive model and its material properties. The design parameter is

$$d \equiv \{\text{We}, R_\infty^*\}, \tag{4}$$

representing the experimental free parameters. Following [10], in the physical context of interest, we regard densities, pressures, and temperatures as constants, though this is not a restriction of the method. We use IMR to simulate forward-time bubble dynamics with known error signatures, which we represent as Gaussian noise in the model error and measurement noise.

2.2 Numerical methods

The state vector is

$$\mathbf{q}(t) = \{R^*, \dot{R}^*, p_b, S^*, T^*, C^*, 1/\text{Ca}, 1/\text{Re}, \alpha\}, \tag{5}$$

where the state parameters represent the bubble-wall radius, velocity, bubble pressure, stress integral, the discretized temperature and vapor concentration fields inside the bubble, the reciprocal-Cauchy and reciprocal-Reynolds numbers, and the strain-stiffening parameter. The discrete-time nonlinear dynamical system takes the form of

$$\mathbf{q}_{k+1} = \mathcal{F}_k(\mathbf{q}_k, \mathbf{d}), \tag{6a}$$

$$R_{k+1}^* = \mathcal{H}(\mathbf{q}_{k+1}), \tag{6b}$$

where \mathcal{F}_k is the nonlinear operator given the time steps, and \mathcal{H} is the linear observation function that maps the state \mathbf{q} to a point in measurement space. In this study, we designate the bubble radius R^* as the primary observable variable due to its direct measurability in experimental setups. For a given time interval $t \in [0, T]$ with N_t time steps, the deterministic model outputs, $\tilde{\mathbf{Q}} = [\mathbf{q}_1 \cdots \mathbf{q}_{N_t}]$, and the corresponding bubble dynamics measurements, $\tilde{\mathbf{Y}} = [R^* \cdots R_{N_t}^*]$, can be collected as

$$\tilde{\mathbf{Q}} = \mathcal{F}(\theta, \mathbf{d}) \quad \text{and} \quad \tilde{\mathbf{Y}} = \mathcal{H}(\tilde{\mathbf{Q}}), \tag{7}$$

Table 1 Dimensionless quantities used in this manuscript

Dimensional	Dimensionless quantity	Quantity name
	$U_c = \sqrt{p_\infty/\rho}$	Characteristic velocity
	$\lambda = R/R_\infty$	Material stretch ratio
t	$t^* = tU_c/R_{\max}$	Time
R	$R^* = R/R_{\max}$	Bubble-wall radius
U	$U^* = U/U_c$	Bubble-wall velocity
R_∞	$R_\infty^* = R_\infty/R_{\max}$	Equilibrium bubble-wall radius
c	$c^* = c/U_c$	Material wave speed
p_b	$p_b^* = p_b/p_\infty$	Bubble-wall pressure
$p_{v, \text{sat}}(T_\infty)$	$p_{v, \text{sat}}^* = p_{v, \text{sat}}(T_\infty)/p_\infty$	Vapor saturation pressure
C	$C^* = 1/(1 + (p_b^*/p_{v, \text{sat}}^* - 1))R_v/R_g$	Vapor concentration
T	$T^* = T/T_\infty$	Temperature
R_{\max}	$We = p_\infty R_{\max}/(2\gamma)$	Weber number
S	$S^* = S/p_\infty$	Stress integral
G	$1/Ca = G/p_\infty$	1/Cauchy number
μ	$1/Re = \mu/(\rho U_c R_{\max})$	1/Reynolds number

Table 2 Summary of constitutive models under consideration

Model \mathcal{M}	Description	Material properties $\phi_{\mathcal{M}}$
	Newtonian Fluid	1/Re
NHE [70]	Neo-Hookean Elastic	1/Ca
NHKV [61]	Neo-Hookean Kelvin–Voigt	1/Re, 1/Ca
qKV [11]	Quadratic Law Kelvin–Voigt	1/Re, α , 1/Ca $_\infty$
Gen. qKV [71]	Generalized qKV	1/Re, α , 1/Ca $_\infty$

where \mathcal{F} represents the nonlinear operator that creates the space-time states at all time instances.

Following a procedure similar to [72], we incorporate the deterministic IMR solver in (7) with the model error ϵ_m and the experimental error ϵ_e to approximate experimental measurements, such that

$$Q_m = \tilde{Q} + \epsilon_m = \mathcal{F}(\theta, d) + \epsilon_m, \quad \text{where } \epsilon_m \sim \mathcal{N}(\mathbf{0}, \Sigma_m), \tag{8a}$$

and

$$Y = Y_m + \epsilon_e = \mathcal{H}(Q_m) + \epsilon_e, \quad \text{where } \epsilon_e \sim \mathcal{N}(\mathbf{0}, \Sigma_e). \tag{8b}$$

The observation function \mathcal{H} is linear, so (8) can be written as

$$Y = \tilde{Y} + \epsilon = \mathcal{H} \circ \mathcal{F}(\theta, d) + \epsilon, \quad \text{where } \epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma), \tag{9}$$

where ϵ is the combined error from the model and the experiments, and the true bubble dynamics, \tilde{Y} , are unobtainable from the measurements. In the following, (9) is used to create synthetic measurements.

3 Simulation-based Bayesian optimal experimental design

The goal of the optimal design procedure is to find a design point, d^* , within a given design space \mathcal{D} that maximizes the expectation of a utility function, $u(d, Y, \theta)$. That is,

$$d^* = \underset{d \in \mathcal{D}}{\operatorname{argmax}} \mathbb{E}\{u(d, Y, \theta)\} = \underset{d \in \mathcal{D}}{\operatorname{argmax}} \int_{\mathcal{Y}} u(d, Y, \theta) p(\theta|d, Y) p(Y|d) d\theta dY, \tag{10}$$

where \mathcal{Y} and Θ represent the parameter spaces for the observations and model parameters. The inference of parameters θ can be obtained based on the prior distribution observations and Bayes’ rule,

$$p(\theta|d, Y) = \frac{\overbrace{p(Y|\theta, d)}^{\text{Likelihood}} \overbrace{p(\theta|d)}^{\text{Prior}}}{\underbrace{p(Y|d)}_{\text{Evidence}}}. \tag{11}$$

The probability $p(\theta)$ can be separated as

$$p(\theta) = p(\mathcal{M})p(\phi_{\mathcal{M}}|\mathcal{M}), \tag{12}$$

which contains the probability of the mathematical constitutive model \mathcal{M} and the probability of the corresponding material parameters. From (9), the likelihood function is

$$p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{d}) = \frac{1}{\sqrt{(2\pi)^{N_t}|\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\mathbf{Y} - \mathcal{H} \circ \mathcal{F}(\boldsymbol{\theta}, \mathbf{d}))^\top (\boldsymbol{\theta}, \mathbf{d}) \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathcal{H} \circ \mathcal{F}(\boldsymbol{\theta}, \mathbf{d}))\right], \quad (13)$$

and the evidence is obtained through integration as

$$p(\mathbf{Y}|\mathbf{d}) = \int_{\Theta} p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{d}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (14)$$

The maximum information gain from the prospective experiment follows from using a relative entropy utility function, which is the same as the Kullback–Leibler (KL) divergence between the posterior and prior [17], so

$$\begin{aligned} u(\mathbf{d}, \mathbf{Y}, \boldsymbol{\theta}) &= D_{\text{KL}}(\text{posterior} \parallel \text{prior}) \\ &= \int_{\Theta} p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{Y}) \log \left[\frac{p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{Y})}{p(\boldsymbol{\theta})} \right] d\boldsymbol{\theta} = u(\mathbf{d}, \mathbf{Y}). \end{aligned} \quad (15)$$

This choice of utility function is not a function of the parameters $\boldsymbol{\theta}$. The expectation of the KL divergence is then

$$E\{u(\mathbf{d}, \mathbf{Y}, \boldsymbol{\theta})\} = \int_{\mathcal{Y}} \int_{\Theta} p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{Y}) \log \left[\frac{p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{Y})}{p(\boldsymbol{\theta})} \right] d\boldsymbol{\theta} p(\mathbf{Y}|\mathbf{d}) d\mathbf{Y} \quad (16)$$

$$= \int_{\mathcal{Y}} \int_{\Theta} \log \left[\frac{p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{d})}{p(\mathbf{Y}|\mathbf{d})} \right] p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{d}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{Y}, \quad (17)$$

where the Bayes' rule in (11) is applied. This quantity is also known as the expected information gain (EIG). Further, $p(\boldsymbol{\theta}|\mathbf{d}) = p(\boldsymbol{\theta})$, as specifying \mathbf{d} does not provide further information regarding $\boldsymbol{\theta}$. In practice, the double integral in (16) cannot be computed analytically and is expensive to approximate. To address this, a double-loop Monte Carlo (DLMC) estimator, also known as the nested MC (NMC) estimator, approximates the EIG [23]. It is

$$\begin{aligned} \text{EIG}(\mathbf{d}) &\approx \mu_{\text{NMC}}(\mathbf{d}) \equiv \frac{1}{N_2} \sum_{j=1}^{N_2} \\ &\log \left[\frac{p(\mathbf{Y}^{(j)}|\boldsymbol{\theta}^{(0,j)}, \mathbf{d})}{\frac{1}{N_1} \sum_{i=1}^{N_1} p(\mathbf{Y}^{(j)}|\boldsymbol{\theta}^{(i,j)}, \mathbf{d})} \right], \end{aligned} \quad (18)$$

where $\boldsymbol{\theta}^{(i,j)} \stackrel{\text{i.i.d.}}{\sim} p(\boldsymbol{\theta})$ and $\mathbf{Y}^{(j)} \stackrel{\text{i.i.d.}}{\sim} p(\mathbf{Y}|\boldsymbol{\theta}^{(0,j)}, \mathbf{d})$. The samples $\boldsymbol{\theta}^{(0,j)}$ are used to approximate the outer loop integral, while $\boldsymbol{\theta}^{(i=1 \rightarrow N_1, j)}$ are used in the inner loop. To

obtain the dependent pair $(\boldsymbol{\theta}^{(i,j)}, \mathbf{Y}^{(i)})$, the importance sampling technique is used: we first draw $\boldsymbol{\theta}^{(i,j)}$ from the prior $p(\boldsymbol{\theta})$, and then draw $\mathbf{Y}^{(i)}$ from the conditional distribution $p(\mathbf{Y}|\boldsymbol{\theta}^{(i,j)}, \mathbf{d})$. In the computation, the samples $\boldsymbol{\theta}^{(i,j)}$ are collected using the sample reused technique [24]. This technique uses a batch of prior samples $\{\boldsymbol{\theta}^{(l)}\}_{l=1}^{N_2}$ for both the inner and outer Monte Carlo sums, reducing the computational cost from $O(N_1 N_2)$ to $O(N_2)$. In the following, we use the notation N_{EIG} to represent the sample size used for approximating the EIG.

