


Hierarchical Bayesian Uncertainty Quantification for a Model of the Red Blood Cell

Athena Economides,¹ Georgios Arampatzis,¹ Dmitry Alexeev¹,¹ Sergey Litvinov,¹ Lucas Amoudruz¹,¹ Lina Kulakova,¹ Costas Papadimitriou²,² and Petros Koumoutsakos^{1,3,*}

¹*Computational Science and Engineering Laboratory, ETH Zürich, CH-8092 Zürich, Switzerland*

²*Department of Mechanical Engineering, University of Thessaly, Volos 38334, Greece*

³*School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*

 (Received 30 October 2020; revised 7 February 2021; accepted 15 February 2021; published 22 March 2021)

Simulations of blood flows in microfluidic devices and physiological systems are gaining importance in complementing experimental and clinical studies. The predictive capabilities of these simulations hinge on the parameters of the red blood cell (RBC) model that are usually calibrated from experimental data. However, these parameter values may vary drastically when calibrated using different experimental quantities or experimental settings. In turn, the results of existing blood flow simulations largely depend on the utilized parameters that have been chosen to validate a particular experiment. We suggest a revision to this type of model calibration to properly integrate experimental data in the computational models and accordingly inform their predictions. In this context, we introduce the calibration of a popular RBC model using data-driven, hierarchical Bayesian inference. We employ data from classical experiments of RBC stretching by optical tweezers and tank treading in shear flows, and distinguish the calibration of the model parameters through single-level and hierarchical Bayesian uncertainty quantification. We find that the optimal model parameters depend not only on the data used for the inference but also on the way the data are used in the inference process. Single-level Bayesian models predict well the data used in their calibration, but are inferior to the hierarchical Bayesian model at predicting previously unseen data. This work demonstrates that the proper integration of experimental data is essential for the development of a robust and transferable RBC model. We believe that the present study can serve as a prototype across scientific fields, in revising the integration of computational models and heterogeneous experimental data.

DOI: [10.1103/PhysRevApplied.15.034062](https://doi.org/10.1103/PhysRevApplied.15.034062)

I. INTRODUCTION

Understanding the governing principles of blood flow is instrumental for the design of future medical devices, the improvement of diagnostic measures, and the development of next-generation therapies for blood-borne diseases. The rheology of blood is known to be governed by the hydrodynamic interactions between neighboring red blood cells (RBCs) and the surrounding plasma [1]. Consequently, the viscoelastic properties of single RBCs in static loads and in shear flows have been studied extensively in experiments [2,3]. In recent years, advanced microfluidic techniques have enabled the detailed study of single RBCs under various conditions [3–5]. Mathematical models can complement these experimental studies and further assist the optimization of microfluidic devices [6].

RBC models can be categorized in terms of continuum, mesoscopic, and molecular-level methods. Continuum models of blood, while able to cover domains from

the size of an artery [7] up to the full circulatory system [8], assume (often Newtonian) stress-strain relations and cannot provide microscopic rheological accuracy that may be of relevance in different parts of the flow field (microcapillaries, stenosis, vessel bifurcations, etc.). Detailed RBC models described by boundary integral methods [9] allow for subcellular resolution, although their computational cost in simulations of multiple cells often implies the usage of low-resolution models [10] or the study of two-dimensional systems [6]. In a similar manner, molecular-level methods are restricted to very small systems such as modeling a part of the RBC membrane [11] and times of the order of nanoseconds. Mesoscopic RBC models can provide submicrometer resolution and have substantially lower computational costs than molecular-level methods [1,12]. However, their accuracy depends critically on the model parameters for the RBC membrane and the lumen, as well as their interaction with the surrounding flow field.

The first mesoscopic RBC model [13] represented the spectrin cytoskeleton by two- and three-body effective potentials. The bending resistance of the membrane's lipid

*petros@seas.harvard.edu

bilayer was described by an energy term. Volume and area constraints were also incorporated in the model based on experimental evidence from healthy RBCs [14]. Following this work, in Refs. [15,16] the authors used the wormlike chain (WLC) model for the description of the nonlinear force-displacement behavior of each spectrin element. They derived expressions for the membrane's material properties, such as the shear modulus, linear area compression modulus, linear elastic Young's modulus, and the Poisson ratio. They studied large cell deformations under stretching and compared the computational results to experimental data [17]. A systematic coarse-graining procedure [18] reduced the degrees of freedom of the model by 2 orders of magnitude. The authors also introduced the coupling of this model with dissipative particle dynamics (DPD) to simulate flows with suspended RBCs. Karniadakis' group [19,20] further extended this coarse-grained model by introducing local area energy and membrane dissipation. Subsequent studies dealt with the validation of the model and simulations of RBCs under various conditions. Single cell studies include stretching [19], twisting torque cytometry (TTC) [20], dynamics of RBCs in high viscosity-contrast shear flows [20], flows in stenotic channels [21], flows in cylindrical microchannels [22], membrane thermal fluctuations at equilibrium [23], flows in microfluidic devices aiming at size-based blood filtering [24], RBC dynamics in shear flows with physiological viscosity contrasts [25], and flow-induced shape transitions [26]. Multicell studies include cell-free layer formation in microvessels [27], Pf-malaria biophysics [28], blood viscosity prediction [29], and platelet transport in microchannels with constrictions [30].

Remarkably, despite the numerous validation studies present in the literature, a consensus for the model parameters has not yet been reached. The choice of the mechanical law describing the elasticity of the spectrin cytoskeleton is at the center of this problem. Moreover, it is well known that the estimates of RBC mechanical moduli depend on the particular mechanical model [19,31,32]. In this study we demonstrate how to remedy this situation by focusing on the WLC model of the spectrin elastic energy. Different studies employing the WLC model report (often significantly) different values for the shear modulus, bending modulus, and the membrane viscosity for the RBC (Table I). Moreover, the respective model predictions are not robust, as no uncertainty bounds are reported for the parameter values.

It is important to note that discrepancies between the results of experimental studies [33] hinder validation studies while insufficient information for the discrimination between models [1,34] contribute to reduced reliability of these simulations. Hence, we believe that it is essential to assess the veracity of predictions from the RBC model, and its applicability as a design and exploration tool that can assist in clinical studies.

TABLE I. Summary of RBC mechanical properties used in the literature. Here T ($^{\circ}\text{C}$) is the temperature, x_0 is the equilibrium-to-maximum spring length in the WLC strain law, μ_0 ($\mu\text{M}/\text{m}$) is the shear modulus, κ_b (10^{-19} J) is the bending rigidity, and η_m (Pa S) is the three-dimensional lipid bilayer viscosity as given in the respective studies. If no parameter value is given, we label the corresponding entry with three center dots (\dots). If Young's modulus is given instead of the shear modulus, we estimate the shear modulus following Ref. [19] as $Y \approx 4\mu_0$ and label the entry with an asterisk (*).

Application	T	x_0^{-1}	μ_0	κ_b	η_m
Single RBC					
Stretching [19]	23	2.2	6.3	2.4	\dots
TTC and shear flow [20]	23	2.2	6.3	4.8	0.022
Cylindrical channel flow [22]	37	2.2	4.73*	3.0	\dots
Equilibrium [23]	23	2.2	2.42	1.43	0.1
Flow in microfluidic device [24]	37	2.2	4.73*	3.0	\dots
Dynamics in shear [25]	37	2.2	4.73*	3.0	\dots
Shape transitions [26]	37	2.2	4.8	3.0	\dots
Multiple RBCs					
Cell-free layer [27]	23	2.2	4.73*	2.4	0.022
Pf-malaria biophysics [28]	37	2.2	6.3	2.4	\dots
Blood viscosity [29]	37	2.2	4.82	3.0	0.0144
Platelet transport [30]	27	2.2	4.5	3.0	\dots

The diversity in the sources of uncertainty and our lack of knowledge on their effect in the final computational prediction necessitates their explicit consideration in the construction of a predictive model. We propose a data-driven, Bayesian uncertainty quantification (UQ) framework to address the challenges inherent to RBC modeling. The framework is conveniently built on hierarchical Bayesian methods [35,36] for quantifying uncertainty due to estimation accuracy and variability from various modeling and experimental effects. This framework integrates data in the modeling process and quantifies the uncertainties in model predictions [37–39]. The increasing importance of UQ in simulations related to this study is reflected in its applications, including sensitivity analysis and error estimation in molecular simulations [40–43], model comparison, and uncertainty quantification in blood flow simulations [44,45].

The ever expanding role of UQ in science and engineering is facilitated as recent hardware and algorithmic advances have helped to overcome their computational cost [38]. In addition, open-source UQ software is facilitating the incorporation of complex computational models into the UQ framework [46–48].

Here we integrate a RBC model and DPD flow simulations, following a hierarchical Bayesian UQ framework

[49] that takes into account the heterogeneity of the experimental data. The resulting *hierarchical* Bayesian model \mathcal{M}_{HB} depends on a set of parameters $\boldsymbol{\vartheta}$ with probability distributions inferred from experiments on cells under stretching and shear flow. We compare the hierarchical Bayesian model with *single-level* Bayesian models \mathcal{M}_i , whose parameters $\boldsymbol{\vartheta}_i$ are inferred based on a distinct experimental data set \mathbf{d}_i . We propagate the parameter uncertainty in the predictions of the RBC model, and compare the results of the single-level and hierarchical Bayesian models. The transferability of the RBC model is assessed in cross-experiment predictions: by testing the predictions of the RBC model on a specific quantity using parameters inferred from a different quantity. In particular, cross-experiment predictions are made for the thickness of RBCs at equilibrium, the extension of RBC diameters under stretching, and the inclination angle of tank-trading RBCs in shear flow, using Bayesian models that have not used these experimental data during the inference.

II. METHODS

The coarse-grained RBC model is described in Sec. II A 1, and the DPD solvent model in Sec. II A 2. In Sec. II B we present the Bayesian UQ framework used to infer probability distributions for the parameters of the RBC model. The inference is performed via sampling, thus requiring large numbers of model evaluations (i.e., RBC simulations). We remedy the high computational cost associated with each model evaluation by constructing surrogates for the output of the RBC model using Gaussian processes (GPs), as described in Sec. II C.

A. Mesoscopic modeling of RBCs

1. Membrane modeling

We discretize the RBC membrane on a triangulated mesh, composed of N_v vertices and N_s links (springs) [20]. The elasticity of the spectrin cytoskeleton is modeled by elastic forces between the springs (for the shear energy) and by a local area conservation constraint (for the stretch energy). The resistance to bending induced by the lipid bilayer is incorporated through an energy potential. Its magnitude depends on the angle between neighboring triangles while the membrane incompressibility is enforced by a global area constraint. The membrane viscosity is modeled by the addition of viscous dissipation on the springs. Finally, the incompressibility of the enclosed hemoglobin solution is represented by a volume constraint term. The total potential energy on the RBC membrane is therefore composed of four terms:

$$U = U_{\text{in plane}} + U_{\text{bending}} + U_{\text{area}} + U_{\text{volume}}.$$

Here $U_{\text{in plane}}$ accounts for the energy of the elastic spectrin network of the RBC membrane, modeled by an attractive

wormlike chain potential and a repulsive potential such that a nonzero equilibrium spring length can be obtained,

$$U_{\text{in plane}} = \sum_{j=1}^{N_s} \left[\frac{k_s l_m (3x_j^2 - 2x_j^3)}{4(1-x_j)} + \frac{k_{p_j}}{l_j} \right],$$

where k_s is a spring constant, l_j is the length of the j th spring, $x_j = l_j/l_m$, l_m is the maximum spring extension, and k_{p_j} is computed such that the total spring force on each spring is zero at equilibrium ($l_j = l_0$). The bending energy term, U_{bending} , accounts for the bending resistance of the lipid bilayer and is modeled as

$$U_{\text{bending}} = k_b \sum_{j=1}^{N_s} [1 - \cos \theta_j],$$

where k_b is a bending coefficient and θ_j is the angle between two adjacent triangles. The terms U_{area} and U_{volume} represent the area and volume conservation constraints, respectively, given by

$$U_{\text{area}} = \frac{K_A (A - A_0)^2}{2A_0} + \sum_{j=1}^{N_t} \frac{k_a (A_j - A_{0j})^2}{2A_{0j}},$$

$$U_{\text{volume}} = \frac{K_V (V - V_0)^2}{2V_0},$$

where A_j and A_{0j} are the current and initial areas of the j th triangle, A and A_0 are the current and initial total membrane areas, V and V_0 are the current and initial volumes enclosed by the membrane, k_a is the stretch modulus controlling the local area compressibility, and K_A and K_V are coefficients for the global area and volume, respectively.

The viscous dissipation of the membrane is modeled through the addition of a dissipative force term on the springs. We use the membrane viscosity formulation presented in Ref. [20] and set the noncentral part of the force to zero ($\gamma^T = 0$), as this term does not conserve the angular momentum. In this case, the dissipative force on a spring connecting vertices i and j is

$$\mathbf{F}_{m,ij}^D = -\gamma^C (\mathbf{v}_{ij} \cdot \mathbf{e}_{ij}) \mathbf{e}_{ij},$$

where γ^T and γ^C are coefficients to the central and noncentral parts of the membrane dissipation force, respectively, \mathbf{e}_{ij} is the unit vector along the membrane vertex centers, $\mathbf{e}_{ij} = \mathbf{r}_{ij}/\|\mathbf{r}_{ij}\|$, and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ with \mathbf{r}_i being the positional vector of vertex i . The model includes an additional fluctuation term to satisfy the fluctuation-dissipation theorem [20,50].

a. Connection to macroscopic mechanical properties. The parameters of the RBC model have been correlated to its macroscopic mechanical properties through a linear analysis of a regular hexagonal network [20]. Based on this analysis, the shear modulus μ_0 , bending rigidity κ_b , and membrane surface viscosity η_m^{2D} are estimated as

$$\begin{aligned}\mu_0 &\approx \frac{\sqrt{3}k_s}{4l_mx_0} \left(\frac{x_0}{2(1-x_0)^3} - \frac{1}{4(1-x_0)^2} + \frac{1}{4} \right) + \frac{3\sqrt{3}k_p}{4l_0^3}, \\ \kappa_b &\approx k_b\sqrt{3}/2, \\ \eta_m^{2D} &\approx \gamma^C\sqrt{3}/4,\end{aligned}$$

where $x_0 = l_{0j}/l_{mj}$ is the equilibrium-to-maximum spring length, and k_p is computed similarly to k_{pj} by using the average equilibrium spring length, $l_0 = [4A_0/(2\sqrt{3}N_v - 4)]^{1/2}$, instead of the individual spring length.

b. Stress-free shape. The set of l_{0j} spring lengths defines the equilibrium, or stress-free, shape of the RBC. The existence of the stress-free shape is attributed to experimental findings of Fischer [51], who studied the shape memory of human RBCs by tracking the final position of membrane points initially forming the rim of the RBC, after the RBC is subjected to a shear flow. The exact stress-free shape is currently a subject of ongoing investigations [52–54]. It has been postulated [52] that the stress-free shape affects the dynamics of the RBC at low shear stresses and the transition threshold from tumbling to the tank-treading regime. Here, we only consider pure tank-treading motion, at shear stresses considerably larger than the transition regime. We follow a popular choice of using a biconcave stress-free shape [24,55], which is hereafter considered as part of our computational model. The reference biconcave shape is based on the experimental findings of Evans and Fung [56] for an isotonic solution.