In practical settings, experiments and data collection for inertial cavitation are carried out separately because hydrogel specimens must be prepared for different experimental setups. Thus, a sequential experimental design is important for this purpose. We assume that the experiment outcomes are conditionally independent, given the latent variables and designs,

$$p(\mathbf{Y}_{1:N_{\text{Des}}}, \boldsymbol{\theta}|\mathbf{d}_{1:N_{\text{Des}}}) = p(\boldsymbol{\theta}) \prod_{n=1}^{N_{\text{Des}}} p(\mathbf{Y}_n|\boldsymbol{\theta}, \mathbf{d}_n). \quad (19)$$

Having conducted experiments 1, 2, \dots , $N_{\text{Des}} - 1$, the design $\mathbf{d}_{N_{\text{Des}}}$ for the prospective experiment can be obtained by replacing the prior, $p(\boldsymbol{\theta})$, with $p(\boldsymbol{\theta}|\mathbf{d}_{1:N_{\text{Des}}-1}, \mathbf{Y}_{1:N_{\text{Des}}-1})$ in (11) such that

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{Y}_{1:N_{\text{Des}}}, \mathbf{d}_{1:N_{\text{Des}}}) &= \frac{p(\mathbf{Y}_{N_{\text{Des}}}|\boldsymbol{\theta}, \mathbf{d}_{N_{\text{Des}}}) p(\boldsymbol{\theta}|\mathbf{Y}_{1:N_{\text{Des}}-1}, \mathbf{d}_{1:N_{\text{Des}}-1})}{p(\mathbf{Y}_{N_{\text{Des}}}|\mathbf{d}_{N_{\text{Des}}})} \\ &= \dots = \frac{p(\boldsymbol{\theta}) \prod_{n=1}^{N_{\text{Des}}} p(\mathbf{Y}_n|\boldsymbol{\theta}, \mathbf{d}_n)}{p(\mathbf{Y}_{1:N_{\text{Des}}}|\mathbf{d}_{1:N_{\text{Des}}})}. \end{aligned} \quad (20)$$

Similar to (18), the EIG for N_{Des} is approximated in a Markovian fashion as

$$\begin{aligned} \text{EIG}(\mathbf{d}_{N_{\text{Des}}}) &\approx \frac{1}{N_{\text{EIG}}} \sum_{j=1}^{N_{\text{EIG}}} \log \\ &\left[\frac{p(\mathbf{Y}_{N_{\text{Des}}}^{(j)}|\boldsymbol{\theta}_{N_{\text{Des}}}^{(0,j)}, \mathbf{d}_{N_{\text{Des}}})}{\frac{1}{N_{\text{EIG}}} \sum_{i=1}^{N_{\text{EIG}}} p(\mathbf{Y}_{N_{\text{Des}}}^{(j)}|\boldsymbol{\theta}_{N_{\text{Des}}}^{(i,j)}, \mathbf{d}_{N_{\text{Des}}})} \right], \end{aligned} \quad (21)$$

where $\boldsymbol{\theta}_{N_{\text{Des}}}^{(i,j)} \stackrel{\text{i.i.d.}}{\sim} p(\boldsymbol{\theta}|\mathbf{Y}_{1:N_{\text{Des}}-1}, \mathbf{d}_{1:N_{\text{Des}}-1})$ and $\mathbf{Y}_{N_{\text{Des}}}^{(j)} \stackrel{\text{i.i.d.}}{\sim} p(\mathbf{Y}|\boldsymbol{\theta}_{N_{\text{Des}}}^{(0,j)}, \mathbf{d}_{N_{\text{Des}}})$. Through this procedure, we conduct an adaptive sequential experiment that iteratively optimizes the selection of the design $\mathbf{d}_{N_{\text{Des}}}$ at each step. For each such step, we solve a sequential optimization problem

$$\mathbf{d}_{N_{\text{Des}}}^* = \underset{\mathbf{d}_{N_{\text{Des}}} \in \mathcal{D}}{\text{argmax}} \text{EIG}(\mathbf{d}_{N_{\text{Des}}}), \quad (22)$$

Given an EIG estimator, different methods can be used for (22), including some specifically developed for BOED [24, 31, 32]. Here, Bayesian optimization (BO) is selected for the subsequent design optimization, given its advantageous features such as sample efficiency, robustness to multi-modality, and inherent capability to handle noisy objective evaluations [73]. Following [41], we use the Ard Matérn 5/2 kernel for Gaussian process (GP) regression and the expected improvement criterion for the acquisition function. Further details are provided in appendix A. [73] showed that the expected improvement can be unreliable with a small number of initial data points. To improve the quality of GP, they recommend selecting several initial design points in a random quasi-uniform arrangement. This strategy creates a more reasonable initial GP model [74–76]. In practice, we initialize BO by evaluating the EIG values at N_{Int} random designs. A total number of N_{BO} BO trials is used to obtain the optimal design.

Algorithm 1 Bayesian optimal experimental design (refer to Fig. 1b for graphical illustration)

Input: prior $p(\boldsymbol{\theta}|\mathbf{d}_{1:N_{\text{Des}}})$, error variance $\boldsymbol{\Sigma}$, EIG sample size N_{EIG}
Output: Next design point $\mathbf{d}_{N_{\text{Des}}+1}^*$

- 1: Evaluate EIG for the N_{Int} random points
- 2: **for** $l = N_{\text{Int}} + 1 : N_{\text{BO}}$ **do**
- 3: Perform Gaussian process regression based on the evaluated values, $\{\text{EIG}(\mathbf{d}_{N_{\text{Des}}+1}^{(l)})\}_{l'=1}^l$
- 4: Obtain next search point, $\mathbf{d}_{N_{\text{Des}}+1}^{(l+1)}$, that maximizes expected improvement
- 5: Evaluate $\text{EIG}(\mathbf{d}_{N_{\text{Des}}+1}^{(l+1)})$
- 6: **end for**
- 7: $\mathbf{d}_{N_{\text{Des}}+1}^* \leftarrow \text{argmax}_{1 \leq l \leq N_{\text{BO}}+1} \{\text{EIG}(\mathbf{d}_{N_{\text{Des}}+1}^{(l)})\}$.
- 1: **function** $\text{EIG}(\mathbf{d}; p(\boldsymbol{\theta}|\mathbf{d}_{1:N_{\text{Des}}}); N_{\text{EIG}})$
- 2: Draw $N_{\text{EIG}} + 1$ samples $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N_{\text{EIG}})})$ from prior $p(\boldsymbol{\theta}|\mathbf{d}_{1:N_{\text{Des}}})$
- 3: Draw N_{EIG} samples $(\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(N_{\text{EIG}})})$ from the likelihood $p(\mathbf{Y}|\boldsymbol{\theta}^{(0)}, \mathbf{d})$ with Gaussian error $\boldsymbol{\Sigma}$
- 4: **for** $i = 1 : N_{\text{EIG}}$ **do**
- 5: Perform the IMR simulation for the design \mathbf{d} using the parameter $\boldsymbol{\theta}^{(i)}$
- 6: Evaluate the likelihood $p(\mathbf{Y}^{(i)}|\boldsymbol{\theta}^{(i)}, \mathbf{d})$ with Gaussian error $\boldsymbol{\Sigma}$
- 7: **end for**
- 8: Calculate the EIG using (21)
- 9: **end function**

An algorithm for IMR-based BOED is outlined in Algorithm 1. This strategy is systematic and identifies the optimal design for the next experiment. The next step involves characterizing the material properties based on the measurements of bubble dynamics.

4 Model inference

4.1 Data assimilation

With data collected from experiments or simulations on inertial cavitation, the remaining task is to find the most accurate model for characterizing bubble dynamics within uncertainty-prone high-strain-rate regimes. Here, we adopt the En4D-Var approach due to its computational efficiency [13, 58]. We assume the variables follow a multivariate Gaussian distribution and use N_{En} ensembles, $\tilde{\mathbf{Q}}_0 = (\tilde{\mathbf{Q}}_0^{(1)}, \dots, \tilde{\mathbf{Q}}_0^{(N_{\text{En}})})$, to approximate this distribution based on a given observed data set, \mathbf{Y}^{D} , and a data assimilation window size. Details of the standard En4D-Var method are provided in appendix B, along with three enhancements introduced here. First, the reciprocal-Cauchy and reciprocal-Reynolds numbers, $1/\text{Ca}$ and $1/\text{Re}$, are incorporated into the state vector in (5) to guarantee Gaussian distributions of the physical quantities, G and μ . Second, the parameter α can be negative, corresponding to strain-softening, when the quasi-static shear modulus, G_{∞} , is overestimated. Third, instead of performing En4D-Var for every measurement independently and then collecting all the posterior ensembles, we consider an iterative-restart strategy to reduce the computational cost and bias from the prior. A similar restart strategy has been used in the restart-EnKF to address the dynamical systems with strong nonlinearity [77–79]. We apply En4D-Var to the data mean, and the measurement noise matrix \mathbf{P}_k at each time step is obtained from the data. After obtaining the posterior ensembles, we restart the data assimilation process by drawing fresh samples from the inflated posterior distribution. Here, the “Relaxation Prior to Spread” (RTPS) scheme addresses the sampling error in ensemble methods due to finite ensemble size [80]. The variances are updated as

$$\sigma_i = \sigma_i^{(\text{post})} + a \left(\sigma_i^{(\text{prior})} - \sigma_i^{(\text{post})} \right), \tag{23}$$

where $a = 0.7$ is an inflation parameter [58]. We repeat this process, and the final posterior distributions are obtained through N_{Runs} complete cycles. Thus, the total number of DA runs required is $N_{\text{DA}} = N_{\text{iter}}N_{\text{Runs}}$.