2. Solvent modeling

The solvent is modeled as a collection of DPD particles [57] with a given density $\rho = n_d m$, where n_d is the number density and m is the DPD particle mass. Particle interactions are governed by pairwise forces \mathbf{F}_{ij} , and their velocity \mathbf{v}_i and position \mathbf{r}_i are updated based on Newton's second law,

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i, \quad \frac{d\mathbf{v}_i}{dt} = \frac{1}{m} \sum_{j \neq i} \mathbf{F}_{ij}.$$

The force \mathbf{F}_{ij} consists of three parts: a conservative force \mathbf{F}_{ij}^C , a dissipative force \mathbf{F}_{ij}^D , and a random force \mathbf{F}_{ij}^R ,

$$\mathbf{F}_{ij} = \mathbf{F}_{ij}^C + \mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R$$

with

$$\begin{aligned}\mathbf{F}_{ij}^C &= a w^C(r_{ij}) \mathbf{e}_{ij}, \\ \mathbf{F}_{ij}^D &= \gamma w^D(r_{ij}) (\mathbf{e}_{ij} \cdot \mathbf{v}_{ij}) \mathbf{e}_{ij}, \\ \mathbf{F}_{ij}^R &= \sigma w^R(r_{ij}) \zeta_{ij} \Delta t^{-1/2} \mathbf{e}_{ij}.\end{aligned}$$

Here a , γ , and σ are coefficients of the respective DPD forces, and affect the material properties of the solvent, w^D , w^R , and w^C are weight functions that are zero for $r > r_c$, r_c is the cutoff radius, and the ζ_{ij} are independent Gaussian random variables with zero mean and unit variance, chosen independently for each pair of particles and at each timestep. The vector between particle centers i and j is denoted by $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and their relative velocity by $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$. The coefficients of the dissipative and random forces are connected through the fluctuation-dissipation theorem [58]. We choose the weighting kernels w^D , w^R , and w^C , defined as

$$\begin{aligned}w^D(r) &= [w^R(r)]^2 = [w(r)]^{2k}, \\ w^C(r) &= w(r),\end{aligned}$$

where k is the envelope parameter and

$$w(r) = \begin{cases} 1 - \frac{r}{r_c}, & r < r_c, \\ 0, & r \geq r_c. \end{cases}$$

The time integration is performed using a modified velocity-verlet algorithm with $\lambda = 1/2$ [57]. The timestep is set through Courant - Friedrichs - Lewy-type conditions for the viscous, sonic, and acceleration timescales, following Ref. [59].

3. Boundary conditions

The no-penetrability condition for the RBC membrane is enforced through the bounce back reflection mechanism for the solvent particles [60]. The RBC is coupled to the fluid flow through dissipation forces between the mesh vertices and surrounding solvent particles [20]. Solid walls are modeled using “frozen” particles, generated by first equilibrating a homogeneous DPD fluid and then freezing the particles that are inside the wall geometry [12]. The “frozen” particles have a prescribed wall velocity and are not subjected to position update. The density profile and viscous dissipation for wall-solvent interactions are maintained through DPD interactions, identical to those used for solvent-solvent interactions.

It is important to note that we model the inner and outer solvents with different particles (having different γ), allowing for different inner and outer solvent viscosities.

4. Software

The simulations are performed with `Mirheo` [61], a high-throughput open-source software with kernels thoroughly optimized for graphics processing units, aimed at microrheology simulations using state-of-the-art RBC-DPD models.

B. Bayesian uncertainty quantification

The objective of our study is to infer values for the parameters of the RBC model using a set of experimental data $\mathbf{d} = (d_1, \dots, d_N) \in \mathbb{R}^N$, where each d_j corresponds to external conditions $\mathbf{x}_j \in \mathbb{R}^{N_x}$. Since the observations may contain noise and the model is an imperfect description of the physical phenomenon, we assume that the parameters of the RBC model will also contain uncertainty. We wish to quantify this uncertainty using the rigorous setting of Bayesian inference.

In Sec. II B 1 we describe the Bayesian inference under the assumption of a single experimental data set. In Sec. II B 2 we extend this framework to account for multiple data sets, by considering more complex Bayesian models.

1. Single-level Bayesian models

We consider a set of experimental data $\mathbf{d} \in \mathbb{R}^N$ where each data point corresponds to external conditions $\mathbf{x}_j \in \mathbb{R}^{N_x}$. We develop a Bayesian model \mathcal{M} that aims to explain all data \mathbf{d} . We assume that the model output is a random variable y and that \mathbf{d} is an observation of this variable. We have

$$y_j = F(\mathbf{x}_j; \boldsymbol{\vartheta}_c) + \sigma \varepsilon_j, \quad j = 1, \dots, N, \quad (1)$$

where F is the output of the computational model that depends on a set of parameters $\boldsymbol{\vartheta}_c$ and a set of input variables or conditions $\mathbf{x} \in \mathbb{R}^{N_x}$, and the ε_j are independent random variables following $\mathcal{N}(0, 1)$ [62]. The additive random term models the collective effect of all sources of errors and uncertainty discussed in the Introduction. The variable σ is the standard deviation of the random term. We call all variables other than $\boldsymbol{\vartheta}_c$ statistical parameters and we denote them by $\boldsymbol{\vartheta}_s$. The vector of all parameters is denoted by $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_c, \boldsymbol{\vartheta}_s)$ and is considered a random variable.

In the context of the current work, the computational model is the RBC model and its output is the diameter extensions or the tank-treading frequency (TTF) in the stretching and shear flow simulations, respectively. The computational model parameters correspond to the mechanical parameters of the RBC model, and the input variables to the conditions of the simulation, e.g., the applied force in the stretching simulations or the shear rate in the shear flow simulations.

Since the parameters $\boldsymbol{\vartheta}$ are random variables, we are interested in finding their distribution conditioned on the

assumption that the output of model \mathcal{M} is equal to the experimental observations, $\mathbf{y} = \mathbf{d}$. Using Bayes' theorem, this conditional distribution is written as (for the sake of a lighter notation, the dependence of the probabilities on the inputs $\mathbf{x}_1, \dots, \mathbf{x}_N$ is omitted)

$$p(\boldsymbol{\vartheta} | \mathbf{y}, \mathcal{M}) = \frac{p(\mathbf{y} | \boldsymbol{\vartheta}, \mathcal{M}) p(\boldsymbol{\vartheta} | \mathcal{M})}{p(\mathbf{y} | \mathcal{M})}, \quad (2)$$

where the terms $p(\mathbf{y} | \boldsymbol{\vartheta}, \mathcal{M})$, $p(\boldsymbol{\vartheta} | \mathcal{M})$, and $p(\mathbf{y} | \mathcal{M})$ are the likelihood function, the prior probability density, and the model evidence factor, respectively. The *prior* beliefs regarding the model parameters are updated in light of the experimental data, to yield a data-informed distribution for the parameters. This prior belief is updated by the experimental data through the *likelihood* function $p(\mathbf{y} | \boldsymbol{\vartheta}, \mathcal{M})$, yielding the *posterior* distribution $p(\boldsymbol{\vartheta} | \mathbf{y}, \mathcal{M})$, which represents our knowledge about the parameters after considering the data.

The likelihood function $p(\mathbf{y} | \boldsymbol{\vartheta}, \mathcal{M})$ represents how probable the observations \mathbf{y} are for a given parameter set $\boldsymbol{\vartheta}$. Under the assumption of Eq. (1), the likelihood function is given by

$$p(\mathbf{y} | \boldsymbol{\vartheta}, \mathcal{M}) = \mathcal{N}[\mathbf{y} | \boldsymbol{\mu}(\boldsymbol{\vartheta}_c), \boldsymbol{\Sigma}(\boldsymbol{\vartheta}_s)], \quad (3)$$

where $\boldsymbol{\mu}(\boldsymbol{\vartheta}_c) = [F(\mathbf{x}_1; \boldsymbol{\vartheta}_c), \dots, F(\mathbf{x}_N; \boldsymbol{\vartheta}_c)]$, $\boldsymbol{\Sigma}(\boldsymbol{\vartheta}_s) = \sigma^2 I_N$, $\boldsymbol{\vartheta}_s = \sigma$, and I_N is the identity matrix in $\mathbb{R}^{N \times N}$.

Once the posterior distribution of the parameters is estimated, we can propagate the model uncertainty to any output quantity of interest \mathbf{y}^{new} :

$$\begin{aligned} p(\mathbf{y}^{\text{new}} | \mathbf{y}, \mathcal{M}) &= \int p(\mathbf{y}^{\text{new}} | \boldsymbol{\vartheta}) p(\boldsymbol{\vartheta} | \mathbf{y}, \mathcal{M}) d\boldsymbol{\vartheta} \\ &\approx \frac{1}{N_s} \sum_{k=1}^{N_s} p(\mathbf{y}^{\text{new}} | \boldsymbol{\vartheta}^{(k)}). \end{aligned} \quad (4)$$

Here the $\boldsymbol{\vartheta}^{(k)} \sim p(\boldsymbol{\vartheta} | \mathbf{y}, \mathcal{M})$ are samples obtained from the posterior distribution.

The relation of the variables involved in the Bayesian inference is usually described through a directed acyclic graph (Fig. 1). The variables that are unobserved, observed, and propagated are denoted by a circle, square, and shaded circle, respectively. We refer to these models as *single-level* Bayesian models, as opposed to the hierarchical ones discussed in the next section.

2. Hierarchical Bayesian models

When multiple data sets are available, one approach is to pool all data together and use the single-level methodology of the previous section. However, this approach has been shown to be inaccurate for atomistic and coarse-grain models [42,63]. The data heterogeneity can be attributed to data

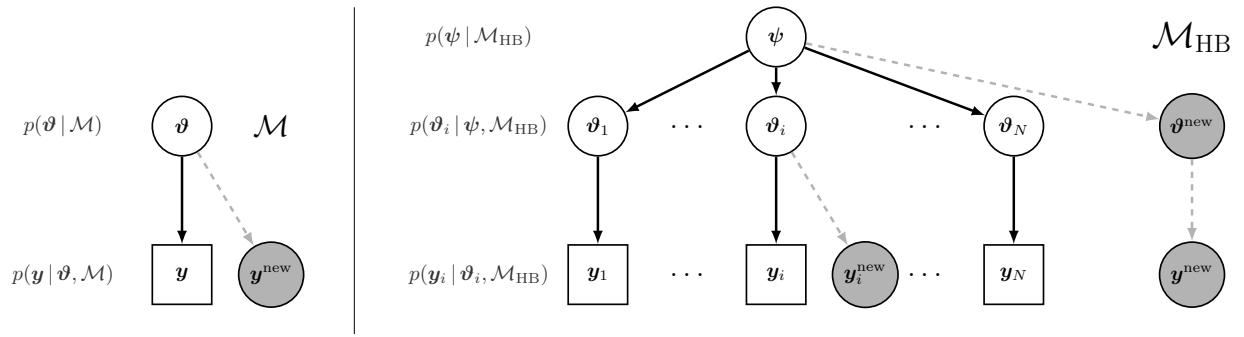


FIG. 1. Directed acyclic graphs describing two Bayesian models. Left: a single-level model with observation variable \mathbf{y} corresponding to experimental data \mathbf{d} . Right: a hierarchical Bayesian model with observations \mathbf{y}_i corresponding to experimental data \mathbf{d}_i . The data may correspond to different experimental conditions or even completely different experiments. Each data set \mathbf{d}_i is represented with different parameters ϑ_i and the parameters are connected through the hyperparameters ψ .

sets reflecting different physical quantities and/or experimental conditions. Model parameters need to account for this heterogeneity and at the same time express the same physical system, the RBC in this study. This dual role for the model parameters is addressed through the *hierarchical Bayesian* (HB) framework.

We denote the collection of data sets as $\vec{\mathbf{d}} = \{\mathbf{d}_1, \dots, \mathbf{d}_{N_d}\}$. Each $\mathbf{d}_i = (d_{i,1}, \dots, d_{i,N_i}) \in \mathbb{R}^{N_i}$ is the data vector corresponding to the data set i , and $d_{i,j}$ corresponds to input conditions $\mathbf{x}_{i,j}$ for $i = 1, \dots, N_d$ and $j = 1, \dots, N_i$. The HB model \mathcal{M}_{HB} introduces a vector of hyperparameters ψ , following a distribution $p(\psi | \vec{\mathbf{d}}, \mathcal{M}_{\text{HB}})$, that encodes the information exchanged between data sets. This hyperparameter vector is then used to specify a prior probability distribution on ϑ_i , $p(\vartheta_i | \psi, \mathcal{M}_{\text{HB}})$, as shown on the right of Fig. 1. Through the HB framework we can hence obtain better informed individual posterior distributions $p(\vartheta_i | \vec{\mathbf{d}}, \mathcal{M}_{\text{HB}})$, using information from all data sets. Finally, the hyperparameter prior can be used to obtain a new general set of parameters ϑ^{new} corresponding to unobserved data. The details of sampling from this Bayesian model are covered in Ref. [63]. For the sake of completeness, we have added a section in Appendix F.

3. Sampling methods

The posterior probability distributions are estimated through BASIS [64], a variant of the transitional Markov chain Monte Carlo algorithm [65]. The BASIS algorithm is deployed inside *korali* [47], a high-performance framework for uncertainty quantification. We apply a special technique to sample the parameters, developed in Ref. [49], that allows us to reduce the computational expenses associated with hierarchical Bayesian inference. The details for the setup of the BASIS algorithm are given in Appendix F.

C. Surrogates for expensive models: Gaussian processes

The computational cost of Bayesian UQ is determined by the sampling of the likelihood function [Eq. (3)]. Here, sampling the respective high-dimensional spaces involves numerous evaluations of the RBC model and each is computationally intensive. We reduce the number of expensive model evaluations through surrogate models based on GP regression [66]. GPs are a widely used approach for generating surrogate models to mitigate the computational cost of Bayesian inference [67–71]. We construct GP surrogates for the following quantities of interest: (i) the thickness of a resting RBC, expressed in terms of the deviation from the analytical approximation of Ref. [56] [Eq. (6)], denoted by z^{eq} , (ii) the axial and (iii) transverse diameter extensions of a RBC under stretching, denoted by z^{ax} and z^{tr} , and (iv) the TTF and (v) inclination angle of RBCs under shear, denoted by z^{sh} and z^{ϕ} . The inputs for each GP are a set of nondimensional RBC parameters ϑ_c^* and conditions \mathbf{x}^* . The GP output z is a Gaussian distribution with mean m_z and standard deviation σ_z . Each GP is trained on N_{sim} simulations, performed at preselected values of ϑ_c^* and \mathbf{x}^* , obtained from a latin hypercube (LH) in the $(\vartheta_c^*, \mathbf{x}^*)$ space. The training process serves for the optimization of the hyperparameters of each GP. Details for the training of the GPs and the LH construction are given in Table IV in Appendix E.

The prediction accuracy of each GP is assessed through cross validation (CV) with repeated random subsampling [72], also known as Monte Carlo CV [73]. During the CV, the GPs are trained on a randomly selected subset comprising 90% of the simulation data, and their prediction is tested on the remaining 10%. This process is repeated 100 times, and statistics are collected. The results show a small prediction error and uncertainty for all surrogates considered (with an average between 1%–4% of the target value), suggesting that the means of the GPs provide an acceptable approximation to the simulation outputs.

Details for the construction and training of the GPs are given in Appendix E.

III. PARAMETRIC INFERENCE

The RBC model of Sec. II A 1 has seven parameters, two of which are the global area and volume constraints, set to $K_A = 0.5 \text{ J/m}^2$ and $K_V = 7.23 \times 10^5 \text{ J/m}^3$. These values are obtained from Ref. [2] and result in global area and volume deviations of less than 0.5% in all considered simulations. There is considerable uncertainty for the value of the stretch (local area) modulus k_a . Experimental studies summarized in Ref. [2] estimate $k_a = 2\text{--}13 \mu\text{N/m}$. However, this estimate corresponds to measurements on bare membrane skeletons, and thus applicability to intact RBCs is unknown. Previous computational studies have assumed either a weakly compressible ($C = k_a/\mu_0 \sim 1$) or an incompressible membrane (see Appendix B). We set $C = 1$ in all simulations, to reduce the computational cost associated with the introduction of an additional parameter in the UQ study. Further studies could unveil the effects of local area conservation, and infer k_a given experimental data.

The remaining parameters are the equilibrium-to-maximum spring length ratio x_0 , the shear modulus μ_0 , the bending modulus κ_b , and the membrane surface viscosity η_m^{2D} . We infer the distribution of these parameters based on experimental data of RBCs (i) under stretching, and (ii) in shear flow, a summary of which is given in Table II and plotted in Fig. 2. Afterwards, the inferred distributions are used to predict data of resting RBCs at equilibrium, and the inclination angle of tank-treading RBCs in shear flow.

A. Stretching

We first perform the inference of parameters related to the elastic properties of the RBC: x_0 , μ_0 , and κ_b . We simulate a RBC under stretching and compare its axial and transverse extensions [Fig. 3(a)] with experimental data

TABLE II. Experimental data considered for the UQ inference. The symbols refer to Fig. 2. The solvent viscosity η_0 is given in units of mPa s, and the assumed cell area A_0 in μm^2 . For all experiments where the area is not explicitly given, we assume that $A_0 = 135 \mu\text{m}^2$ [56]. The last column denotes the data set ID, as used in the UQ. Details for the extraction of data from the respective references are given in Appendix D.

Reference	Symbol	η_0	A_0	ID
Stretching				
Mills <i>et al.</i> [17]	O	d ₁
Suresh <i>et al.</i> [74]	×	d ₂
Shear flow				
Fischer <i>et al.</i> [75]	...	11	135	...
Fischer <i>et al.</i> [75]	...	18	135	...
Fischer <i>et al.</i> [75]	☆	31	135	d ₃
Fischer <i>et al.</i> [75]	...	59	135	...
Fischer [76]	...	23	178	...
Tran-Son-Toy [77]	...	20	135	...
Tran-Son-Toy <i>et al.</i> [78]	...	35	141	...
Fischer [79] (donor 3)	◇	28.9	135	d ₄
Fischer <i>et al.</i> [80]	...	12.9	135	...
Fischer <i>et al.</i> [80] (donors 1, 2)	△	28.9	135	d ₅
Fischer <i>et al.</i> [80]	...	53.9	135	...
Fischer <i>et al.</i> [80]	...	109.3	135	...

of healthy RBCs stretched with optical tweezers [17,74]. In the experiments, a RBC is stretched by two silica microbeads attached at opposite sides of its rim and pulled in opposite directions by an externally applied force F_{ext} . We simulate this setup by following the procedure proposed by Sigüenza *et al.* [34], assuming a circular contact area between the bead and the RBC with diameter d_c . A stretching force is then applied only to the RBC vertices located on the edge of the contact area, mimicking the rigidity of the beads. We note that d_c is not determined experimentally. As suggested by an anonymous reviewer, a possible treatment could be to assume uncertainty about

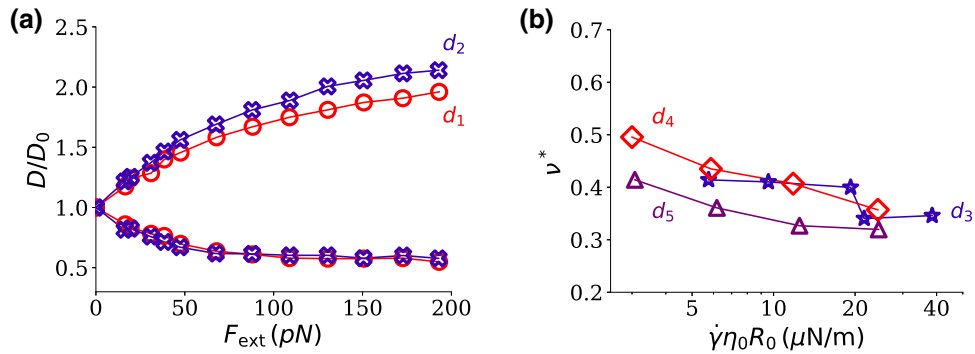


FIG. 2. (a) Experimental data sets considered for the inference on stretching. The y axis shows the axial and transverse extensions normalized by the diameter at rest, D/D_0 . (b) Experimental data sets considered for the inference on shear flow. The y axis shows the dimensionless TTF, $\nu^* = 4\pi/(\dot{\gamma}T_{\text{tt}})$, with T_{tt} the tank-treading period. Error bars are omitted for clarity, and can be obtained from the respective references in Table II.

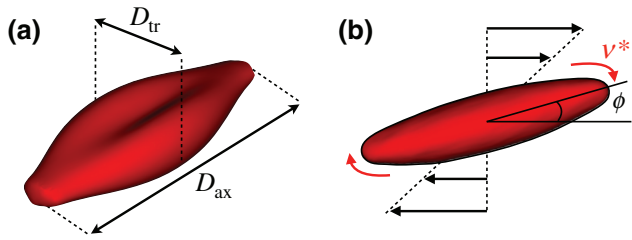


FIG. 3. Simulation snapshots from RBCs under (a) stretching and (b) in shear flow, depicting the axial (D_{ax}) and transverse (D_{tr}) diameters, dimensionless TTF (v^*), and inclination angle (ϕ).

the value of d_c and account for this in the Bayesian inference. To reduce the computational cost associated with the inclusion of an additional parameter in the inference, we fix $d_c = 2 \mu\text{m}$ [34].

The elasticity and extension of a stretched RBC can be described by four dimensionless numbers: the equilibrium-to-maximum spring ratio x_0 , the Föppl-von Kármán (FvK) number, the dimensionless local compressibility (C) (here fixed at $C = 1$; see Sec. III), and the nondimensional force F^* . We define $\text{FvK} = \mu_0 R_0^2 / \kappa_b$ and $F^* = F_{\text{ext}} / (\mu_0 R_0)$, with $R_0 = [A_0 / (4\pi)]^{1/2}$ the effective radius of the RBC. The nondimensional extensions are defined as $D_{\text{ax}}^* = D_{\text{ax}} / D_0$ and $D_{\text{tr}}^* = D_{\text{tr}} / D_0$, with D_0 the diameter of a RBC at rest.

The computational cost of the Bayesian inference is reduced by forming two surrogate models for the axial and transverse extensions using GPs. The GPs are trained on a set of simulations with preselected parameters obtained from a LH [81] (details in Table IV in Appendix E). Since the stretching depends on the aforementioned set of dimensionless numbers, the simulations are performed in the region 0.2–0.7 for x_0 , 20–450 for FvK, and 0–30 for F^* .

The mean of the GP predictions for D_{ax}^* and D_{tr}^* is shown in Fig. 4. The diameter extension is characterized by two regimes [31]: at low applied forces the extension is linearly dependent on the applied force [82], while at larger forces the RBC extension follows a nonlinear, strain-hardening response. The extensions in the nonlinear regime are highly affected by x_0 , as shown in Fig. 4 for $F^* = 15$ and $F^* = 30$, with lower x_0 (softer cells) yielding larger D_{ax}^* . Conversely, the combination of increasing D_{ax}^* and global area and volume conservation results in decreasing D_{tr}^* with decreasing x_0 . The FvK number has negligible effect on the extensions, indicated by the nearly horizontal isolines of D_{ax}^* and D_{tr}^* . Our results corroborate the findings of Siguéza *et al.* [34], who demonstrated that the inclusion of bending rigidity changes the cross-section shape of the RBC but does not have a considerable effect on the axial and transverse extensions.

The experimental data sets considered in the inference [17,74] are shown in Fig. 2(a) and Table II. The collection

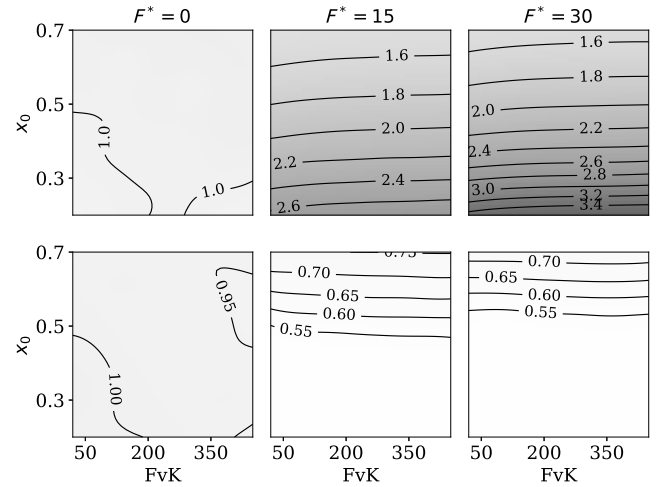


FIG. 4. Two-dimensional (2D) slices from the 3D GP surrogate of the relative extensions D_{ax}^* (top) and D_{tr}^* (bottom) of a RBC under stretching. The color corresponds to the extensions, with low (high) values represented by white (gray). Lines represent isolines of the extensions. Columns correspond to different extensional forces F^* , denoted at the top of the figure.

of data for the axial and transverse extensions will be denoted as $\mathbf{d}_i = \{\mathbf{d}_i^{\text{ax}}, \mathbf{d}_i^{\text{tr}}\}$, where $i = 1$ corresponds to the data of Mills *et al.* [17] and $i = 2$ to the data of Suresh *et al.* [74].

The statistical model for the stretching setup is defined as

$$y_{ij}^{\text{ax}} = m_{z^{\text{ax}}}(\mathbf{x}_{ij}, \boldsymbol{\vartheta}_{c,i}) + \sigma_{\text{ax},i} \varepsilon_{ij}^{\text{ax}},$$

$$y_{ij}^{\text{tr}} = m_{z^{\text{tr}}}(\mathbf{x}_{ij}, \boldsymbol{\vartheta}_{c,i}) + \sigma_{\text{tr},i} \varepsilon_{ij}^{\text{tr}},$$

where \mathbf{x}_{ij} , $i = 1, 2$, $j = 1, \dots, N_i$, is the input of the j th data point within the i th data set, and $\varepsilon_{ij}^{\text{ax}}$, $\varepsilon_{ij}^{\text{tr}}$ are independent random variables following $\mathcal{N}(0, 1)$. Additionally, $\boldsymbol{\vartheta}_{c,i} = (x_{0,i}, \mu_{0,i}, \kappa_{b,i})$ and $\mathbf{x}_{ij} = F_{ij}^* = F_{\text{ext},ij} / (\mu_{0,i} R_{0,i})$ for $i = 1, 2$. The relation of the variables involved in the inference is described through the directed acyclic graph in Fig. 5. We note that introducing correlation between the error random variables $\varepsilon_{ij}^{\text{ax}}$, $\varepsilon_{ij}^{\text{tr}}$ showed no significant improvement in the posterior distributions and the propagation. Therefore, the complete parameter set used in the inference for stretching is defined by $\boldsymbol{\vartheta}_i = (x_{0,i}, \mu_{0,i}, \kappa_{b,i}, \sigma_{\text{ax},i}, \sigma_{\text{tr},i})$.

We adopt a uniform distribution for all priors for $\boldsymbol{\vartheta}$, $p(\boldsymbol{\vartheta} | \mathcal{M}_i)$, so that $p(x_0 | \mathcal{M}_i) = \mathcal{U}(0.2, 0.7)$ [a random variable ε following a uniform distribution with lower and upper bounds α and β is denoted as $\varepsilon \sim \mathcal{U}(\alpha, \beta)$], $p(\mu_0 | \mathcal{M}_i) = \mathcal{U}(2, 6) \mu\text{N/m}$, $p(\kappa_b | \mathcal{M}_i) = \mathcal{U}(1.5, 10) \times 10^{-19} \text{ J}$, $p(\sigma_{\text{ax}} | \mathcal{M}_i) = \mathcal{U}(0, 0.5)$, and $p(\sigma_{\text{tr}} | \mathcal{M}_i) = \mathcal{U}(0, 0.5)$. These bounds cover the range of experimentally estimated values for μ_0 and κ_b as summarized in Appendix A of Ref. [2], while the bounds for x_0 are based on previous studies of the same RBC model [19,20].

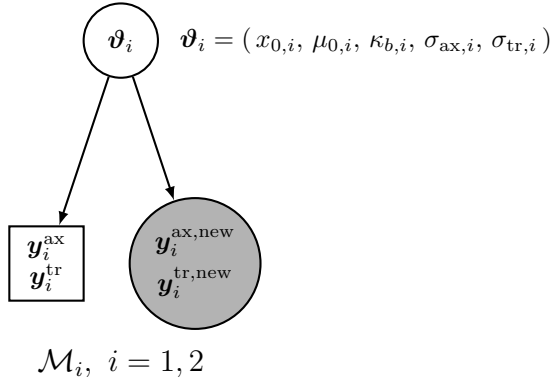


FIG. 5. Single-level Bayesian model \mathcal{M}_i used in the inference with stretching data.

The results of the inference are depicted in Figs. 6(a) and 6(b). The marginal distributions of x_0 and μ_0 are characterized by narrow peaks, suggesting that x_0 and μ_0 are inferred with high certainty from the stretching data. This, in turn, indicates that D_{ax}^* and D_{tr}^* are highly affected by x_0 and μ_0 , in agreement with earlier observations from the GP stretching surrogates [19,31,82]. In contrary, κ_b is found to be unidentifiable from the stretching data. This is shown in Figs. 6(a) and 6(b) by the spread of the posterior distribution along the entire range of κ_b , and a high log likelihood across a wide range of κ_b values. Lastly, the noise associated with the prediction of the axial data is smaller than for the transverse, as seen by the marginal distributions of σ_{ax} and σ_{tr} . This is possibly because the axial extensions (larger values) have a larger contribution to the likelihood than the transverse (smaller values).

The uncertainty of ϑ_i , $i = 1, 2$, is propagated to the model output $p(\mathbf{y}_i^{new} | \mathbf{d}_i, \mathcal{M}_i)$ using Eq. (4). The results are shown in Figs. 6(c) and 6(d) by the expected value of the prediction along F^* and the uncertainty associated with four credible intervals (CIs). The 90% CIs are relatively narrow, implying a strong belief in the model prediction. Additionally, the experimental data lie within the bounds of the 90% CIs.

B. Shear flow

We report the inference of the RBC parameters x_0 , μ_0 , κ_b , and η_m^* under dynamic flow conditions. We choose to simulate a simple shear flow setup [Fig. 3(b)], due to the large number of available experimental data sets with measurements for the TTF of RBCs in shear flow [75–80].

In the simulation, a single RBC is suspended in a solution of viscosity η_o , enclosing a liquid with viscosity η_i . The simulations are performed in a cubic domain with length $L = 10R_0$, chosen to minimize wall-induced confinement effects [26,83–85]. A linear shear flow with shear

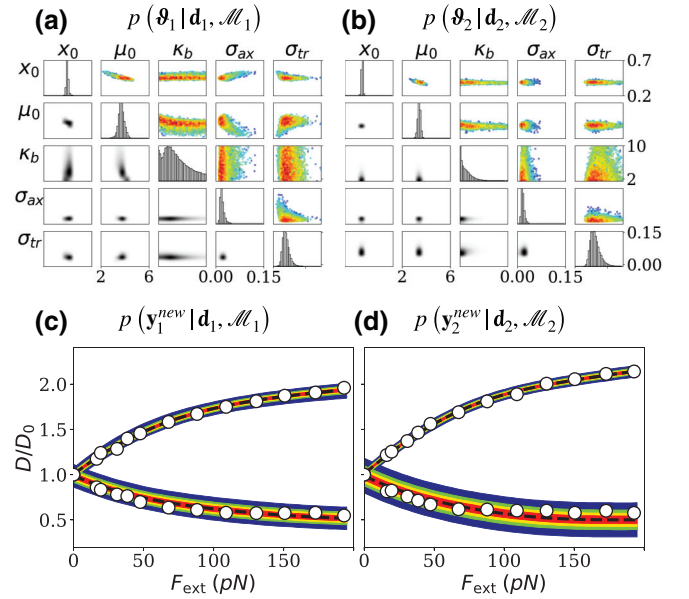


FIG. 6. Summary of the single-level Bayesian inference based on the stretching data sets. (a),(b) Marginal distribution with respect to each parameter (*diagonal*), two-dimensional projections of the posterior colored by the probability density [black (white) denotes high (low) density] (*below diagonal*), and samples from the posterior distribution colored by the log likelihood [red (blue) denotes high (low) log likelihood] (*above diagonal*). For tiles in every even column (e.g., μ_0 , σ_{ax}), the x -axis range is given at the bottom of the figure. For tiles in every odd row (e.g., x_0 , κ_b , σ_{tr}), the y -axis range is given at the right of the figure. (c),(d) Propagation of the uncertainty in ϑ to the prediction of D_{ax}^* and D_{tr}^* , $p(\mathbf{y}_i^{new} | \mathbf{d}_i, \mathcal{M}_i)$, $i = 1, 2$. The colored areas denote the 99% (blue), 90% (green), 75% (yellow), and 50% (red) credible intervals. Circles correspond to the experimental data. The dashed line corresponds to the expected (mean) prediction.

rate $\dot{\gamma}$ is created by two planar walls, moving at a constant velocity $\mathbf{v} = \{\dot{\gamma}L/2, 0, 0\}$ in opposite directions, with periodic boundary conditions along x and z . Technical details regarding the implementation of the flow setup with the DPD method are given in Appendix A.