4.2 Model probability

After performing data assimilation for available models, the next step is to choose models that best represent the experimental measurements. The most straightforward way is to select the model with the least-squares error. This strategy, however, does not account for the uncertainty in measurements. To tackle this, we calculate the probability of each model from the En4D-Var outputs using the Bayesian model

selection framework [59, 60]. Given the measurement data \mathbf{Y}^D , the marginal likelihood of each model \mathcal{M} can be calculated as

$$p(\mathcal{M}|\mathbf{Y}^D, \mathbf{d}) = \frac{p(\mathcal{M})}{p(\mathbf{Y}^D|\mathbf{d})} \int_{\Theta} p(\mathbf{Y}^D|\mathcal{M}, \boldsymbol{\phi}_{\mathcal{M}}, \mathbf{d}) p(\boldsymbol{\phi}_{\mathcal{M}}|\mathcal{M}, \mathbf{d}) d\boldsymbol{\phi}_{\mathcal{M}}. \quad (24)$$

Similar to (18), importance sampling can be used to approximate the marginal likelihood as

$$p(\mathcal{M}|\mathbf{Y}^D, \mathbf{d}) \approx \frac{1}{N_{\text{En}}} \sum_{i=1}^{N_{\text{En}}} p(\mathbf{Y}^D|\mathcal{M}, \boldsymbol{\phi}_{\mathcal{M}}^{(i)}, \mathbf{d}), \quad (25)$$

where $\boldsymbol{\phi}_{\mathcal{M}}^{(i)} \sim p(\boldsymbol{\phi}_{\mathcal{M}}|\mathcal{M}, \mathbf{d})$. If one assumes the models can fully represent the experiments, then $\sum_{\mathcal{M}} p(\mathcal{M}) = 1$. The posterior probability of the model \mathcal{M} can then be normalized as

$$p(\mathcal{M}|\mathbf{Y}^D, \mathbf{d}) \propto p(\mathcal{M}|\mathbf{Y}^D, \mathbf{d}) / \sum_{\mathcal{M}} p(\mathcal{M}|\mathbf{Y}^D, \mathbf{d}). \quad (26)$$

The obtained posterior distribution,

$$p(\boldsymbol{\theta}|\mathbf{Y}_{1:N_{\text{Des}}}, \mathbf{d}_{1:N_{\text{Des}}}) = p(\mathcal{M}|\mathbf{Y}_{1:N_{\text{Des}}}, \mathbf{d}_{1:N_{\text{Des}}}) p(\boldsymbol{\phi}_{\mathcal{M}}|\mathcal{M}, \mathbf{Y}_{1:N_{\text{Des}}}, \mathbf{d}_{1:N_{\text{Des}}}), \quad (27)$$

is subsequently used to update the prior for Algorithm 1 to obtain the next optimal design point.

Algorithm 2 Model inference (refer to Fig. 1c for graphical illustration)

-
- Input:** target design \mathbf{d} , prior distribution $p((\boldsymbol{\phi}_{\mathcal{M}}, \mathcal{M})|\mathbf{d}_{1:N_{\text{Des}}})$
 - Output:** posterior distribution $p((\boldsymbol{\phi}_{\mathcal{M}}, \mathcal{M})|\mathbf{Y}^D, \mathbf{d})$
 - 1: Collect data \mathbf{Y}^D at the design \mathbf{d} with error $\boldsymbol{\Sigma}$
 - 2: **for** each model \mathcal{M} **do**
 - 3: **for** $l = 1 : N_r$ **do**
 - 4: Draw N_{En} samples $(\tilde{\boldsymbol{\theta}}_0^{(1)}, \dots, \tilde{\boldsymbol{\theta}}_0^{(N_{\text{En}})})$ from the prior distribution $p(\boldsymbol{\phi}_{\mathcal{M}}|\mathcal{M}, \mathbf{d}_{1:N_{\text{Des}}})$
 - 5: Generate N_{En} initial ensembles $\tilde{\mathbf{Q}}_0 = (\tilde{\mathbf{Q}}_0^{(1)}, \dots, \tilde{\mathbf{Q}}_0^{(N_{\text{En}})})$
 - 6: Perform En4D-Var with N_{iter} iterations to update the ensembles $\tilde{\mathbf{Q}}_0$
 - 7: Perform covariance inflation and update the prior distribution
 - 8: **end for**
 - 9: Calculate the marginal likelihood $p(\mathcal{M}|\mathbf{Y}^D, \mathbf{d})$ using (25)
 - 10: **end for**
 - 11: Normalize the model probability to obtain the posterior distribution $p((\boldsymbol{\phi}_{\mathcal{M}}, \mathcal{M})|\mathbf{Y}^D, \mathbf{d})$
 - 12: Update the prior $p((\boldsymbol{\phi}_{\mathcal{M}}, \mathcal{M})|\mathbf{d}_{1:N_{\text{Des}}+1})$ for Algorithm 1 to obtain the next design point
-

An algorithm for IMR-based model inference is presented in Algorithm 2. Together with Algorithm 1, these form a

complete loop for the simulation-based characterization of soft matter, as illustrated in Fig. 1.

5 Results

We demonstrate the proposed framework by using the IMR method to create two datasets. The underlying models for these datasets are chosen to mimic the viscoelastic behavior of stiff and soft PA hydrogels [11]. The details are summarized in (3). To align these simulation-based datasets with real-world experimental measurements, we introduce synthetic error to accommodate different sources of error. These include uncertainties in measurement errors and aliasing in the bubble response. The standard deviations of this synthetic noise, $\sigma = |R^* - 1|/50 + t^*/160$, are tailored to depend on time and state, qualitatively reflecting experimental measurements [10, 11, 14]. A longer duration of measurement or being closer to bubble collapse will result in a larger error, as illustrated in Fig. 4. A set of measurements containing 100 $R(t)$ curves is collected for each design. We aim to accurately characterize the underlying model with a minimum requirement of design iterations using the optimal sequential design process, as shown in Fig. 1. We consider two candidate models, NHKV and Gen. qKV, to demonstrate the proposed framework at a reasonable computational cost. The design is initialized with a probability of 50%–50% for these two models. To better represent real-world experiments, the optimization problems for the design parameters are restricted within the ranges $We \in [100, 1000]$ and $R_{\infty}^* \in [0.14, 0.3]$ [10, 11, 13, 14]. In the computation, the data assimilation window is set up to the first two peaks of the bubble collapse. For each set of measurements, En4D-Var is run $N_{\text{runs}} = 3$ times (with 2 restarts), using 5 iterations for each run and an ensemble size of $N_{\text{En}} = 48$. This choice of ensemble size follows [58]. Later, we will show that the above setup is enough to characterize the underlying model of synthetic data.

Computations are performed on PSC Bridges2 using dual AMD 64-core CPUs (SKU 7742, Rome). The default wall-clock time is 30 min, and the memory per core is 1GB. The CPU hours required vary between the two models due to differences in automatic time step requirements needed to address stiffness near bubble collapse. Generally, a single design consisting of $N_{\text{BO}} = 15$ BO trials with $N_{\text{EIG}} = 1000$ EIG samples and its subsequent DA process requires approximately 200 CPU core hours when all simulations are performed using the qKV model and an additional 200 CPU core hours when using the NHKV model. Quantitative assessment metrics include EIG, root-mean-square error (RMSE) of the $R(t)$ data and the relative error of the material properties.

Table 3 Summary of synthetic datasets. The characterization of these parameters from experimental data are demonstrated in [11]

	Material	Model \mathcal{M}	Parameters $\phi_{\mathcal{M}}$			
			G_{∞} (kPa)	G (kPa)	μ (Pas)	α
Case 1	Stiff PA	qKV	2.77	–	0.186	0.48
Case 2	Soft PA	NHKV	0.57	8.31	0.093	–

5.1 qKV for stiff PA

We first consider qKV as the underlying model to approximate the behavior of stiff PA [11]. We assume that the quasistatic shear modulus can be measured with $G_{\infty} = 2.77 \pm 0.3\text{kPa}$, which has a higher variance than the experimental measurements. The prior distributions of the material properties are set as $G = 15.09 \pm 4.35\text{kPa}$, $\mu = 0.209 \pm 0.18\text{Pas}$ for NHKV and $\mu = 0.286 \pm 0.186\text{Pas}$, $\alpha = 0.28 \pm 0.48$ for Gen. qKV. The latter has around 50% error compared to the underlying truth, which has been shown as a reasonable offset to validate the performance of DA [58]. Truncated Gaussian distributions [81] ensures $\mu > 0$ such that the material properties are physically interpretable.