The dimensionless numbers controlling the dynamics of RBCs [32,85,86] and capsules [87–89] in simple shear flow, as well as their stability and transition from one state to another [89–91], are the capillary (Ca) number, the FvK number, the viscosity ratio (λ), the dimensionless membrane viscosity (η_m^*), and the stress-free shape [53,54]. In this study we define $\text{Ca} = \dot{\gamma}\eta_o R_0 / \mu_0$, $\lambda = \eta_i / \eta_o$, and $\eta_m^* = \eta_m^{2D} / (\eta_i R_0)$ (η_m^* is defined with respect to the inner viscosity, since the outer solvent viscosity varies between experimental data sets).

The computational cost of the inference is reduced significantly by the GP surrogates for the dimensionless TTF, $v^* = 4\pi / (\dot{\gamma} T_{tt})$, where T_{tt} is the tank-treading period [33,85]. Similarly to the stretching surrogate GPs, the TTF GP is trained on a set of simulations with preselected

parameters obtained from a LH. The simulations are performed in the region 0.2–0.7 for x_0 , 20–450 for FvK, 0–20 for η_m^* , and 0.5–20 for Ca, covering the region of FvK numbers [86] and η_m^* [78,92] relevant to RBCs. The viscosity ratio is set to $\lambda = 0.32$ (see the discussion in Appendix C).

The Reynolds number $Re = \dot{\gamma} R_0^2 \rho / \eta_o$ in the shear flow experiments is approximately 10^{-4} . The simulation of low Reynolds number flows with DPD implies very small time steps (due to the usage of an explicit time-integration scheme), leading to excessively long simulation run times. A common practice is to scale down the viscosities ($\eta_o, \eta_i, \eta_m^{2D}$) by multiplying with a factor $f_{sc} < 1$, while at the same time, increasing $\dot{\gamma}$ by dividing it with f_{sc} , such that the Ca number is preserved. While this kind of scaling preserves the Ca number, it increases quadratically the Reynolds (Re) number. In all simulations we maintain $Re < 0.1$ such that inertial effects are negligible [84], and $Ma < Re$ to preserve the separation of viscous and sonic time scales [93].

The predictions from the GP surrogate of the TTF are shown in Figs. 7 and 8. The trends observed in ν^* are qualitatively similar to the computational results of Yazdani *et al.* [33] for membranes as a function of λ , and as a function of η_m^* [85]. In particular, ν^* follows a nonmonotonic trend as a function of η_m^* , with an initial decrease and a subsequent increase. Additionally, the η_m^* corresponding to

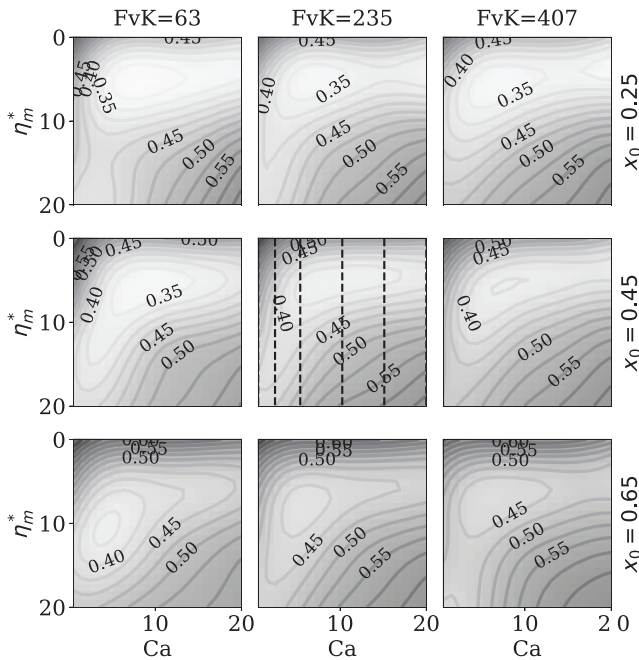


FIG. 7. 2D slices from the 4D GP surrogate of the TTF. The color corresponds to ν^* , with low (high) values represented by white (gray). Lines represent isolines of ν^* . Columns and rows correspond to fixed FvK and x_0 values, shown on the top and right of the figure. The dashed lines in the center tile correspond to the slices shown in Fig. 8.

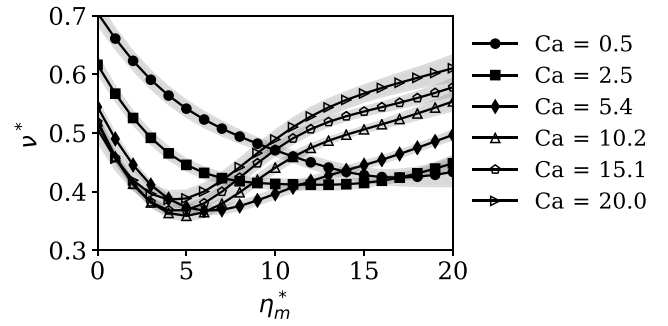


FIG. 8. TTF over η_m^* at $x_0 = 0.45$ and FvK = 235. Points denote predictions from the GP surrogate. Shaded areas show the respective uncertainty in the prediction.

the minimum ν^* decreases with increasing Ca. At low η_m^* , ν^* seems to be insensitive to the Ca number for large Ca ≥ 10 . At high η_m^* , ν^* increases with Ca, a trend not shown in the results of Refs. [33,85], who however performed simulations for up to Ca ≈ 0.5 corresponding to the minimum Ca considered in our study. We also observe the emergence of wrinkles with increasing η_m^* , and their disappearance with increasing bending (thus decreasing FvK number), consistent with the observations in Ref. [85].

The data sets considered in the inference are shown in Fig. 2(b) and Table II. Details for the extraction of the data from the respective references are given in Appendix D, while a discussion on qualitative differences observed between these data sets can be found in Ref. [33]. The dimensionless number characterizing the shear stress sustained by a RBC in shear flow is the Ca number. However, as μ_0 is one of the parameters that will be inferred, we present the experimental data in terms of $\dot{\gamma} \eta_o R_0$.

The collection of TTF data sets is denoted as $\mathbf{d}_i = \mathbf{d}_i^{\text{sh}}$, $i = 3, \dots, 5$. The statistical model for shear flow is

$$y_{ij}^{\text{sh}} = m_{z^{\text{sh}}}(\mathbf{x}_{ij}, \boldsymbol{\vartheta}_{c,i}) + \sigma_{\text{sh},i} \varepsilon_{ij}, \quad (5)$$

where $\varepsilon_{ij} \sim \mathcal{N}(0, 1)$ and $\boldsymbol{\vartheta}_{c,i} = (x_{0,i}, \mu_{0,i}, \kappa_{b,i}, \eta_{m,i}^*)$ for $i = 3, \dots, 5$ and $j = 1, \dots, N_i$. The input \mathbf{x}_{ij} , $i = 3, 4, 5$, $j = 1, \dots, N_i$, corresponds to the input of the j th data point within the i th data set, and is equal to $\text{Ca}_{ij} = \dot{\gamma}_{ij} \eta_{0,i} R_{0,i} / \mu_{0,i}$. The complete parameter set used in the inference for shear flow is $\boldsymbol{\vartheta}_i = (x_{0,i}, \mu_{0,i}, \kappa_{b,i}, \eta_{m,i}^*, \sigma_{\text{sh},i})$. The relation of the variables involved in the inference is graphically shown in Fig. 9.

The prior distributions for the parameters are chosen to follow a uniform distribution. In particular, $p(x_0 | \mathcal{M}_i) = \mathcal{U}(0.2, 0.7)$, $p(\mu_0 | \mathcal{M}_i) = \mathcal{U}(2, 6)$ $\mu\text{N/m}$, $p(\kappa_b | \mathcal{M}_i) = \mathcal{U}(1.5, 10) \times 10^{-19}$ J, $p(\eta_m^* | \mathcal{M}_i) = \mathcal{U}(0, 20)$, and $p(\sigma_{\text{sh}} | \mathcal{M}_i) = \mathcal{U}(0, 0.1)$. These bounds cover the range of experimentally estimated values for μ_0 and κ_b as summarized in Appendix A of Ref. [2], while for η_m^* , we base the prior on estimates of the membrane surface viscosity, ranging between $0.05\text{--}0.8 \times 10^{-6}$ Pa s m [78,92].

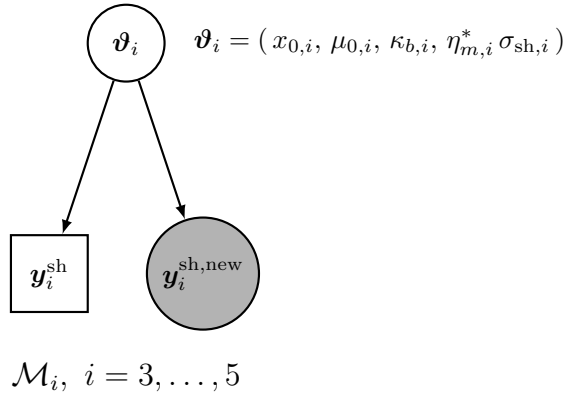


FIG. 9. Single-level Bayesian model \mathcal{M}_i used in the inference with TTF data.

The posterior distribution $p(\vartheta_i | \mathbf{d}_i, \mathcal{M}_i)$, $i = 3, \dots, 5$, is shown in Figs. 10(a)–10(c). The marginal distributions of η_m^* and the two-dimensional projections of the posterior along (x_0, η_m^*) , (μ_0, η_m^*) , and (κ_b, η_m^*) show that all TTF data sets contain information to identify a unimodal posterior for η_m^* . Additionally, the uncertainty in η_m^* is low

for $i = 3, 4$, shown by the small spread of the marginal distribution and the narrow regions of high likelihood centered around $\eta_m^* \approx 8-9$. On the other hand κ_b is unidentifiable, as shown by the nearly flat marginal distributions and high likelihood extending over the entire range of κ_b . We find no evidence in the literature that κ_b has a significant effect on the TTF, which corroborates our Bayesian inference results. Lastly, it is not clear whether x_0 and μ_0 are identifiable from the TTF data. The posterior for data set \mathbf{d}_4 has higher likelihood in the inner region of the parameter space (see the corresponding marginal distributions and 2D projections along x_0 and μ_0). In contrast, for data sets \mathbf{d}_3 and \mathbf{d}_5 , high-likelihood regions appear at the boundaries of the parameter space. Between data sets with the same number of data points, \mathbf{d}_4 and \mathbf{d}_5 , the highest model evidence [Eq. (2)] corresponds to \mathbf{d}_4 . This indicates that the GP surrogate of the TTF can explain better the data of \mathbf{d}_4 .

The conditional distributions of the model output, $p(\mathbf{y}_i^{\text{new}} | \mathbf{d}_i, \mathcal{M}_i)$ for $i = 3, \dots, 5$, are presented in Figs. 10(d)–10(f). The expected value of the prediction over $\dot{\gamma}\eta_0 R_0$ passes through the experimental data of \mathbf{d}_3 and \mathbf{d}_4 [Figs. 10(d) and 10(e)], whereas it over predicts \mathbf{d}_5 [Fig. 10(f)]. In particular, the data points of \mathbf{d}_5 lie on

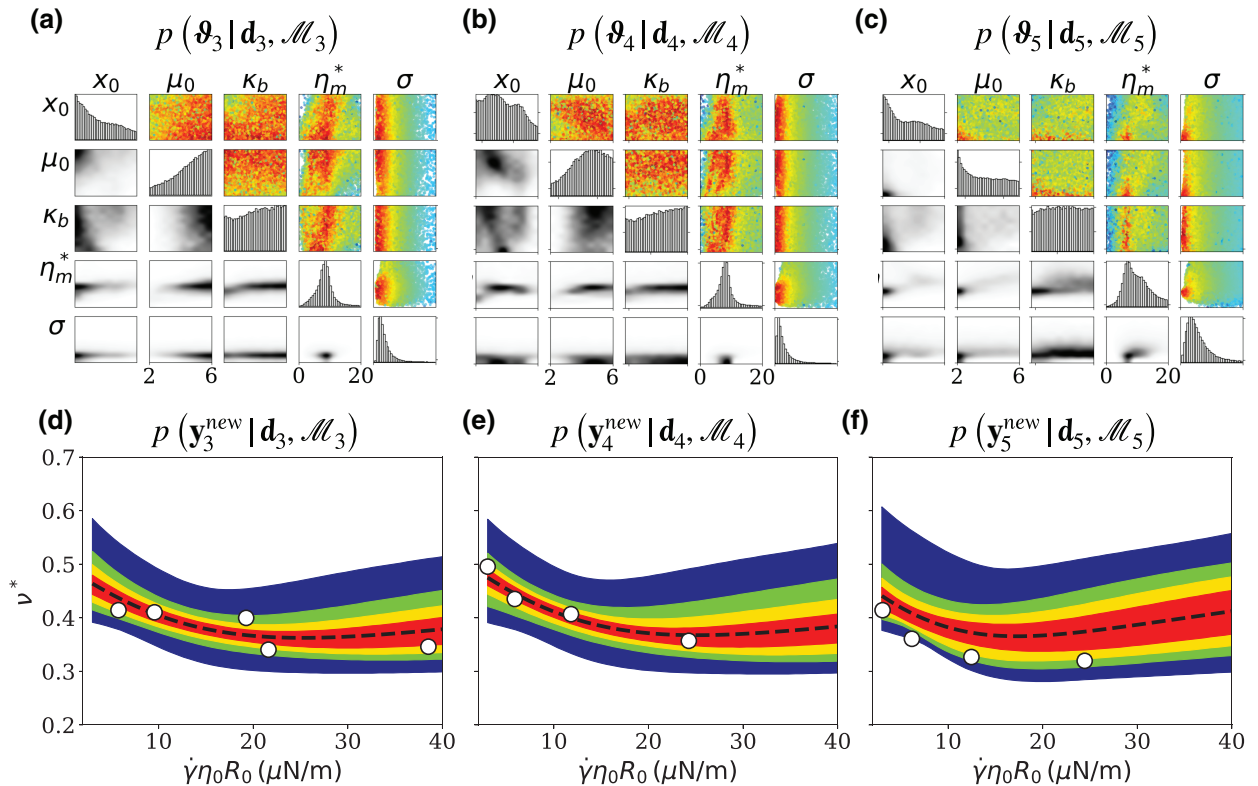


FIG. 10. Summary of the single-level Bayesian inference based on the TTF data sets. (a)–(c) Posterior distributions $p(\vartheta_i | \mathbf{d}_i, \mathcal{M}_i)$, $i = 3, \dots, 5$. Figure descriptions follow Figs. 6(a) and 6(b). (d)–(f) Propagation of the uncertainty in ϑ to the prediction of TTF, $p(\mathbf{y}_i^{\text{new}} | \mathbf{d}_i, \mathcal{M}_i)$, $i = 3, \dots, 5$. The colored areas denote the 99% (blue), 90% (green), 75% (yellow), and 50% (red) credible intervals. Circles correspond to the experimental data. The dashed line is the expected (mean) prediction.