We first show the results of the simulation-based BOED in Sect. 3 using the aforementioned prior distributions as an example. Figure 2a shows the BO outputs for different sample sizes used to estimate the EIG. In general, the observed EIG increases as the sample size grows. Even without additional noise, the meaningful uncertainty in the initial model selection and material properties leads to a potential for gaining information through experiments. As a result, the EIG values are comparable for the same EIG sample size. Note that the goal of BO is to explore specific regions of the design space while maximizing the objective function based on available evaluations. However, there is no guarantee that subsequent evaluations will always yield better function values. This is because, in addition to exploiting known high-value areas, BO also emphasizes reducing uncertainty in the exploration process, which may lead to evaluations that are not immediately optimal but provide valuable information for future optimization. With an initialization of 10 random trials, only a few more trials are necessary to accurately identify the optimal EIG values. These values are shown in Fig. 2b for different sample sizes, where a decreasing trend in the slope can be observed. A similar trend can be observed for the relative error in the optimal design parameters. In the current scenario, both the NHKV and qKV models have two free modeling parameters. To estimate each EIG accurately with a relative error of less than 5%, a sample size exceeding 1000 is required, which is commensurate with about 13 CPU core hours of computational time. This cost increases exponentially with the number of parameters [24]; for example, the Upper-Convected Maxwell model [82] with three free parameters could raise the cost to approximately 52 CPU core

hours, while the Oldroyd-B model [83] with four free parameters will require hundreds of CPU core hours. Since a library of constitutive models, ranging from simple to complex, may be needed to accurately represent real-world data, sampling efficiency in global optimization becomes a critical factor in selecting an optimizer, a need that BO effectively addresses. Additionally, BO provides a means to balance exploration and exploitation, offering a more accurate estimation of the entire phase space. For these reasons, we use BO for optimization, even for the relatively simple case presented in this work. Note that the EIG serves as a guiding variable for identifying the optimal design, and its actual value is not meaningful in this context. Based on these observations, we will estimate the EIG using a sample size of $N_{\text{EIG}} = 1000$ and perform $N_{\text{BO}} = 15$ trials for BO in the design process to achieve a reasonable balance between accuracy and computational efficiency.

Next, we collect measurements at the optimal design and perform data assimilation to obtain the posterior distributions. For example, Fig. 3 shows the DA outputs using the initial prior distributions for Gen. qKV. As expected, the variance of the ensembles decreases with more DA iterations. Despite an initial guess of approximately 50% error, using En4D-Var enables accurate identification of the true material properties. It can be observed that the restart strategy with covariance inflation enhances the posterior distributions with more restart runs, decreasing the RMSE of the ensemble $R(t)$ curves. Compared to the standard En4D-Var in [58], drawing fresh samples when restarting helps avoid local minima due to initial bias in ensembles.

The final step of the sequential design is to calibrate the model probabilities based on the measurements and posteriors. Figure 4 shows the RMSE and the model probabilities by comparing the posterior $R(t)$ curves to the measurements for two designs. For each design, both models show favorable bubble dynamics compared to the average of measurements, resulting in similar RMSE. Still, considering the variance present in these data, the likelihoods of these two models offer a different perspective on model selection. For the design in Fig. 4a, c, NHKV and Gen. qKV show comparable model probability. Conversely, for the other case in Fig. 4b, d, the preference for Gen. qKV over NHKV is unequivocal. These findings are also visually corroborated. Magnified regions near the second bubble collapse, where the differences between the two models are most pronounced, are

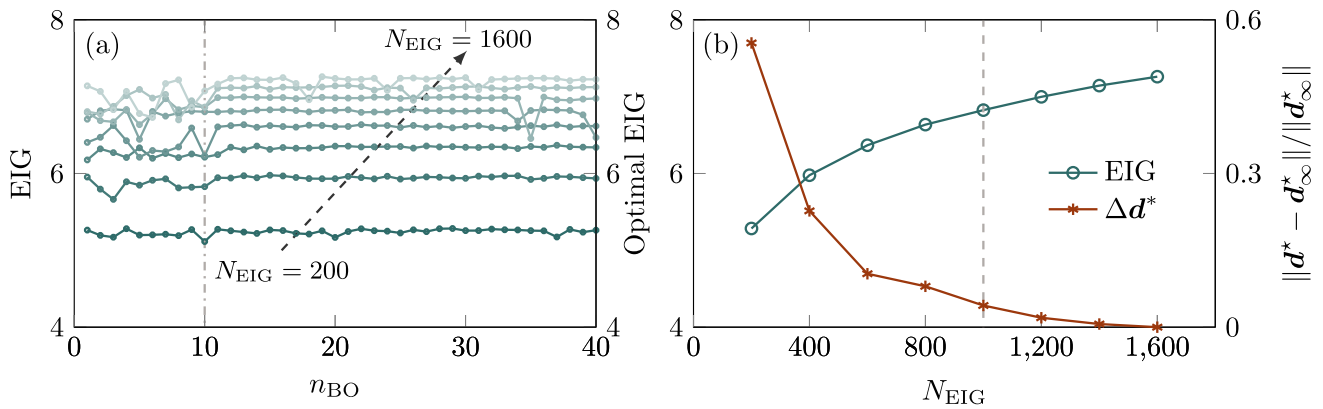
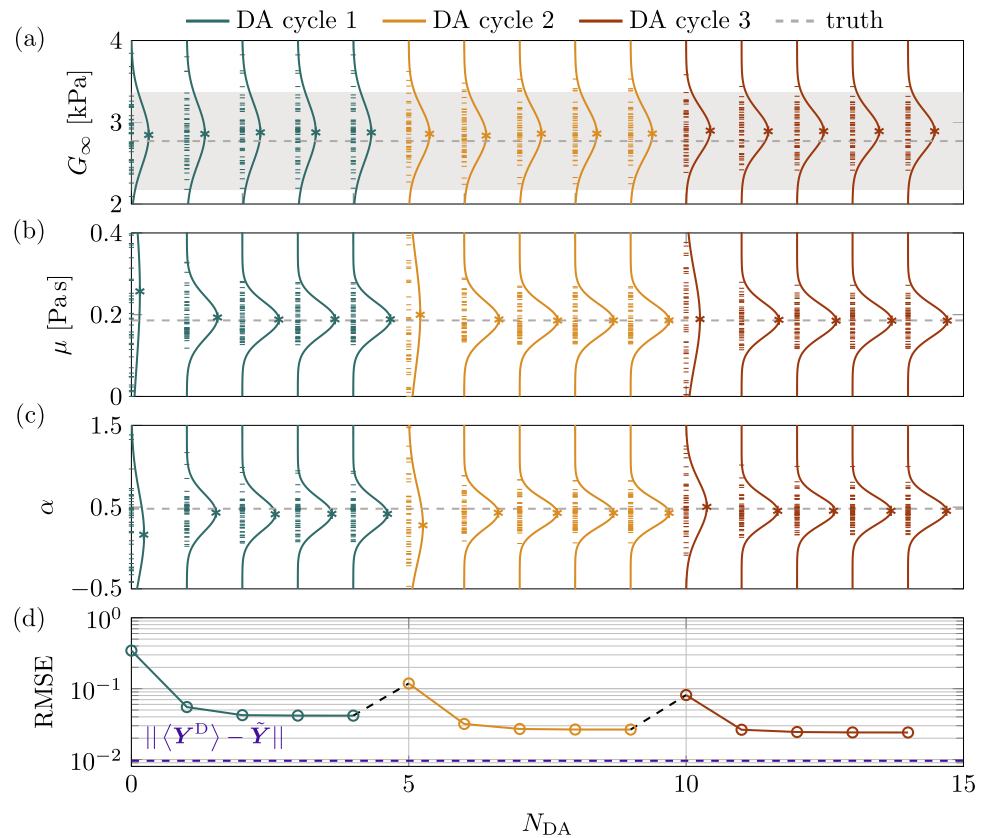


Fig. 2 BO output trajectories as a function of n_{BO} trials (a) and the optimal BO outputs (b) for EIG sample sizes $N_{EIG} = 200, 400, \dots, 1600$. The dot-dashed line in (a) indicates the onset of the BO process with 10 initial trials. The relative difference between the optimal design, d^* ,

using N_{EIG} samples and the design using 1600 samples, d_{∞}^* , is shown in (b). The optimal EIG for the EIG sample size used later, $N_{EIG} = 1000$, is highlighted in (b)

Fig. 3 DA outputs over the total En4D-var iteration number N_{DA} : ensembles for **a** G_{∞} ; **b** μ ; **c** α ; and **d** RMSE of the bubble dynamics curves (see examples in Fig. 4). The shaded area in (a) represents the 95% confidence interval for the G_{∞} measurements. The solid curves in (a–c) represent Gaussian distributions approximated from the 48 ensembles, with their respective mean values marked as stars. In (d), the error between the mean of the measurements and the unobtainable truth, $\|\langle Y^D \rangle - \tilde{Y}\|$, is shown for comparison



provided in Fig. 8. These model probabilities are next used to update the prior distribution to estimate the optimal EIG for the next design point. The processes shown in Figs. 2, 3 and 4 are then repeated.

Figure 5 shows the results for the sequential BOED. As the number of measurements increases, we observe a trend of increased exploration of parameter space and higher model probability for Gen. qKV, leading to a decreasing EIG.

Conversely, the total EIG continues to rise due to inherent measurement uncertainties. The initial EIG values for both the optimal and randomly selected design parameters, which follow a uniform distribution, are notably high, reflecting a meaningfully large discrepancy between our chosen prior distribution and the actual underlying distribution. As a result, experiments on any design yield substantial knowledge gains, leading to a larger EIG. Accurate identification of

Fig. 4 Posterior bubble dynamics trajectories and their marginal likelihoods: **a, c** $R_{\max} = 9.85 \times 10^{-4}\text{m}$ and $R_{\infty}^* = 0.2887$; **b, d** $R_{\max} = 3.87 \times 10^{-4}\text{m}$ and $R_{\infty}^* = 0.15$

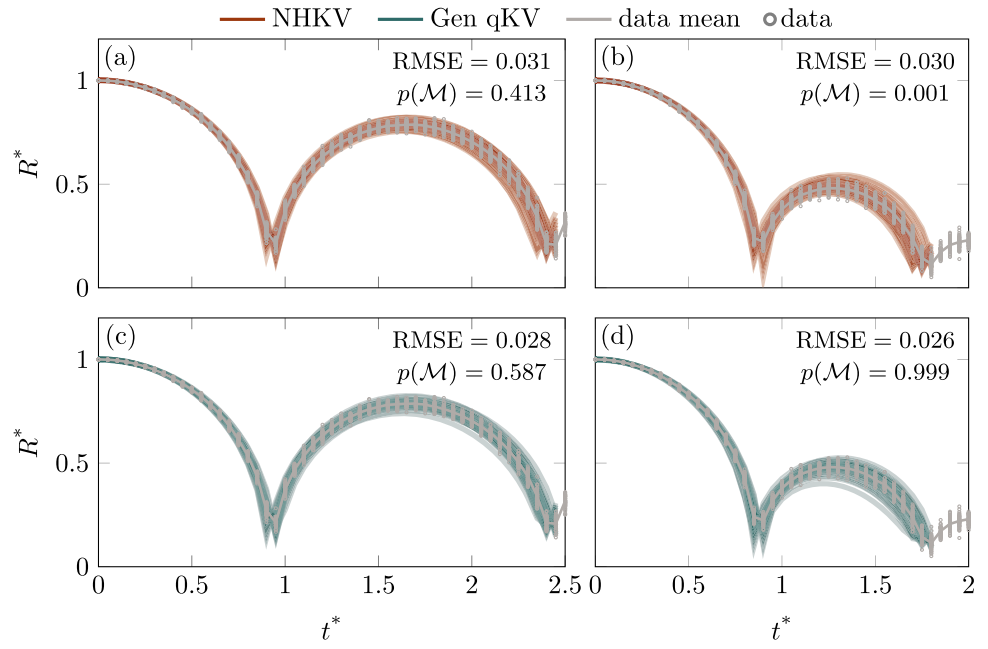
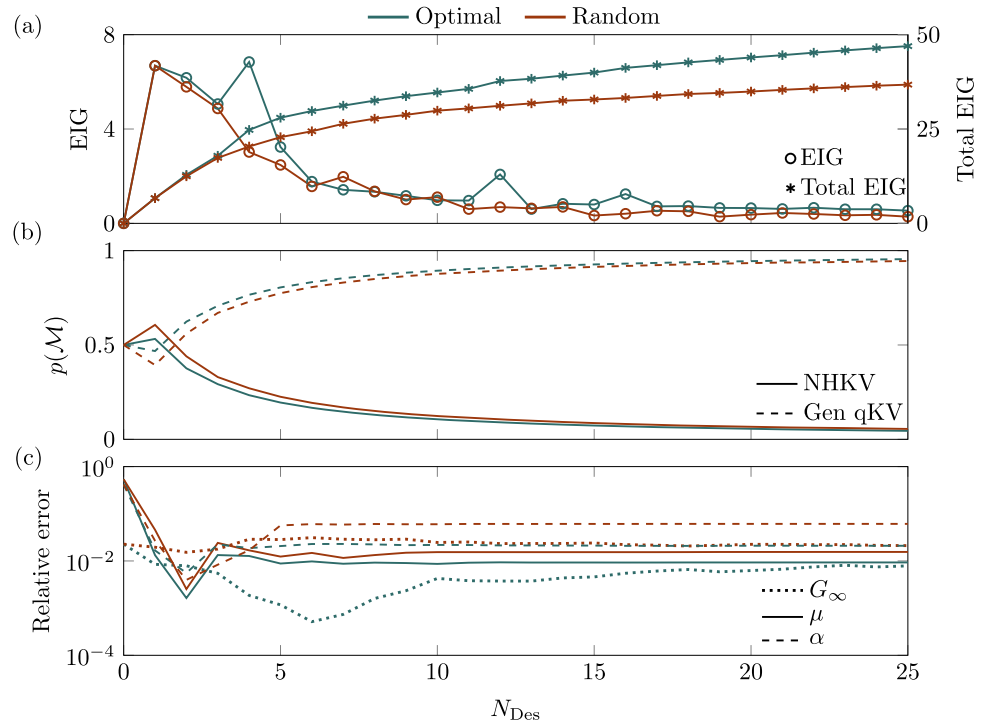


Fig. 5 Sequential BOED outputs over the design number N_{Des} : **a** EIG and total EIG; **b** model probabilities; and **c** relative error of the mean material properties



the material properties for Gen. qKV can be seen from Fig. 5c in terms of the relative error. Here, the distributions of material properties are cumulatively updated to incorporate the results from all the DA analyses up to the N_{Des} th simulation. The mean material properties converge across approximately 10 designs, coinciding with a reduction in their variances to levels deemed negligible (see the posterior entropy shown in Fig. 9b). The convergence of G_{∞} to the ground truth suggests that the Gen. qKV model effectively reduces to the

standard qKV model with a constant quasistatic shear modulus. These findings indicate that the sequential approach effectively characterizes the underlying qKV model despite multiple sources of error. The optimal sequential BOED demonstrates superior EIG, model probability, and relative error performance compared to the random design. This performance demonstrates that the proposed approach can accurately and efficiently characterize the underlying soft material from bubble dynamics.

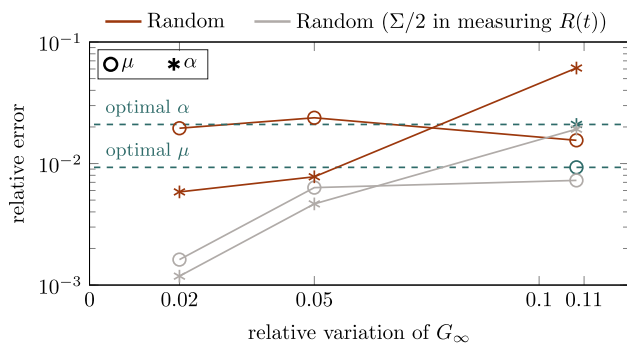


Fig. 6 Convergence study of different error sources

Finally, we examine the effects of different error sources within the system, as shown in Fig. 6. The accuracy of the stiff-straining parameter, α , improves as the variation of the quasistatic shear modulus, G_∞ , decreases to the real experimental error of 2%, see, e.g., [10]. At the same time, these parameters collectively represent a material's resistance to shearing deformation under shearing stress (see 2).

A more accurate determination of the viscosity, μ , requires smaller measurement errors in the bubble radius, R . This correspondence can also be inferred from 2 due to the coupled contributions of μ and R to the stress integral. For example, the optimal design is conducted considering high measurement noise in both R and G_∞ , and the outputs demonstrate notable improvement compared to the random design. By reducing the error in both sources, we anticipate accurately identifying the two parameters with a relative error of approximately 0.1%, as is the case for the random design.

5.2 NHKV for soft PA

Next, we consider NHKV as the underlying model to approximate the behavior of soft PA [11]. Consistent with the previous case, we initialize the prior distributions of the NHKV material properties as $G = 12.00(635)$ kPa and $\mu = 0.14 \pm 0.073$ Pas, resulting in a 50% error against the truth. For Gen. qKV, we assume that the quasistatic shear modulus is measured with $G_\infty = 0.57 \pm 0.06$ kPa and the prior distributions are set as $\mu = 0.08 \pm 0.05$ Pas, $\alpha = 0.96 \pm 0.48$.

We repeat the process shown in Sect. 5.1 and show the sequential BOED results for the synthetic soft PA in Fig. 7. The overall trend is qualitatively similar to those presented in Fig. 5. These 12 designs accurately characterize the underlying NHKV model and its material properties. Although the optimal designs are chosen by maximizing information gains instead of minimizing errors in material properties, they yield improved results for μ and comparable outcomes for G relative to the random design. Collectively, Sects. 5.1 and 5.2 illustrate that the proposed method can accurately and effi-

ciently characterize the mechanical behaviors of different soft materials.

6 Limitations of present work

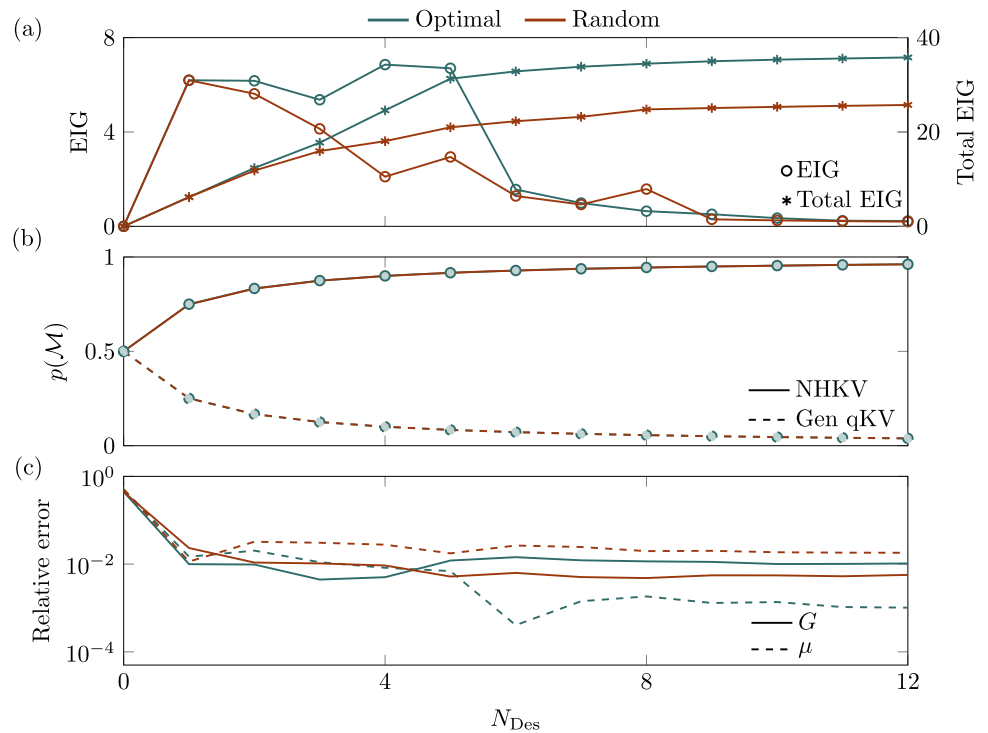
The application of En4D-Var for data assimilation is computationally efficient if the material properties, such as shear modulus and viscosities, follow a multivariate normal distribution. Consequently, its performance deteriorates when the soft materials under characterization do not adhere to this assumption. Other Bayesian parameter inference methods, such as Markov chain Monte Carlo (MCMC) sampling, can address this issue but often require many samples to compute posterior estimates with acceptable accuracy. As suggested by [84], a minimum sample size for an effective MCMC process is 10^4 , higher than the En4D-Var ensemble size used in this work, $N_{\text{En}} = 48$. Therefore, balancing the number of measurements required for posterior sampling and the constraints imposed on the distributions of the material properties becomes necessary for analyzing real experimental data. Conducting a prior assessment of the test samples could potentially aid in achieving this balance.