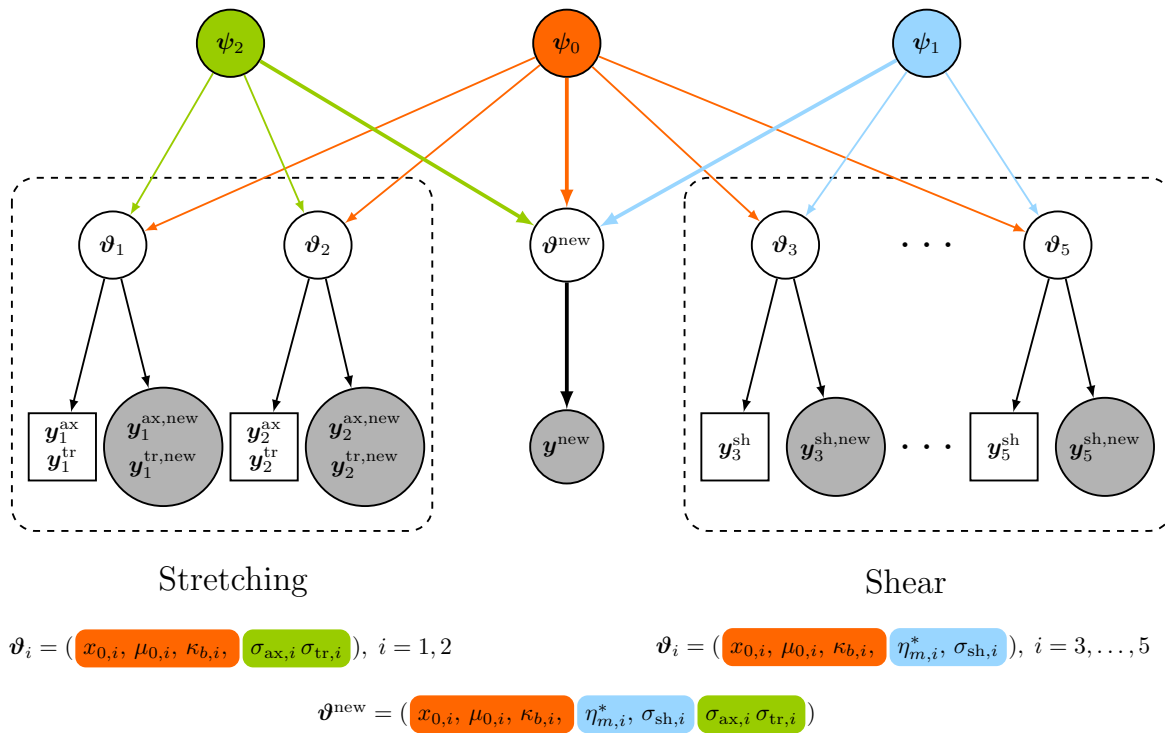


FIG. 11. Directed acyclic graph for the hierarchical Bayesian model \mathcal{M}_{HB} used in this study. Details are given in Sec. III C.

the edge of the 99% credible interval, reflecting that high-likelihood regions of the posterior, Fig. 10(c), are at the boundaries of the parameter space.

C. Hierarchical Bayesian inference

In this section we infer the RBC parameters x_0 , μ_0 , κ_b , η_m^* and statistical parameters σ_{sh} , σ_{ax} , σ_{tr} by constructing a hierarchical Bayesian model \mathcal{M}_{HB} (Fig. 11). The complete parameter set in \mathcal{M}_{HB} is $\boldsymbol{\vartheta} = (x_0, \mu_0, \kappa_b, \eta_m^*, \sigma_{\text{sh}}, \sigma_{\text{ax}}, \sigma_{\text{tr}})$. For the inference, we use two data sets from the stretching experiment, $(y_i^{\text{ax}}, y_i^{\text{tr}}) = \mathbf{d}_i$, $i = 1, 2$, and three data sets from the shear flow experiment, $y_i^{\text{sh}} = \mathbf{d}_i$, $i = 3, \dots, 5$ (Table II).

In \mathcal{M}_{HB} , the distribution of parameters entering both the stretching and TTF surrogates are controlled by the hyperparameter vector $\boldsymbol{\psi}_0$, shown with orange in Fig. 11. Here η_m^* is an input only to the TTF surrogate, and along with σ_{sh} , is controlled by the hyperparameter vector $\boldsymbol{\psi}_1$ denoted with blue, thus sharing information between all TTF data sets. Similarly, the distribution of σ_{ax} and σ_{tr} is described by $\boldsymbol{\psi}_2$, denoted with green. The parameters $\boldsymbol{\vartheta}$ are related to the hyperparameters $\boldsymbol{\psi} = (\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2)$ through the prior probability distribution $p(\boldsymbol{\vartheta} | \boldsymbol{\psi})$. Details for the distributions used in this study and the obtained hyperparameter distribution $p(\boldsymbol{\psi} | \bar{\mathbf{d}}, \mathcal{M}_{\text{HB}})$ can be found in Appendix F and Fig. 20.

After inferring $p(\boldsymbol{\psi} | \bar{\mathbf{d}}, \mathcal{M}_{\text{HB}})$, a new distribution of computational and statistical parameters $\boldsymbol{\vartheta}^{\text{new}}$ can be

computed, shown by the arrows in Fig. 11 from $\boldsymbol{\psi}_0$, $\boldsymbol{\psi}_1$, and $\boldsymbol{\psi}_2$ towards $\boldsymbol{\vartheta}^{\text{new}}$. The distribution for $\boldsymbol{\vartheta}^{\text{new}}$, $p(\boldsymbol{\vartheta}^{\text{new}} | \bar{\mathbf{d}}, \mathcal{M}_{\text{HB}})$, is shown in Fig. 12. The RBC parameters x_0 , μ_0 , and η_m^* are well identified, shown by the unimodal 1D and 2D marginal distributions for x_0 , μ_0 , and η_m^* . This can be explained from the results of the single-level models, where the parameters controlling the static RBC response, x_0 and μ_0 , are identified well through the stretching data, while η_m^* that controls the dynamic

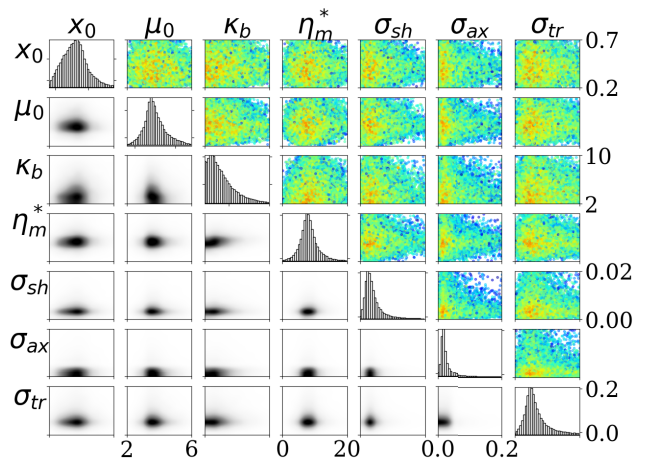


FIG. 12. Posterior distribution for $\boldsymbol{\vartheta}^{\text{new}}$ from the hierarchical Bayesian inference. Figure description follows Figs. 6(a) and 6(b).

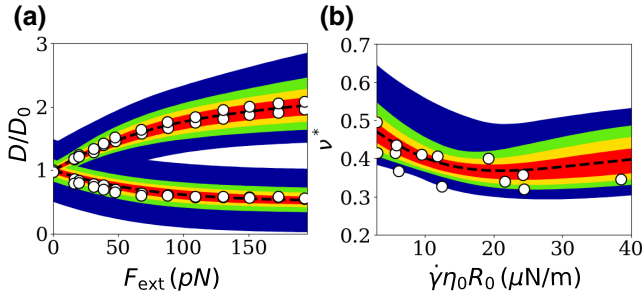


FIG. 13. Propagation of the uncertainty in ϑ^{new} using \mathcal{M}_{HB} to make predictions of (a) the axial and transverse extensions of a RBC under stretching and (b) the TTF of a RBC in shear flow. The colored areas denote the 99% (blue), 90% (green), 75% (yellow), and 50% (red) credible intervals. Circles correspond to *all* the experimental data. The dashed line corresponds to the expected (mean) prediction.

response is identified well through the TTF data. On the other hand, κ_b remains unidentifiable in \mathcal{M}_{HB} , similar to the inference by the single-level models.

We can now use \mathcal{M}_{HB} to make predictions for any quantity of interest, by evaluating the conditional distribution $p(\mathbf{y}^{\text{new}} | \mathbf{d}, \mathcal{M}_{\text{HB}})$, using Eq. (4). In Fig. 13 we present the predictions for the RBC extensions in stretching and the TTF in shear flow, described by the mean prediction of \mathcal{M}_{HB} and the respective credible intervals.

IV. DISCUSSION

A. Comparison of posterior distributions and expected values of RBC parameters

We combine multiple, heterogeneous experimental data sets with the RBC model under the single-level Bayesian and HB frameworks. We compare six Bayesian models: two single-level models from the stretching extensions, three single-level models from the TTF data, and a hierarchical model using all data sets. Using these models, we can now address issues and concerns raised in the

TABLE III. EV and SD of the RBC model parameters, as estimated from the Bayesian models of the inference. The SD is reported in parentheses as a percentage of the respective expected value.

Model	x_0	μ_0 ($\mu\text{N/m}$)	κ_b (10^{-19} J)	η_m^*
\mathcal{M}_1	0.453 (2)	3.745 (7)	4.682 (46)	...
\mathcal{M}_2	0.387 (2)	3.444 (3)	2.684 (48)	...
\mathcal{M}_3	0.384 (36)	4.581 (22)	6.032 (40)	8.571 (33)
\mathcal{M}_4	0.418 (32)	4.357 (23)	6.032 (40)	8.472 (36)
\mathcal{M}_5	0.395 (37)	3.707 (33)	5.792 (42)	10.124 (38)
\mathcal{M}_{HB}	0.405 (23)	3.777 (17)	3.801 (43)	8.446 (33)

Introduction, namely, the variation of RBC parameter values among studies in the literature, discrepancies between experimental data sets, uncertainty in the predictions of the RBC model, and transferability of the predictions of the RBC model between different setups.

In Fig. 14 we show the marginal distribution of the RBC parameters from all Bayesian models, and in Table III we give the corresponding expected values (EVs) and standard deviation (SDs). We observe that the models using the stretching data, \mathcal{M}_1 and \mathcal{M}_2 , yield highly certain predictions for x_0 and μ_0 with peaked marginal distributions [Figs. 14(a) and 14(b)] and thus small SDs, between 2%–7%. On the other hand, models \mathcal{M}_3 – \mathcal{M}_5 , using the TTF data, yield larger SDs around the EVs, between 20%–30%, also seen in the wide marginal distributions of x_0 and μ_0 in Figs. 14(a) and 14(b). Models \mathcal{M}_3 – \mathcal{M}_5 give a unimodal marginal distribution for η_m^* [Fig. 14(d)], with EV around 8–10 and SD between 33%–38%. Lastly, we conclude that κ_b cannot be identified from either the stretching extensions or the TTF data, as the respective SDs denote high uncertainty in its EV, ranging within 40%–50%.

We note how different types of experiments, as well as how different data sets of the same experiment type, can lead to differences in the posterior distribution and EVs of

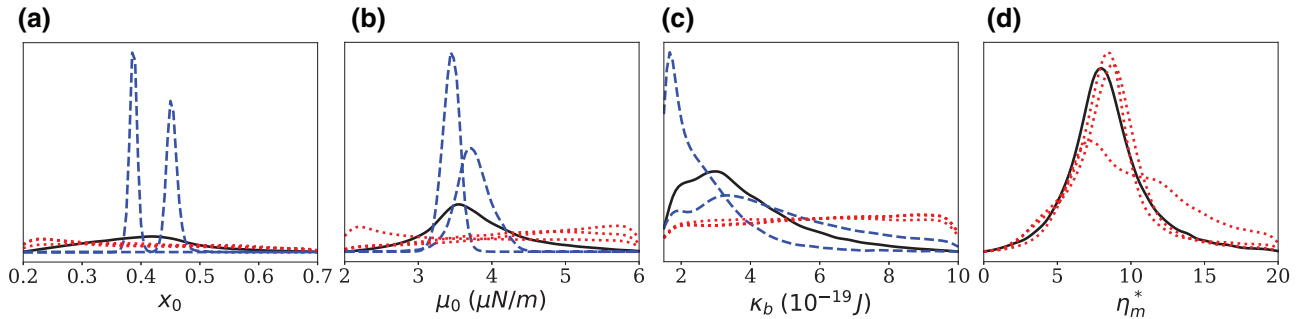


FIG. 14. Marginal posterior probability distributions from the results of the single-level and hierarchical models for the RBC parameters (a) x_0 , (b) μ_0 , (c) κ_b , and (d) η_m^* . Solid black lines represent the hierarchical model, $p(\vartheta^{\text{new}} | \mathbf{d}, \mathcal{M}_{\text{HB}})$. Dashed blue lines represent single-level models for stretching, $p(\vartheta_i | \mathbf{d}_i, \mathcal{M}_i)$, $i = 1, 2$. Dotted red lines represent single-level models for TTF, $p(\vartheta_i | \mathbf{d}, \mathcal{M}_i)$, $i = 3, \dots, 5$.

the RBC parameters. The HB framework provides a solution to this issue as we no longer need to choose among data sets for the calibration of a model, but rather use *all* data sets to estimate a probability distribution for a general parameter set $p(\vartheta^{\text{new}} | \mathbf{d}, \mathcal{M}_{\text{HB}})$. This distribution is broad enough to explain all experimental data, as seen in Fig. 13. The EVs of the RBC parameters from the distribution of ϑ^{new} are shown in the last line of Table III.

B. Transferability of distributions between experimental setups

We assess the transferability of the model by comparing its output with experimental data that have not been used during the inference process. We first present the propagation of uncertainty from the single-level Bayesian model of TTF to the prediction of the RBC extensions in stretching. This propagation is possible as the parameters that enter the stretching surrogate are a subset of those entering the TTF. We then test the predictive capability of the TTF and the hierarchical model on two previously unseen quantities, namely the inclination angle of RBCs in shear flow and the thickness of RBCs at rest.

When propagating the uncertainty of the parameters ϑ_i to obtain predictions \mathbf{y}^{new} of a different quantity of interest, the error term σ_i is scaled with respect to $\max(\mathbf{y}^{\text{new}})/\max(\mathbf{y}_i)$. Taking a relative error magnitude instead of an absolute is necessary as the new quantity of interest does not necessarily have the same units, or order of magnitude as that used in the inference.

1. Stretching prediction

We first demonstrate the results of three uncertainty propagation computations for the prediction of the stretching extensions $\mathbf{y}^{\text{ax}}, \mathbf{y}^{\text{tr}}$. The propagation is performed using

Eq. (4) with the parameter distributions obtained from each of the single-level Bayesian models of TTF, \mathcal{M}_i , $i = 3, \dots, 5$. The results are presented in Fig. 15.

In all cases the mean prediction is a reasonable approximation of the data. However, all propagations are characterized by large uncertainty, especially at large F_{ext} . This can be explained by the uncertainty of the three single-level models, in the EV of x_0 . In this RBC model, the nonlinear regime of the force-extension curve is controlled by x_0 . The posterior distributions of the single-level TTF models have a large uncertainty in x_0 , shown by the large SDs in Table III, which leads to the wide confidence intervals seen in Fig. 15. We therefore conclude that the nonlinear regime of the force-extension curve cannot be predicted well from the information included only in the TTF data.

The extension at low forces (linear force-extension regime) is controlled by μ_0 , as larger μ_0 imply stiffer cells and thus smaller extensions. The confidence intervals at low forces are narrower than at large forces, suggesting that the linear regime is predicted with higher certainty.

Comparing the single-level predictions with the prediction of \mathcal{M}_{HB} in Fig. 13(a), we find that the predictions of the single-level models are inferior to those from \mathcal{M}_{HB} . In the latter, all data points are included in the 90% CI, and the uncertainty of the prediction at large F_{ext} is smaller and does not increase with F_{ext} , as in Fig. 15. The HB model therefore gives a better approximation of the data for both extensions and has smaller prediction uncertainty.

2. Inclination angle prediction

When RBCs tank tread, they acquire an inclination angle ϕ [Fig. 3(b)] that varies with the applied shear rate and solvent viscosity [80]. Here we attempt to predict the mean

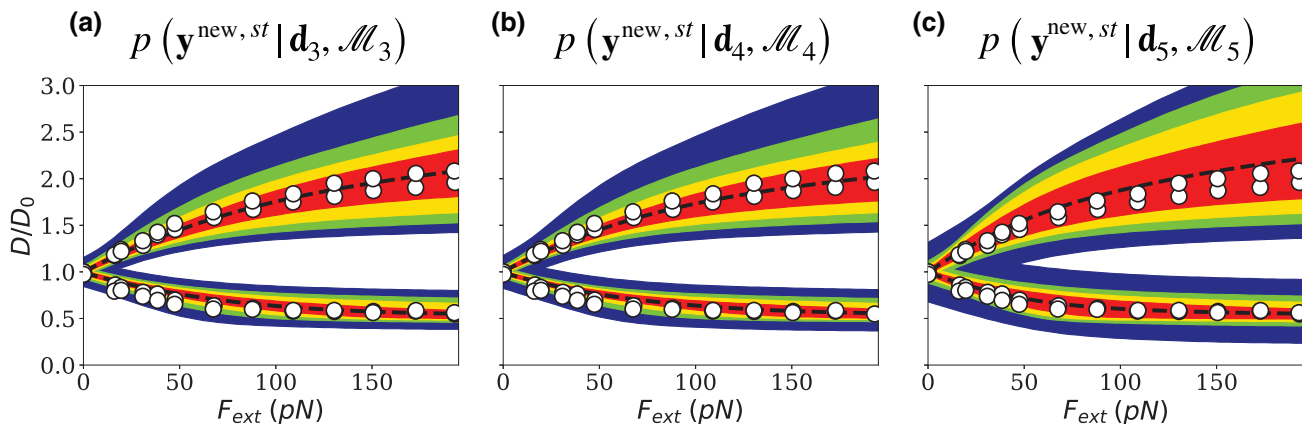


FIG. 15. Propagation of uncertainty from single-level models of TTF to the prediction of RBC extensions in stretching. The black dotted line represents the expected (mean) value. The colored areas denote the 99% (blue), 90% (green), 75% (yellow), and 50% (red) credible intervals. Circles correspond to all experimental data of stretching [17,74].

inclination angle by propagating the parameter distributions of the Bayesian models of TTF and \mathcal{M}_{HB} , using Eq. (4). The predictions are compared to the fitted functions in Fig. S8 of Ref. [80] for $\eta_o = 23.9$ mPas. The propagation is performed using a GP surrogate for ϕ , computed from the same set of simulations used to construct the TTF surrogate.

In Fig. 16(a) we show one of the three propagations from the TTF models, as a representative case, corresponding to $p(\mathbf{y}^{\text{new},\phi} | \mathbf{d}_3, \mathcal{M}_3)$, and in Fig. 16(b) we show the propagation from the hierarchical model $p(\mathbf{y}^{\text{new},\phi} | \bar{\mathbf{d}}, \mathcal{M}_{\text{HB}})$. We observe that the predictions of the Bayesian models capture the general trend of decreasing ϕ with increasing shear rate; however, the mean predictions under predict the experimental data. Additionally, the experimental data are included in the 99% CI of the \mathcal{M}_{HB} , as opposed to \mathcal{M}_3 .

3. Equilibrium shape prediction

We now examine the prediction of the RBC resting shape with respect to the measurements of Evans and Fung [56]. The comparison of the resting shapes is based on the thickness of the membrane H , along the radial direction. The experimental thickness is described by Evans and Fung [56] as

$$H(r) = \left(1 - \frac{r^2}{R^2}\right)^{1/2} \left(C_0 + C_2 \frac{r^2}{R^2} + C_4 \frac{r^4}{R^4}\right), \quad (6)$$

where r is the radial distance from the center, R is the RBC radius, and C_0, C_2, C_4 are constants estimated experimentally [56]. We use the values from Ref. [56] corresponding to the case of a RBC submerged in an isotonic solution. The comparison between the experimental data and the

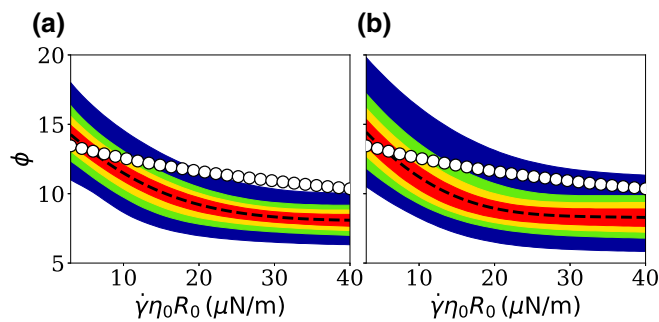


FIG. 16. Propagation of uncertainty for the prediction of the mean inclination angle of a single RBC in shear flow. (a) Prediction from the single-level model of TTF $p(\mathbf{y}^{\text{new},\phi} | \mathbf{d}_4, \mathcal{M}_4)$. (b) Prediction from the hierarchical model $p(\mathbf{y}^{\text{new},\phi} | \bar{\mathbf{d}}, \mathcal{M}_{\text{HB}})$. The black dotted line represents the expected (mean) value. The colored areas denote the 99% (blue), 90% (green), 75% (yellow), and 50% (red) credible intervals. Circles are generated from the fit of experimental data of the inclination angle [80]—see the text for details.

simulations is performed using the average deviation of the simulated RBC thickness with respect to the experimental thickness, as given by Eq. (6). The average thickness deviation is used as an indicator for the goodness of the resting shape prediction.

Given a fixed stress-free shape, we find that the resting shape of the RBC depends only on the FvK number, as seen in Fig. 17. At large FvK ≈ 400 , strain energy dominates over bending, and the resting shape has a closer resemblance (and thus smaller deviation) to the experimental biconcave shape. This is expected, as the stress-free shape used in this study is a discretization of the biconcave shape of Eq. (6), on a triangulated mesh. On the other hand, as the bending modulus increases (FvK number decreases), the dimples of the RBC become shallower, leading to a flattened RBC at FvK ≈ 20 , and thus larger deviations of the thickness from Eq. (6). The average thickness deviation with respect to the FvK number is fitted with a one-dimensional GP (details in Appendix E), and the mean of the GP is shown with a line over the simulation points in Fig. 17.

The results of the uncertainty propagation are shown in Fig. 18. The TTF Bayesian models $\mathcal{M}_3, \mathcal{M}_4, \mathcal{M}_5$ predict the largest average thickness deviation, with EVs at 0.031, 0.032, and 0.035 μm , respectively. The stretching Bayesian models $\mathcal{M}_1, \mathcal{M}_2$ and the HB model \mathcal{M}_{HB} predict lower EVs, at 0.030, 0.020, and 0.026 μm , respectively. An explanation for this outcome is that the stretching data include the resting shape of the RBC, corresponding to the data point where $F_{\text{ext}} = 0$. Therefore, the parameter distributions obtained from the stretching Bayesian models are generated taking into account the resting shape of the RBC. This can also be seen in the respective marginal distributions of κ_b [dashed blue lines in Fig. 14(c)], where higher probability is given to smaller values of κ_b , thus limiting the flattening effects of κ_b on

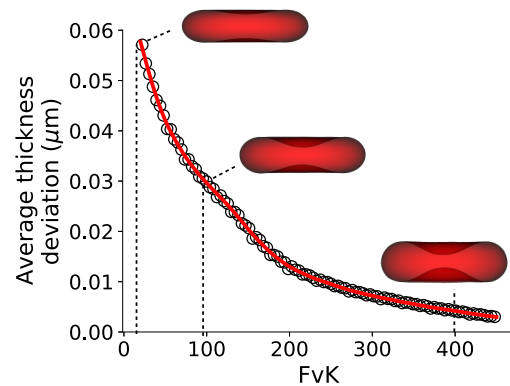


FIG. 17. GP surrogate for the deviation of the RBC thickness at rest from the shape of Evans and Fung [56]. Circles correspond to the simulation results. Line corresponds to the prediction of the GP surrogate.

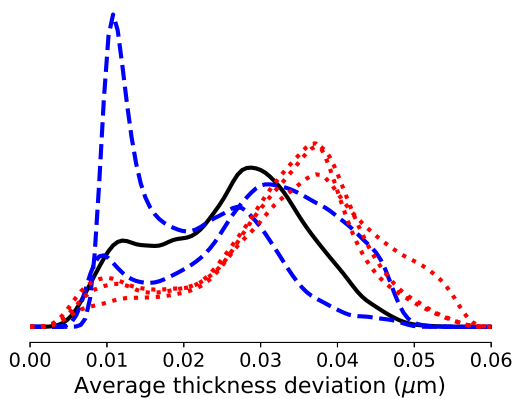


FIG. 18. Predictions of the average thickness deviation. Solid (black) line corresponds to the prediction from the hierarchical model $p(\mathbf{y}^{\text{new,eq}} | \bar{\mathbf{d}}, \mathcal{M}_{\text{HB}})$. Dashed (blue) lines correspond to the predictions from the single-level models of stretching $p(\mathbf{y}^{\text{new,eq}} | \mathbf{d}_i, \mathcal{M}_i)$, $i = 1, 2$. Dotted (red) lines correspond to the predictions from the single-level models of the TTF $p(\mathbf{y}^{\text{new,eq}} | \mathbf{d}_i, \mathcal{M}_i)$, $i = 3, \dots, 5$.

the resting shape. This result also demonstrates that if the stress-free shape is biconcave then the bending modulus of the RBC cannot be realistically estimated, as a “perfect” resting shape would correspond to $\kappa_b = 0$ J, and as such, strengthens the usage of a stress-free shape other than the resting one [53,54].

V. CONCLUSIONS

This study presents a systematic investigation of the incorporation of experimental data in the calibration of the most frequently used computational model of the red blood cell. We justify the need to revise existing practices that rely on calibration of the model parameters without providing cross-experimental validation. We present a systematic Bayesian inference framework using classical and hierarchical models to remedy the RBC model calibration. We consider the calibration stage as an important part of the whole modeling process. We combine multiple data sets originating from different experimental setups using a hierarchical Bayesian model. The hierarchical Bayesian model accommodates information from all data sets in the inference of the model’s parameters. Lastly, we propagate the parameter uncertainties in forward predictions to test the transferability of the model.

We observe that the examined RBC model has enough flexibility to fit experimental data of various types. When employing single-level Bayesian inference, the predictions are accurate and the predictive uncertainties are small (indicating confidence in the prediction), if the predicted data are included in the inference. However, if the propagation is performed on data that are not included in the inference, the predictions are not always an accurate approximation for the data. Hence, in contrast to existing

studies and calibration practices, the present work introduces a hierarchical Bayesian inference to obtain a general probability distribution for the parameters that bridge across multiple heterogeneous experimental data sets.

The traditional approach for the search of optimal RBC model parameters relies on performing a one-at-a-time calibration of each model parameter to a particular set of data. This approach, however, does not assess the predicting capabilities of the model on new or unseen data, nor provides a measure of the uncertainty in the predictions of the computational model. The HB framework provides a remedy to this problem through a rigorous and reproducible procedure. The need for explicit weighting of different experiments when inferring the model parameters is circumvented in the HB framework, as all data sets have an equal contribution to the parameter inference. Another feature of the HB framework is that it is easily extendable in terms of including more data sets. Finally, the unique feature provided by the framework is the prediction of unseen experiments with prediction uncertainty.

In conclusion, our results indicate that the transferability of the computational model strongly depends on the choice and handling of the experimental data. The accuracy of the predictions reflects the relevance of the physical mechanisms governing the experimental data used to infer the model parameters. By comparing single-level and hierarchical Bayesian models, we conclude that the latter give the best compromise of mean prediction and model uncertainty. Moreover, differences between experimental measurements are addressed through the HB framework, as the uncertainty of the parameters in the HB model incorporates the diversity of the experimental data. As such, the current practice of using different parameters to fit a particular set of data is rendered obsolete.

We believe that the results and lessons learned from the present study go beyond the specific RBC model. They reflect the need to properly incorporate experimental data and make them an integral part of a modeling process that can provide predictions with quantified uncertainties.

ACKNOWLEDGMENTS

We would like to thank Xin Bian for invaluable discussions on RBC simulations. We acknowledge support by the European Research Council (ERC Advanced Grant No. 341117) and computational resources: granted by the Swiss National Supercomputing Center (CSCS) under project IDs “ch7” and “928” (used for the RBC simulations) and by the Euler cluster at ETH Zürich (used for the UQ inference).

Author contributions are as follows. Conceptualization: C.P., P.K. Investigation: A.E., G.A., S.L., L.K., C.P., P.K. Data curation: A.E., L.K. RBC simulations: A.E., S.L. UQ inference: A.E., G.A., C.P., P.K. Analysis: A.E., D.A. RBC software: D.A., L.A., S.L. UQ software: G.A., A.E.,

L.K., L.A. RBC/DPD validation: A.E., D.A. Visualization: A.E., G.A. Funding acquisition & resources: P.K. Project administration: P.K. Writing—original draft: A.E. Writing—review and editing: A.E., G.A., L.A., S.L., D.A., C.P., P.K.

APPENDIX A: SIMULATION DETAILS

(Note that in the following we omit the units when giving values in the simulation system to ease text readability.) The basic mass, length, and energy scales in the simulation are set by specifying $\rho^{(s)} = 10$, $R_0^{(s)} = 6$, and $k_B T^{(s)} = 0.1$, where $\rho^{(s)}$ is the mass density of the fluid, $R_0^{(s)}$ the radius of a sphere having the same area as the RBC, and $k_B T^{(s)}$ the thermal energy. Through these, we define the *length*, *mass*, and *time* conversion factors between the physical [superscript “(p)”] and simulation [superscript “(s)”] systems as

$$\begin{aligned} U_L &= R_0^{(p)}/R_0^{(s)} = 5.463 \times 10^{-7} \text{ m}, \\ U_M &= U_L^3 \rho^{(p)}/\rho^{(s)} = 1.630 \times 10^{-17} \text{ kg}, \\ U_T &= (U_L^2 U_M k_B T^{(s)}/k_B T^{(p)})^{1/2} = 1.091 \times 10^{-5} \text{ s}; \end{aligned}$$

assuming that the RBC has an area $A_0^{(p)} = 135 \mu\text{m}^2$ and volume $V_0 = 94 \mu\text{m}^3$ [56], the water density at room temperature is $\rho^{(p)} = 1000 \text{ kg/m}^3$ and $k_B T^{(p)} = 4.087 \times 10^{-21} \text{ J}$. The effective RBC length is $R_0^{(p)} = 3.278 \mu\text{m}$, computed from $R_0 = [A_0/(4\pi)]^{1/2}$. The DPD number density is $n_d = 10$ and the mass of a single DPD particle is $m = 1$, as computed from $\rho = mn_d$. For the RBC membrane, we use a mesh with $N_v = 2562$ vertices. The total membrane mass is $M_m = \rho A_0$, assuming a two-dimensional membrane surface, and the mass of each membrane particle is $m_m = M_m/N_v$.