The proposed approach necessitates knowledge of the underlying theoretical models as a prior for optimal design and parameter inference. Specifically, in our context, this information includes the constitutive models used within the spherical bubble dynamics equations. While the modal probability calculation yields the marginal likelihoods for each constitutive model under consideration, it does not provide further insights beyond these models. If all the available models inadequately represent the experimental measurements, data-driven modeling approaches, such as system identification or operator inference methods, offer a viable strategy for exploring alternative models.

7 Conclusions

This study presents a computational approach for the optimal design of experiments to accelerate the discovery of material properties. To create synthetic data that aligns with real experiments, we used inertial microcavitation rheometry (IMR) for accurate and efficient bubble dynamics simulations. By incorporating appropriate noise to account for model error and measurement noises, these simulations serve as predictions of bubble dynamics trajectories under specific experimental conditions during the optimal design phase and as synthetic measurements during the parameter inference phase. We formulated the optimization problem within a Bayesian statistical framework to design experiments that provide the most informative data about unknown material properties. Bayesian optimization (BO) is selected as the

Fig. 7 Sequential BOED outputs over the design number N_{Des} : **a** EIG and total EIG; **b** model probabilities $p(\mathcal{M})$; and **c** relative error of the mean material properties



optimizer to maximize the expected information gain (EIG) in the design space, effectively providing sample efficiency in the global optimization of the noisy, high-computational-cost EIG, despite the lack of gradient information. The constitutive models and associated viscoelastic properties are then determined from the measurements using a hybrid ensemble-based 4D-Var method (En4D-Var). By iterating these two processes sequentially, we demonstrated accurate and efficient characterizations of two types of synthetic polyacrylamide (PA) gels. The larger error in each source of synthetic data compared to real experimental measurements evidences the robustness of the IMR-based design approach, underscoring its potential applicability to actual experimental designs.

A Bayesian optimization (BO)

The core of Bayesian Optimization (BO) is to build a surrogate model of the target function using a Gaussian Process (GP) regression and iteratively select points to evaluate based on this model. The GP can model a rich distribution over functions and depends entirely on the choice of the covariance function. After testing different kernels, we chose the Ard Matérn 5/2 kernel [41] with

$$K_{M52}(\mathbf{d}, \mathbf{d}') = \psi_0 \left(1 + \sqrt{5r^2(\mathbf{d}, \mathbf{d}') + \frac{3}{2}r^2(\mathbf{d}, \mathbf{d}')} \right)$$

$$\exp \left(-\sqrt{5r^2(\mathbf{d}, \mathbf{d}')} \right), \tag{A.1}$$

where $r^2(\mathbf{d}, \mathbf{d}') = \sum_{j=1}^2 (d_j - d'_j)^2 / \psi_j$ is the squared distance between the points in parameter space, ψ_0 is the covariance amplitude, and $\psi_{\{1,2\}}$ are the length scales associated with each design parameter. With the predictive mean function μ_y and variance function σ_y^2 , the expected improvement criterion is used for the acquisition function

$$EI(\mathbf{d}; \xi) = (\mu_y - EIG(\mathbf{d}_{best}) - \xi) \Phi \left(\frac{\mu_y - EIG(\mathbf{d}_{best}) - \xi}{\sigma_y} \right) + \sigma_y \varphi \left(\frac{\mu_y - EIG(\mathbf{d}_{best}) - \xi}{\sigma_y} \right), \tag{A.2}$$

where $\varphi(\cdot)$ is the probability density function of the normal distribution $\mathcal{N}(0, 1)$ and $\Phi(\cdot)$ is its cumulative distribution function. In practice, an exploration-exploitation parameter of $\xi = 0.01$ is used. This parameter favors high exploitation while allowing for slightly more exploration compared to the standard expected improvement ($\xi = 0$). This choice aligns with the observation that the evaluation of the EIG at a given design does not vary significantly when $N_{EIG} = 1000$ samples are used for estimation. In the computation, we use the MATLAB function `fitrgp` to train the GPR model. The input arguments include the EIG evaluations and the kernel and acquisition function described above. Based on the available evaluations, the `OptimizeHyperparameters` option minimizes the cross-validation loss for `fitrgp`

by varying normalization-related hyperparameters, such as length scales and variances, and decides whether to standardize the data. As a result, some of the presented examples are standardized throughout the process if standardization reduces the cross-validation error. All data are normalized. The ease of implementing this model makes it an attractive choice for our framework. It enables efficient design optimization while maintaining flexibility in hyperparameter tuning.

B Ensemble-based four-dimensional variational method (En4D-Var)

The En4D-Var filter can be broken down into a forecast and an analysis step. Given the initial N_{En} ensembles

$$\tilde{\mathbf{Q}}_0 = [\tilde{\mathbf{Q}}_0^{(1)} \dots \tilde{\mathbf{Q}}_0^{(N_{\text{En}})}], \quad (\text{B.1})$$

the states can be propagated in time using (6) and the corresponding ensemble bubble radii at time step k can be represented as

$$\tilde{\mathbf{Y}}_k = [R_k^{*(1)} \dots R_k^{*(N_{\text{En}})}]. \quad (\text{B.2})$$

For a given observed data set, \mathbf{Y}^{D} , and a data assimilation window size, N_t , the cost function of En4D-Var is

$$J(\mathbf{Q}_0) = \frac{1}{2N_t} \sum_{k=1}^{N_t} \left\| \mathbf{Y}_k^{\text{D}} - \mathbf{Y}_k(\mathbf{Q}_0) \right\|_{\mathbf{P}_k}^2 + \frac{1}{2} \left\| \mathbf{Q}_0 - \langle \tilde{\mathbf{Q}}_0 \rangle \right\|_{\mathbf{C}_0}^2. \quad (\text{B.3})$$

The norms for the input and output spaces are

$$\|\mathbf{Y}_k\|_{\mathbf{P}_k}^2 \equiv \mathbf{Y}_k^{\top} \mathbf{P}_k^{-1} \mathbf{Y}_k \quad \text{and} \quad \|\mathbf{Q}_0\|_{\mathbf{C}_0}^2 \equiv \mathbf{Q}_0^{\top} \mathbf{C}_0^{-1} \mathbf{Q}_0, \quad (\text{B.4})$$

where \mathbf{P}_k is the measurement noise covariance matrix at time step k , and $\mathbf{C}_0 = \tilde{\mathbf{Q}}_0' \tilde{\mathbf{Q}}_0'^{\top}$ is the initial ensemble covariance defined using the initial state perturbation matrix,

$$\tilde{\mathbf{Q}}_0' = \frac{1}{\sqrt{N_{\text{En}} - 1}} [\tilde{\mathbf{Q}}_0^{(1)} - \langle \tilde{\mathbf{Q}}_0 \rangle \dots \tilde{\mathbf{Q}}_0^{(N_{\text{En}})} - \langle \tilde{\mathbf{Q}}_0 \rangle], \quad (\text{B.5})$$

where $\langle \cdot \rangle$ is the ensemble average. The optimization for the cost function in (B.3) is carried out using the form $\mathbf{Q}_0 = \tilde{\mathbf{Q}}_0 + \tilde{\mathbf{Q}}_0' \cdot \mathbf{w}$ to restrict the solution to the subspace spanned by the scaled perturbation matrix around the initial ensembles using the correction coefficient \mathbf{w} . This process is equivalent to finding the minimizer

$$\mathbf{w}_{\text{opt}} = \underset{\mathbf{w}}{\text{argmin}} J_w(\mathbf{w}) \quad (\text{B.6})$$

for the cost function

$$J_w(\mathbf{w}) = \frac{1}{2N_t} \sum_{k=1}^{N_t} \left\| \mathbf{Y}_k^{\text{D}} - \langle \tilde{\mathbf{Y}}_k \rangle - \langle \tilde{\mathbf{Y}}_k' \cdot \mathbf{w} \rangle \right\|_{\mathbf{P}_k}^2 + \frac{1}{2} \mathbf{w}^{\top} \mathbf{w}, \quad (\text{B.7})$$

where the scaled output perturbation matrix takes the form of

$$\tilde{\mathbf{Y}}_k' = \frac{1}{\sqrt{N_{\text{En}} - 1}} [\tilde{\mathbf{Y}}_k^{(1)} - \langle \tilde{\mathbf{Y}}_k \rangle \dots \tilde{\mathbf{Y}}_k^{(N_{\text{En}})} - \langle \tilde{\mathbf{Y}}_k \rangle]. \quad (\text{B.8})$$

In practice, we follow [53] to seek the optimal correction coefficient \mathbf{w}_{opt} iteratively using a Gauss–Newton method,

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \mathbf{H}_i^{-1} \nabla J_i(\mathbf{w}_i), \quad (\text{B.9})$$

where $i < N_{\text{iter}}$ is the iteration index, and \mathbf{H} and ∇J represent approximations of the Hessian and gradient of J . They can be found with

$$\mathbf{H}_i = (N_{\text{En}} - 1) \mathbf{I} + \frac{1}{N_t} \sum_{k=1}^{N_t} \tilde{\mathbf{Y}}_k'^{\top} \mathbf{P}_k^{-1} \tilde{\mathbf{Y}}_k', \quad (\text{B.10})$$

$$\nabla J_i = -\frac{1}{N_t} \sum_{k=1}^{N_t} \tilde{\mathbf{Y}}_k'^{\top} \mathbf{P}_k^{-1} (\mathbf{Y}_k^{\text{D}} - \langle \tilde{\mathbf{Y}}_k \rangle) + (N_{\text{En}} - 1) \mathbf{w}_i. \quad (\text{B.11})$$

By combining the En4D-Var method with the subsequent marginal likelihood calculation, we establish a framework for model inference based on the data. Figure 8 shows the zoom-in regions of Fig. 4 near the second bubble collapse, where the differences between the two models are most pronounced. In Fig. 8a, the Gen. qKV model performs slightly better than the NHKV model, as reflected in their similar model probabilities. In Fig. 8b, the NHKV model fails to capture the second collapse, whereas some instances of the Gen. qKV model do. This discrepancy, reflected in their model probabilities, results in more confidence in the Gen. qKV model.