The DPD dissipative kernel power is set to $k = 0.125$. The relation between the solvent viscosity and the DPD parameters α and γ is estimated in advance based on double Poiseuille flows in a domain of $20 \times 22 \times 20$, with periodic boundary conditions in x and z , and planar walls in y [94]. We find that the empirical relation $\alpha r_c/k_B T = 100$, with $r_c = 1$ the cutoff radius, can be used to simultaneously maintain $\text{Ma} < \text{Re} < 0.1$ and to keep the fluid in the Newtonian regime for the Re and Ca ranges used in this study, similarly to previous studies [95,96]. The speed of sound is estimated through the relation $c^2 = \partial p/\partial \rho$, with the equation of state taken from the numerical results of Groot and Warren [57].

The TTF of the RBCs in the shear flow simulations is estimated in two ways. In the first way the TTF is estimated by taking the average of the Fourier transform of the individual membrane particle positions in time. We assume that the leading frequency of the averaged Fourier transform corresponds to the TTF. In the second way the TTF is estimated using bead tracking by taking the inverse of

the average time required for each membrane particle to complete one full revolution. All simulations satisfy the conditions that the difference between the two TTF computations is less than 5%, and that the TTF has converged over time to a stationary value.

APPENDIX B: LOCAL AREA COMPRESSIBILITY

There seems to be considerable uncertainty regarding the strength of the stretch (local area) modulus k_a . Computational studies from Refs. [26,83,85,88,89] have assumed a weakly compressible membrane. Weak local compressibility can be justified as the phospholipid plasma membrane can accommodate local area changes by a reorganization of the lipid distribution on the membrane [83]. Other analytical [87] and computational studies [33,86,91,97] have assumed both local and global area incompressibility. Yazdani *et al.* [33] demonstrated that increasing the resistance to local area dilatation increases the TTF of tank-treading RBCs. Our preliminary studies for tank-treading RBCs corroborate this result as we have also observed a systematic increase in the TTF by increasing C from 1 to 2000, the latter corresponding to the case $k_a = K_A$. The TTF computed from our simulations within the specified region of dimensionless numbers is not able to capture the lower TTF values of the experimental data for $C = 2000$. We have therefore used $C = 1$ to avoid the introduction of an additional parameter, noting however that further studies are needed to elucidate the effect of local area compressibility.

APPENDIX C: VISCOSITY RATIO IN SHEAR FLOW DATA SETS

There exist multiple experimental studies that provide measurements for the TTF of single RBCs in simple shear flow. However, each TTF data set is reported at a different solvent viscosity η_o . Considering the uncertainty regarding the cytoplasmic viscosity at room temperature $\eta_i \approx 7\text{--}10 \text{ mPa s}$ [26,78,80,98,99] and its variability among RBCs, there is a range of possible λ corresponding to each TTF data set, shown in Fig. 19. We perform simulations at a single $\lambda = 0.32$, which lies within the possible λ range of data sets [75,76,79,80], denoted with red squares in Fig. 19. However, the cell used in Ref. [76] reports a large surface area (see Table II), possibly corresponding to a reticulocyte, and is thus not considered in our study.

APPENDIX D: EXTRACTION OF TTF DATA FROM THE LITERATURE

The sources of the TTF data sets used in the inference, Fig. 2(b), are as follows.

- (i) Data set \mathbf{d}_3 taken from Fig. 4 of Ref. [75] for $\eta_o = 31 \text{ mPa s}$.

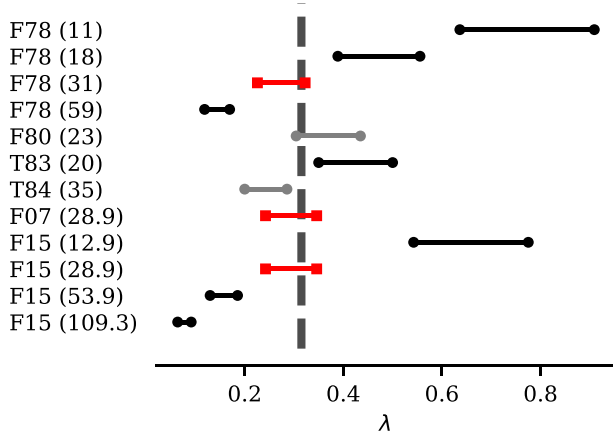


FIG. 19. Possible range for λ in the shear flow experiments, due to the uncertainty in the cytoplasmic viscosity at room temperature ($\eta_i \approx 7\text{--}10$ mPa s). The y axis denotes the data set origin, in the format “AuthorYear(η_o),” with η_o in mPa s. Square (red) markers correspond to data sets considered in the UQ. Gray markers correspond to data sets with large cell surface areas. Dashed line corresponds to simulations, performed at $\lambda = 0.32$.

(ii) Data set \mathbf{d}_4 taken from the mean values* of donor 3 from Refs. [79,80].

(iii) Data set \mathbf{d}_5 taken from the mean values* of donors 1 and 2 from Refs. [79,80].

*Note for \mathbf{d}_4 and \mathbf{d}_5 . The data from Ref. [79] are a subset of the data given in Ref. [80]: the former corresponds to the data of donor 3, while the latter to the collective data of donors 1, 2, 3. We digitize the data of Fig. 4(b) of Ref. [79] and Fig. S11 of Ref. [80] (for $\eta_o = 28.9$ mPa s.) We identify and separate the data of donor 3 from the collective data of donors 1, 2, 3. Two independent data sets are generated: (i) donor 3, and (ii) donors 1 and 2. As all other data sets used in our inference correspond to mean values of the data, we compute the local mean values of the TTF and the shear rate, as plotted in Fig. 2(b).

APPENDIX E: SURROGATE MODELS

1. Mathematical background

We denote by $\mathbf{t}_M = (t_1, \dots, t_M)$ the vector of M observed outputs of the computational model F with input $\mathbf{x}^{(i)}$ and parameters $\vartheta_c^{(i)}$, $i = 1, \dots, M$. We denote by z the GP. Assuming that the observations can be represented by the GP and an additive error term, we write

$$t_i = z(\boldsymbol{\zeta}_i) + \epsilon_i, \quad i = 1, \dots, M, \quad (\text{E1})$$

where $\boldsymbol{\zeta}_i = \{\mathbf{x}^{(i)}, \vartheta_c^{(i)}\}$ and the ϵ_i are independent random variables following $\mathcal{N}(0, \beta^{-1})$.

Let $\mathbf{D}_{M+1} = \{t_1, \dots, t_M, \boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_{M+1}\}$. The prediction t_{M+1} of the GP model for a new input $\boldsymbol{\zeta}_{M+1}$ given the

TABLE IV. Construction of GP surrogates. Here N_{sim} denotes the total number of simulations performed for each setup. The “inputs” denote the quantities varied in each setup.

Setup	N_{sim}	Inputs	Output(s)
Equilibration	100	(x_0, FvK)	E_{th}
Stretching	300	(x_0, FvK, F^*)	$D_{\text{ax}}^*, D_{\text{tr}}^*$
Shear flow	200	$(x_0, \text{FvK}, \eta_m^*, \text{Ca})$	ν^*, ϕ

data \mathbf{D}_{M+1} is a random variable that follows the conditional distribution

$$p(t_{M+1} | \mathbf{D}_{M+1}) = \mathcal{N}[t_{M+1} | m_z(\mathbf{D}_{M+1}), \sigma_z^2(\mathbf{D}_{M+1})], \quad (\text{E2})$$

where

$$\begin{aligned} m_z(\mathbf{D}_{M+1}) &= \mathbf{k}_{M+1}^\top \mathbf{C}_M^{-1} \mathbf{t}_M, \\ \sigma_z^2(\mathbf{D}_{M+1}) &= c_{M+1} - \mathbf{k}_{M+1}^\top \mathbf{C}_M^{-1} \mathbf{k}_{M+1}, \end{aligned} \quad (\text{E3})$$

where the i th element of \mathbf{k}_{M+1} is given by $k(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_{M+1})$ for $i = 1, \dots, M$ and k is the kernel function. The kernel is a user-defined function that reflects the property that points which are close in the input space are expected to have more strongly correlated outputs. The scalar c_{M+1} is defined as $c_{M+1} = k(\boldsymbol{\zeta}_{M+1}, \boldsymbol{\zeta}_{M+1}) + \beta^{-1}$. The elements of the covariance matrix \mathbf{C}_M are given by $(\mathbf{C}_M)_{ij} = k(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j) + \beta^{-1} \delta_{ij}$ for $i, j = 1, \dots, M$, where δ_{ij} is the Kronecker delta function. Usually, the kernel function depends on a set of hyperparameters that define its properties. The values of the hyperparameters, as well as the value for β , are learned through an optimization process such that the likelihood of \mathbf{t}_M is maximized [66].

2. Construction of GPs

The GPs are built from a set of training points, corresponding to a set of simulations with preselected parameters chosen to form a LH. The LHs are constructed with the `MaxPro` design criterion [81]. The number of simulations in each LH along with their respective inputs and outputs are listed in Table IV.

To construct a GP, we need to first specify the form of the covariance matrix. We use a covariance matrix composed of the sum of two functions. The first is a function modeling white noise, while the second is chosen to be one of the following functions: squared exponential, Matern with $\nu = 5/2$, and Matern with $\nu = 3/2$. For the GP of each quantity of interest (equilibrium thickness deviation, axial and transverse extensions, TTF, and inclination angle), we choose the function yielding the smallest training and prediction errors (see the next paragraph). Specifically, we choose the Matern function with $\nu = 5/2$ for the GPs of the axial and transverse extensions, TTF, and inclination angle. In turn, for the equilibrium thickness, we

opt to use the Matern function with $\nu = 3/2$. The GPs are constructed using the open-source library `libgp` [100].

The training of the GPs is performed using all LHs points. After the training is completed, we compute the average training error E_t , as the L^1 norm between the GP prediction and the simulation output, averaged from all LHs points. For the chosen covariance functions, $E_t = 0.04\% - 1.61\%$.

In addition, we perform a cross-validation study where 90% of the simulation data are used to train the GPs (training data set), and the remaining 10% is used to evaluate the accuracy of the GP predictions (testing data set). This type of testing is used to examine whether the GP is over fitting the simulation output. The average prediction error E_p is computed similarly to E_t , but now using only the GP predictions and simulation outputs of the testing data set. The process is repeated 100 times and statistics are collected, yielding $E_p = 0.28\% - 3.47\%$.

APPENDIX F: HIERARCHICAL BAYESIAN INFERENCE

1. Mathematical background

In the hierarchical Bayesian framework, the data are grouped into separate data sets, reflecting different experimental types, conditions, or laboratory of origin. We denote the collection of data sets as $\vec{\mathbf{d}} = \{\mathbf{d}_1, \dots, \mathbf{d}_{N_d}\}$. Each $\mathbf{d}_i = (d_{i,1}, \dots, d_{i,N_i}) \in \mathbb{R}^{N_i}$ is the data vector corresponding to the data set i , and $d_{i,j}$ corresponds to input conditions $\mathbf{x}_{i,j}$ for $i = 1, \dots, N_d$ and $j = 1, \dots, N_i$. The observable variable of the model that corresponds to data \mathbf{d}_i is denoted by \mathbf{y}_i and the collection of all observables as $\vec{\mathbf{y}}$.

Our goal is to obtain samples from the posterior distribution, $p(\vartheta_i | \vec{\mathbf{y}}, \mathcal{M}_{\text{HB}})$, where $\vec{\mathbf{d}} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$,

$$p(\vartheta_i | \vec{\mathbf{y}}, \mathcal{M}_{\text{HB}}) = \int p(\vartheta_i | \boldsymbol{\psi}, \vec{\mathbf{y}}, \mathcal{M}_{\text{HB}}) p(\boldsymbol{\psi} | \vec{\mathbf{y}}, \mathcal{M}_{\text{HB}}) d\boldsymbol{\psi}. \quad (\text{F1})$$

The dependency assumptions from Fig. 1 allow us to simplify $p(\vartheta_i | \boldsymbol{\psi}, \vec{\mathbf{y}}, \mathcal{M}) = p(\vartheta_i | \boldsymbol{\psi}, \mathbf{y}_i, \mathcal{M})$, and Eq. (F1) can be rewritten using Bayes' theorem as

$$p(\vartheta_i | \vec{\mathbf{y}}, \mathcal{M}) = \int \frac{p(\mathbf{y}_i | \vartheta_i, \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}) p(\vartheta_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}})}{p(\mathbf{y}_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}})} \times p(\boldsymbol{\psi} | \vec{\mathbf{y}}, \mathcal{M}_{\text{HB}}) d\boldsymbol{\psi}. \quad (\text{F2})$$

Since $p(\mathbf{y}_i | \vartheta_i, \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}) = p(\mathbf{y}_i | \vartheta_i, \mathcal{M}_{\text{HB}})$, Eq. (F2) simplifies to

$$p(\vartheta_i | \vec{\mathbf{y}}, \mathcal{M}_{\text{HB}}) = p(\mathbf{y}_i | \vartheta_i, \mathcal{M}_{\text{HB}}) \times \int \frac{p(\vartheta_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}})}{p(\mathbf{y}_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}})} p(\boldsymbol{\psi} | \vec{\mathbf{y}}, \mathcal{M}_{\text{HB}}) d\boldsymbol{\psi}. \quad (\text{F3})$$

Finally, the posterior distribution in Eq. (F1) can be approximated as

$$p(\vartheta_i | \vec{\mathbf{y}}, \mathcal{M}_{\text{HB}}) \approx \frac{p(\mathbf{y}_i | \vartheta_i, \mathcal{M}_{\text{HB}})}{N_s} \sum_{k=1}^{N_s} \frac{p(\vartheta_i | \boldsymbol{\psi}^{(k)}, \mathcal{M}_{\text{HB}})}{p(\mathbf{y}_i | \boldsymbol{\psi}^{(k)}, \mathcal{M}_{\text{HB}})}, \quad (\text{F4})$$

where $\boldsymbol{\psi}^{(k)} \sim p(\boldsymbol{\psi} | \vec{\mathbf{y}}, \mathcal{M}_{\text{HB}})$ for $k = 1, \dots, N_s$ and N_s is sufficiently large. Thus, in order to obtain ϑ_i samples, we first have to sample the probability distribution $p(\boldsymbol{\psi} | \vec{\mathbf{y}}, \mathcal{M}_{\text{HB}})$, which, according to Bayes' theorem, is equal to

$$p(\boldsymbol{\psi} | \vec{\mathbf{y}}, \mathcal{M}) = \frac{p(\vec{\mathbf{y}} | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}) p(\boldsymbol{\psi} | \mathcal{M}_{\text{HB}})}{p(\vec{\mathbf{y}} | \mathcal{M}_{\text{HB}})}, \quad (\text{F5})$$

where $p(\boldsymbol{\psi} | \mathcal{M}_{\text{HB}})$ is the prior probability density function on $\boldsymbol{\psi}$ and $p(\vec{\mathbf{y}} | \mathcal{M}_{\text{HB}})$ is the normalizing constant. Exploiting the dependency assumption of Fig. 1 (main text) we see that

$$p(\vec{\mathbf{y}} | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}) = \prod_{i=1}^N p(\mathbf{y}_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}), \quad (\text{F6})$$

and the likelihood of the i th data set can be expressed according to the total probability theorem as

$$p(\mathbf{y}_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}) = \int p(\mathbf{y}_i | \vartheta_i, \mathcal{M}_{\text{HB}}) p(\vartheta_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}) d\vartheta_i. \quad (\text{F7})$$

TABLE V. Prior distributions in the HB model. The elements of the hyperparameter vector $\boldsymbol{\psi}$ are $\psi_{x_0,1}$, $\psi_{x_0,2}$, $\psi_{\mu_0,1}$, $\psi_{\mu_0,2}$, $\psi_{\kappa_b,1}$, $\psi_{\kappa_b,2}$, $\psi_{\eta_m^*,1}$, $\psi_{\eta_m^*,2}$, $\psi_{\sigma_{\text{sh},1}}$, $\psi_{\sigma_{\text{sh},2}}$, $\psi_{\sigma_{\text{ax},1}}$, $\psi_{\sigma_{\text{ax},2}}$, $\psi_{\sigma_{\text{tr},1}}$, $\psi_{\sigma_{\text{tr},2}}$. Values for μ_0 , $\psi_{\mu_0,1}$, $\psi_{\mu_0,2}$ are expressed in $\mu\text{N/m}$, and for κ_b , $\psi_{\kappa_b,1}$, $\psi_{\kappa_b,2}$ in 10^{-19} J .