The trajectory of the design parameter used to explore the stiff PA in Sect. 5.1 over the design number N_{Des} is shown in Fig. 9a. The optimal design tends to explore regions with a larger maximum bubble radius but a smaller equilibrium radius, resulting in a larger stretch ratio. The variance of the multivariate variable, Σ , is shown in Fig. 9b in terms of the posterior entropy, defined as $\log |\Sigma|/2 + 3(1 + \log 2\pi)/2$. A reduction in the variances to levels considered negligible can be observed.

Acknowledgements The authors acknowledge support from the U.S. Department of Defense, the Army Research Office under Grant No. W911NF-23-10324 (PMs Drs. Denise Ford and Robert Martin). This work used PSC Bridges2 and NCSA Delta through allocation PHY210084 (PI Bryngelson) from the Advanced Cyberinfrastructure

Fig. 8 Magnified regions of Fig. 4 near the second bubble collapse: **a** Fig. 4a, c with $p(\text{NHKV}) = 0.413$ and $p(\text{Gen qKV}) = 0.587$; **b** Fig. 4b, d with $p(\text{NHKV}) = 0.001$ and $p(\text{Gen qKV}) = 0.999$

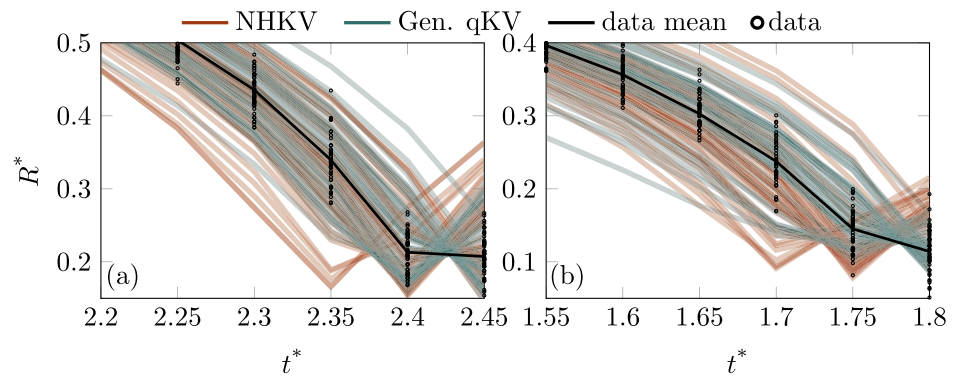
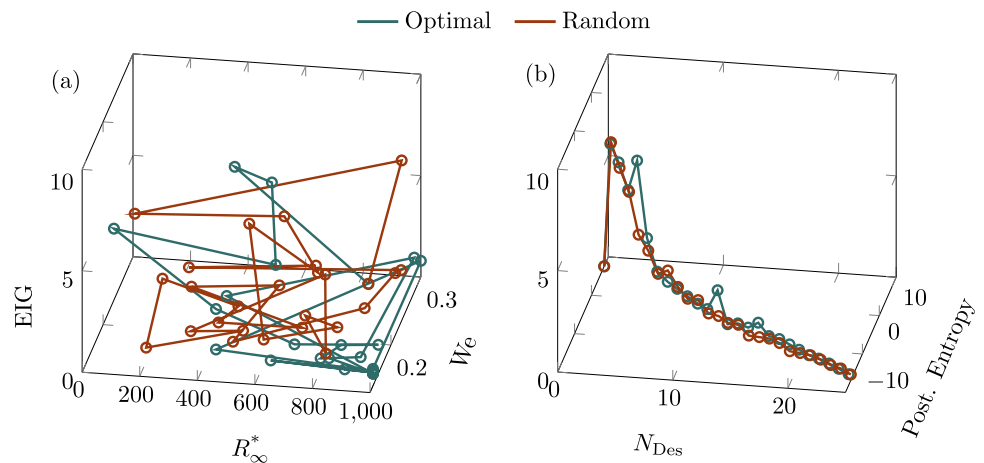


Fig. 9 Sequential BOED outputs for exploring the stiff PA in Sect. 5.1 over the design number N_{Des} : **a** trajectory of the design parameter; **b** posterior entropy



Coordination Ecosystem: Services & Support (ACCESS) program [85], which is supported by National Science Foundation Grants #2138259, #2138286, #2138307, #2137603, and #2138296.

Data availability TC: Formal analysis, Methodology, Software, Investigation, Data Curation, Validation, Visualization, Writing—original draft, Writing—review & editing. JBE: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Writing—review & editing. SHB: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing—review & editing.

Code availability Code available at https://github.com/InertialMicrocavitationRheometry/IMR_Bayesian_design

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Mancia L, Vlaisavljevich E, Yousefi N, Rodriguez M, Ziemlewicz TJ, Lee FT, Henann D, Franck C, Xu Z, Johnsen E (2019) Modeling tissue-selective cavitation damage. *Phys Med Biol* 64:225001
- Vlaisavljevich E, Maxwell A, Mancia L, Johnsen E, Cain C, Xu Z (2016) Visualizing the histotripsy process: bubble cloud-cancer cell interactions in a tissue-mimicking environment. *Ultrasound Med Biol* 42:2466–2477
- Brujan E-A, Vogel A (2006) Stress wave emission and cavitation bubble dynamics by nanosecond optical breakdown in a tissue phantom. *J Fluid Mech* 558:281–308
- Bailey MR, Khokhlova VA, Sapozhnikov OA, Kargl SG, Crum LA (2003) Physical mechanisms of the therapeutic effect of ultrasound (a review). *Acoust Phys* 49:369–388
- Brennen CE (2015) Cavitation in medicine. *Interface Focus* 5:20150022
- Arora PD, Narani N, McCulloch CAG (1999) The compliance of collagen gels regulates transforming growth factor- β induction of α -smooth muscle actin in fibroblasts. *Am J Pathol* 154:871–882
- Chen WW, Song B (2010) Split Hopkinson (Kolsky) Bar: design. Springer science & business media, testing and applications
- Lin DC, Shreiber DI, Dimitriadis EK, Horkay F (2009) Spherical indentation of soft matter beyond the Hertzian regime: numerical and experimental validation of hyperelastic models. *Biomech Model Mechanobiol* 8:345–358
- Style RW, Hyland C, Boltyanskiy R, Wettlaufer JS, Dufresne ER (2013) Surface tension and contact with soft elastic solids. *Nat Commun* 4:2728
- Estrada JB, Barajas C, Henann DL, Johnsen E, Franck C (2018) High strain-rate soft material characterization via inertial cavitation. *J Mech Phys Solids* 112:291–317
- Yang J, Cramer HC III, Franck C (2020) Extracting non-linear viscoelastic material properties from violently-collapsing cavitation bubbles. *Extreme Mech Lett* 39:100839

12. Buyukozturk S, Spratt J-S, Henann DL, Colonius T, Franck C (2022) Particle-assisted laser-induced inertial cavitation for high strain-rate soft material characterization. *Exp Mech* 62:1037–1050
13. Mancia L, Yang J, Spratt J-S, Sukovich JR, Xu Z, Colonius T, Franck C, Johnsen E (2021) Acoustic cavitation rheometry. *Soft Matter* 17:2931–2941
14. Yang J, Cramer HC III, Bremer EC, Buyukozturk S, Yin Y, Franck C (2022) Mechanical characterization of agarose hydrogels and their inherent dynamic instabilities at ballistic to ultra-high strain-rates via inertial microcavitation. *Extreme Mech Lett* 51:101572
15. Bremer-Sai EC, Yang J, McGhee A, Franck C (2024) Ballistic and blast-relevant, high-rate material properties of physically and chemically crosslinked hydrogels. *Exp Mech* 64:587–592
16. López-Fagundo C, Bar-Kochba E, Livi LL, Hoffman-Kim D, Franck C (2014) Three-dimensional traction forces of Schwann cells on compliant substrates. *J R Soc Interface* 11:20140247
17. Lindley DV (1956) On a measure of the information provided by an experiment. *Ann Math Stat* 27:986–1005
18. Lewi J, Butera R, Paninski L (2009) Sequential optimal design of neurophysiology experiments. *Neural Comput* 21:619–687
19. Cavagnaro DR, Myung JI, Pitt MA, Kujala JV (2010) Adaptive design optimization: a mutual information-based approach to model discrimination in cognitive science. *Neural Comput* 22:887–905
20. Long Q, Scavino M, Tempone R, Wang S (2013) Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Comput Methods Appl Mech Eng* 259:24–39
21. Ryan EG, Drovandi CC, Thompson MH, Pettitt AN (2014) Towards Bayesian experimental design for nonlinear models that require a large number of sampling times. *Comput Stat Data Anal* 70:45–60
22. Hamada M, Martz HF, Reese CS, Wilson AG (2001) Finding near-optimal Bayesian experimental designs via genetic algorithms. *Am Stat* 55:175–181
23. Ryan KJ (2003) Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *J Comput Graph* 12:585–603
24. Huan X, Marzouk YM (2013) Simulation-based optimal Bayesian experimental design for nonlinear systems. *J Comput Phys* 232:288–317
25. Myung JI, Cavagnaro DR, Pitt MA (2013) A tutorial on adaptive design optimization. *J Math Psychol* 57:53–67
26. Du X, Wang H (2024) Efficient estimation of expected information gain in Bayesian experimental design with multi-index Monte Carlo. *Stat Comput* 34:1–15
27. Foster A, Jankowiak M, Bingham E, Horsfall P, Teh YW, Rainforth T, Goodman N (2019) Variational Bayesian optimal experimental design. *Adv Neural Inf Process Syst* 32
28. Foster A, Jankowiak M, O’Meara M, Teh YW, Rainforth T (2020) A unified stochastic gradient approach to designing Bayesian-optimal experiments. *AISTATS* 2959–2969
29. Ryan EG, Drovandi CC, McGree JM, Pettitt AN (2016) A review of modern computational algorithms for Bayesian optimal design. *Int Stat Rev* 84:128–154
30. Rainforth T, Foster A, Ivanova DR, Bickford Smith F (2024) Modern Bayesian experimental design. *Stat Sci* 39:100–114
31. Müller P (2005) Simulation-based optimal design. *Handbook Stat* 25:509–518
32. Amzal B, Bois FY, Parent E, Robert CP (2006) Bayesian-optimal design via interacting particle systems. *J Am Stat Assoc* 101:773–785
33. Huan X, Marzouk YM (2014) Gradient-based stochastic optimization methods in Bayesian experimental design. *Int J Uncertain Quan* 4
34. Carlon AG, Dia BM, Espath L, Lopez RH, Tempone R (2020) Nesterov-aided stochastic gradient methods using Laplace approximation for Bayesian design optimization. *Comput Methods Appl Mech Eng* 363:112909
35. Karimi A, Taghizadeh L, Heitzinger C (2021) Optimal Bayesian experimental design for electrical impedance tomography in medical imaging. *Comput Methods Appl Mech Eng* 373:113489
36. Kleinegesse S, Gutmann MU (2020) Bayesian experimental design for implicit models by mutual information neural estimation. In: *ICML*, pp 5316–5326
37. Häse F, Aldeghi M, Hickman RJ, Roch LM, Aspuru-Guzik A (2021) Gryffin: an algorithm for Bayesian optimization of categorical variables informed by expert knowledge. *Appl Phys Rev* 8
38. Brochu E, Cora VM, De Freitas N (2010) A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*
39. Greenhill S, Rana S, Gupta S, Vellanki P, Venkatesh S (2020) Bayesian optimization for adaptive experimental design: a review. *IEEE Access* 8:13937–13948
40. Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N (2015) Taking the human out of the loop: a review of Bayesian optimization. *Proc IEEE* 104:148–175
41. Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. *Adv Neural Inf Process Syst* 25
42. Müller P, Berry DA, Grieve AP, Smith M, Krams M (2007) Simulation-based sequential Bayesian design. *J Stat Plan Infer* 137:3140–3150
43. Kim W, Pitt MA, Lu Z-L, Steyvers M, Myung JI (2014) A hierarchical adaptive approach to optimal experimental design. *Neural Comput* 26:2465–2492
44. Huan X, Marzouk YM (2016) Sequential Bayesian optimal experimental design via approximate dynamic programming. *arXiv preprint arXiv:1604.08320*
45. Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res Oceans* 99:10143–10162
46. Evensen G (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn* 53:343–367
47. Sakov P, Counillon F, Bertino L, Lisæter KA, Oke PR, Korabely A (2012) Topaz4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Sci* 8:633–656
48. Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble Kalman filter technique. *Mon Weather Rev* 126:796–811
49. Burgers G, Van Leeuwen PJ, Evensen G (1998) Analysis scheme in the ensemble Kalman filter. *Mon Weather Rev* 126:1719–1724
50. Whitaker JS, Hamill TM (2002) Ensemble data assimilation without perturbed observations. *Mon Weather Rev* 130:1913–1924
51. Aanonsen SI, Nøvdal G, Oliver DS, Reynolds AC, Vallès B (2009) The ensemble Kalman filter in reservoir engineering—a review. *SPE J* 14:393–412
52. Evensen G, Van Leeuwen PJ (2000) An ensemble Kalman smoother for nonlinear dynamics. *Mon Weather Rev* 128:1852–1867
53. Bocquet M, Sakov P (2013) An iterative ensemble Kalman smoother. *Q J R Meteorol Soc* 140:1521–1535
54. Sakov P, Oliver DS, Bertino L (2012) An iterative EnKF for strongly nonlinear systems. *Mon Weather Rev* 140:1988–2004
55. Liu C, Xiao Q, Wang B (2008) An ensemble-based four-dimensional variational data assimilation scheme. part i: Technical formulation and preliminary test. *Mon Weather Rev* 136:3363–3373

56. Gustafsson N, Bojarova J (2014) Four-dimensional ensemble variational (4D-En-Var) data assimilation for the high-resolution limited area model (HIRLAM). *Nonlinear Proc Geoph* 21:745–762
57. Carrassi A, Bocquet M, Bertino L, Evensen G (2018) Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *Wiley Interdiscip Rev Clim* 9:e535
58. Spratt J-S, Rodriguez M, Schmidmayer K, Bryngelson SH, Yang J, Franck C, Colonius T (2021) Characterizing viscoelastic materials via ensemble-based data assimilation of bubble collapse observations. *J Mech Phys Solids* 152:104455
59. Wasserman L (2000) Bayesian model selection and model averaging. *J Math Psychol* 44:92–107
60. Chipman H, George EI, McCulloch RE, Clyde M, Foster DP, Stine RA (2001) The practical implementation of Bayesian model selection. *Lect Notes-Monogr Ser* 65–134
61. Gaudron R, Warnez MT, Johnsen E (2015) Bubble dynamics in a viscoelastic medium with nonlinear elasticity. *J Fluid Mech* 766:54–75
62. Yang X, Church CC (2005) A model for the dynamics of gas bubbles in soft tissue. *J Acoust Soc Am* 118:3595–3606
63. Wilson CT, Hall TL, Johnsen E, Mancina L, Rodriguez M, Lundt JE, Colonius T, Henann DL, Franck C, Xu Z et al (2019) Comparative study of the dynamics of laser and acoustically generated bubbles in viscoelastic media. *Phys Rev E* 99:043103
64. Barajas C, Johnsen E (2017) The effects of heat and mass diffusion on freely oscillating bubbles in a viscoelastic, tissue-like medium. *J Acoust Soc Am* 141:908–918
65. Keller JB, Miksis M (1980) Bubble oscillations of large amplitude. *J Acoust Soc Am* 68:628–633
66. Akhatov I, Lindau O, Topolnikov A, Mettin R, Vakhitova N, Lauterborn W (2001) Collapse and rebound of a laser-induced cavitation bubble. *Phys Fluids* 13:2805–2819
67. Nigmatulin RI, Khabeev NS, Nagiev FB (1981) Dynamics, heat and mass transfer of vapour-gas bubbles in a liquid. *Int J Heat Mass Transf* 24:1033–1044
68. Fung Y-C (2013) *Biomechanics: mechanical properties of living tissues*. Springer, Berlin
69. Knowles JK (1977) The finite anti-plane shear field near the tip of a crack for a class of incompressible elastic solids. *Int J Fract* 13:611–639
70. Toyjanova J, Hannen E, Bar-Kochba E, Darling EM, Henann DL, Franck C (2014) 3d viscoelastic traction force microscopy. *Soft Matter* 10:8095–8106
71. Spratt J-S (2024) Numerical simulations of cavitating bubbles in elastic and viscoelastic materials for biomedical applications. California Institute of Technology
72. Freund JB, Ewoldt RH (2015) Quantitative rheological model selection: good fits versus credible models using Bayesian inference. *J Rheol* 59:667–701
73. Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J Global Optim* 13:455–492
74. Kandasamy K, Vysyaraju KR, Neiswanger W, Paria B, Collins CR, Schneider J, Poczos B, Xing EP (2020) Tuning hyperparameters without grad students: scalable and robust Bayesian optimisation with dragonfly. *J Mach Learn Res* 21:1–27
75. Benjamin L, Brian K, Guilherme O, Eytan B (2019) Constrained Bayesian optimization with noisy experiments. *Bayesian Anal* 14:495–519
76. Eriksson D, Poloczek M (2021) Scalable constrained Bayesian optimization. In: *AISTATS*, pp. 730–738
77. Zafari M, Reynolds AC (2005) Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter. In: *SPE J*. SPE-95750
78. Gu Y, Oliver DS (2007) An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE J* 12:438–446
79. Hendricks Franssen H-J, Kinzelbach W (2008) Real-time groundwater flow modeling with the ensemble Kalman filter: joint estimation of states and parameters and the filter inbreeding problem. *Water Resour Res* 44
80. Whitaker JS, Hamill TM (2012) Evaluating methods to account for system errors in ensemble data assimilation. *Mon Weather Rev* 140:3078–3089
81. Robert CP (1995) Simulation of truncated normal variables. *Stat Comput* 5:121–125
82. Allen JS, Roy RA (2000) Dynamics of gas bubbles in viscoelastic fluids. ii. Nonlinear viscoelasticity. *J Acoust Soc Am* 108:1640–1650
83. Oldroyd JG (1950) On the formulation of rheological equations of state. *Lond Ser A Math Phys Eng Sci* 200:523–541
84. Kruschke J (2014) *Doing Bayesian data analysis*. Academic Press, Cambridge
85. Boerner TJ, Deems S, Furlani TR, Knuth SL, Towns J (2023) Access: advancing innovation: NSF’s advanced cyberinfrastructure coordination ecosystem: services & support. In: *PEARC*, pp 173–176

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.