Priors for ϑ	Priors for $\boldsymbol{\psi}$
$p(x_0 \boldsymbol{\psi}) = \mathcal{U}(\psi_{x_0,1}, \psi_{x_0,1} + \psi_{x_0,2})$	$\psi_{x_0,1} \sim \mathcal{U}(0.2, 0.7)$ $\psi_{x_0,2} \sim \mathcal{U}(0, 0.5)$
$p(\mu_0 \boldsymbol{\psi}) = \mathcal{U}(\psi_{\mu_0,1}, \psi_{\mu_0,1} + \psi_{\mu_0,2})$	$\psi_{\mu_0,1} \sim \mathcal{U}(2, 6)$ $\psi_{\mu_0,2} \sim \mathcal{U}(0, 4)$
$p(\kappa_b \boldsymbol{\psi}) = \mathcal{U}(\psi_{\kappa_b,1}, \psi_{\kappa_b,1} + \psi_{\kappa_b,2})$	$\psi_{\kappa_b,1} \sim \mathcal{U}(1.5, 10)$ $\psi_{\kappa_b,2} \sim \mathcal{U}(0, 8.5)$
$p(\eta_m^* \boldsymbol{\psi}) = \mathcal{U}(\psi_{\eta_m^*,1}, \psi_{\eta_m^*,1} + \psi_{\eta_m^*,2})$	$\psi_{\eta_m^*,1} \sim \mathcal{U}(0, 20)$ $\psi_{\eta_m^*,2} \sim \mathcal{U}(0, 20)$
$p(\sigma_{\text{sh}} \boldsymbol{\psi}) = \mathcal{U}(\psi_{\sigma_{\text{sh},1}}, \psi_{\sigma_{\text{sh},1}} + \psi_{\sigma_{\text{sh},2}})$	$\psi_{\sigma_{\text{sh},1}} \sim \mathcal{U}(0, 0.1)$ $\psi_{\sigma_{\text{sh},2}} \sim \mathcal{U}(0, 0.1)$
$p(\sigma_{\text{ax}} \boldsymbol{\psi}) = \mathcal{U}(\psi_{\sigma_{\text{ax},1}}, \psi_{\sigma_{\text{ax},1}} + \psi_{\sigma_{\text{ax},2}})$	$\psi_{\sigma_{\text{ax},1}} \sim \mathcal{U}(0, 0.5)$ $\psi_{\sigma_{\text{ax},2}} \sim \mathcal{U}(0, 0.5)$
$p(\sigma_{\text{tr}} \boldsymbol{\psi}) = \mathcal{U}(\psi_{\sigma_{\text{tr},1}}, \psi_{\sigma_{\text{tr},1}} + \psi_{\sigma_{\text{tr},2}})$	$\psi_{\sigma_{\text{tr},1}} \sim \mathcal{U}(0, 0.5)$ $\psi_{\sigma_{\text{tr},2}} \sim \mathcal{U}(0, 0.5)$

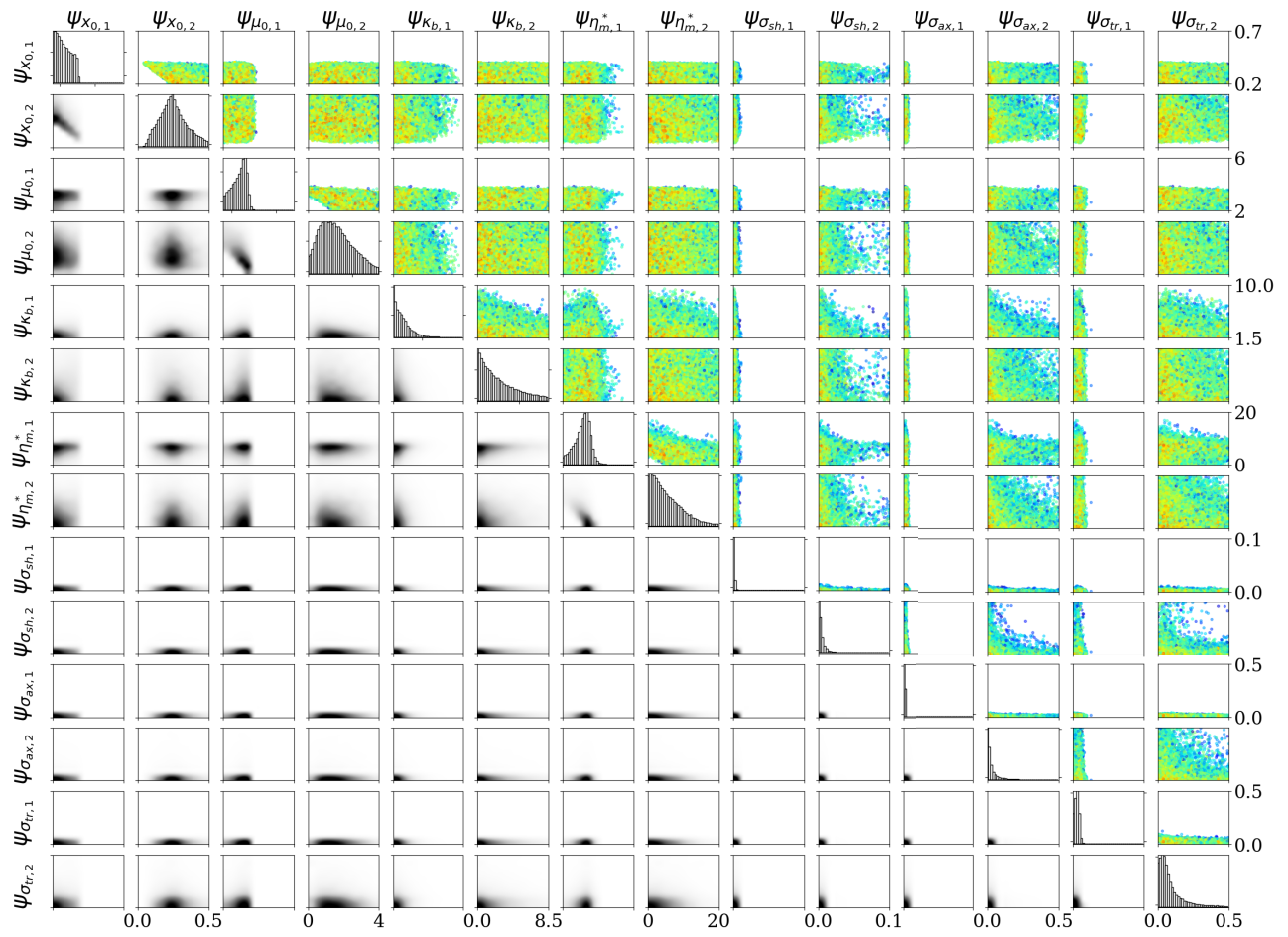


FIG. 20. Posterior distribution for the hyperparameters of the hierarchical model, $p(\boldsymbol{\psi} | \bar{\mathbf{d}}, \mathcal{M}_{\text{HB}})$. Marginal distribution with respect to each parameter (*diagonal*). Two-dimensional projections of the posterior colored by the probability density [black (white) denotes high (low) density] (*below diagonal*). Samples from the posterior distribution coloured by the log likelihood [red (blue) denotes high (low) log likelihood] (*above diagonal*). For tiles in the columns of $\boldsymbol{\psi}_{\bullet,2}$, the x -axis range is given at the bottom of the figure. For tiles in the rows of $\boldsymbol{\psi}_{\bullet,1}$, the y -axis range is given at the right of the figure.

Here we introduce the model \mathcal{M}_i that corresponds to the model describing only the i th data set; see Fig. 1. The posterior distribution of this model will be used as an instrumental density for important sampling. Under the modeling assumption $p(\mathbf{y}_i | \boldsymbol{\vartheta}_i, \mathcal{M}_{\text{HB}}) = p(\mathbf{y}_i | \boldsymbol{\vartheta}_i, \mathcal{M}_i)$ (see Ref. [49]) and using Bayes' theorem, Eq. (F7) is written as

$$p(\mathbf{y}_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}) = \int \frac{p(\boldsymbol{\vartheta}_i | \mathbf{y}_i, \mathcal{M}_i) p(\mathbf{y}_i | \mathcal{M}_i)}{p(\boldsymbol{\vartheta}_i | \mathcal{M}_i)} p(\boldsymbol{\vartheta}_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}) d\boldsymbol{\vartheta}_i, \quad (\text{F8})$$

or, equivalently, as

$$p(\mathbf{y}_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}) = p(\mathbf{y}_i | \mathcal{M}_i) \int \frac{p(\boldsymbol{\vartheta}_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}})}{p(\boldsymbol{\vartheta}_i | \mathcal{M}_i)} p(\boldsymbol{\vartheta}_i | \mathbf{y}_i, \mathcal{M}_i) d\boldsymbol{\vartheta}_i. \quad (\text{F9})$$

Finally, Eq. (F7) can be approximated as

$$p(\mathbf{y}_i | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}}) \approx \frac{p(\mathbf{y}_i | \mathcal{M}_i)}{N_s} \sum_{k=1}^{N_s} \frac{p(\boldsymbol{\vartheta}_i^{(k)} | \boldsymbol{\psi}, \mathcal{M}_{\text{HB}})}{p(\boldsymbol{\vartheta}_i^{(k)} | \mathcal{M}_i)}, \quad (\text{F10})$$

where $\boldsymbol{\vartheta}_i^{(k)} \sim p(\boldsymbol{\vartheta}_i | \mathbf{y}_i, \mathcal{M}_i)$ for $k = 1, \dots, N_s$ and N_s is sufficiently large. Note that, in general, N_s can be different for each data set \mathbf{y}_i . The advantage of this approach is that the likelihoods $p(\mathbf{y}_i | \boldsymbol{\vartheta}_i, \mathcal{M}_i)$, $i = 1, \dots, N$, which are the most expensive part of the computations, are not reevaluated for each $\boldsymbol{\psi}$.

2. Inference setup

The sampling of all obtained distributions is performed using the sampling algorithm BASIS [64], through the *korali* software [47]. In all inferences we used 50 000

samples per generation. In the single-level Bayesian models of stretching, the inference is run with burn-in $BI = 4$ threshold for the coefficient of variation $COV_{\max} = 0.3$ and scaling $\beta^2 = 0.03$. In the single-level Bayesian models of TTF, the inference is run with $BI = 10$, $COV_{\max} = 0.05$, and β^2 equal to 0.02 and 0.01 for the respective data sets \mathbf{d}_3 , \mathbf{d}_4 , and \mathbf{d}_5 . The inference for the hyperparameters of the hierarchical model is run with $BI = 10$, $COV_{\max} = 0.3$, and $\beta^2 = 0.03$. The prior $p(\boldsymbol{\vartheta} | \boldsymbol{\psi})$ for each of the elements in $\boldsymbol{\vartheta}$ is assumed to follow a uniform distribution with lower and upper bounds as listed in Table V. The hyperparameters $\boldsymbol{\psi}$ are assumed to follow a uniform prior distribution as given in Table V. The inference of the new, general set of parameters obtained from the HB model, $\boldsymbol{\vartheta}^{\text{new}}$, is run with $BI = 10$, $COV_{\max} = 0.3$, and $\beta^2 = 0.04$.

-
- [1] J. B. Freund, Numerical simulation of flowing blood cells, *Annu. Rev. Fluid Mech.* **46**, 67 (2014).
- [2] H. W. G. Lim, M. Wortis, and R. Mukhopadhyay, Red blood cell shapes and shape transformations: Newtonian mechanics of a composite membrane, *Soft Matter* **4**, 83 (2008).
- [3] C. Minetti, V. Audemar, T. Podgorski, and G. Coupier, Dynamics of a large population of red blood cells under shear flow, *J. Fluid Mech.* **864**, 408 (2019).
- [4] I. T. Li, T. Ha, and Y. R. Chemla, Mapping cell surface adhesion by rotation tracking and adhesion footprinting, *Sci. Rep.* **7**, 1 (2017).
- [5] A. Kihm, L. Kaestner, C. Wagner, and S. Quint, Classification of red blood cell shapes in flow using outlier tolerant machine learning, *PLoS Comput. Biol.* **14**, 1 (2018).
- [6] G. Kabacaoğlu and G. Biros, Optimal design of deterministic lateral displacement device for viscosity-contrast-based cell sorting, *Phys. Rev. Fluids* **3**, 124201 (2018).
- [7] P. Crosetto, P. Reymond, S. Deparis, D. Kontaxakis, N. Stergiopoulos, and A. Quarteroni, Fluid-structure interaction simulation of aortic blood flow, *Comput. Fluids* **43**, 46 (2011).
- [8] A. Randles, E. W. Draeger, T. Ooppelstrup, L. Krauss, and J. A. Gunnels, in *SC '15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, Austin, TX, USA* (2015), p. 1.
- [9] S. K. Veerapaneni, A. Rahimian, G. Biros, and D. Zorin, A fast algorithm for simulating vesicle flows in three dimensions, *J. Comput. Phys.* **230**, 5610 (2011).
- [10] G. Kabacaoğlu, B. Quaipe, and G. Biros, Low-resolution simulations of vesicle suspensions in 2D, *J. Comput. Phys.* **357**, 43 (2018).
- [11] H.-Y. Chang, X. Li, H. Li, and G. E. Karniadakis, Md/dpd multiscale framework for predicting morphology and stresses of red blood cells in health and disease, *PLoS Comput. Biol.* **12**, e1005173 (2016).
- [12] D. Rossinelli, G. Karniadakis, M. Fatica, I. Pivkin, P. Koumoutsakos, Y. Tang, K. Lykov, D. Alexeev, M. Bernaschi, P. Hadjidoukas, M. Bisson, W. Joubert, and C. Conti, in *SC '15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, Austin, TX, USA* (2015), p. 1.
- [13] S. K. Boey, D. H. Boal, and D. E. Discher, Simulations of the erythrocyte cytoskeleton at large deformation. I. Microscopic models, *Biophys. J.* **75**, 1573 (1998).
- [14] E. A. Evans and R. Skalak, *Mechanics and Thermodynamics of Biomembranes* (CRC Press, Boca Raton, Florida, 1980).
- [15] J. Li, M. Dao, C. T. Lim, and S. Suresh, Spectrin-level modeling of the cytoskeleton and optical tweezers stretching of the erythrocyte, *Biophys. J.* **88**, 3707 (2005).
- [16] M. Dao, J. Li, and S. Suresh, Molecularly based analysis of deformation of spectrin network and human erythrocyte, *Mater. Sci. Eng. C* **26**, 1232 (2006).
- [17] J. Mills, L. Qie, M. Dao, C. Lim, and S. Suresh, Nonlinear elastic and viscoelastic deformation of the human red blood cell with optical tweezers, *Mol. Cell. Biomech.* **1**, 169 (2004).
- [18] I. V. Pivkin and G. E. Karniadakis, Accurate Coarse-Grained Modeling of red Blood Cells, *Phys. Rev. Lett.* **101**, 118105 (2008).
- [19] D. A. Fedosov, B. Caswell, and G. E. Karniadakis, Systematic coarse-graining of spectrin-level red blood cell models, *Comput. Methods Appl. Mech. Eng.* **199**, 1937 (2010).
- [20] D. A. Fedosov, B. Caswell, and G. E. Karniadakis, A multiscale red blood cell model with accurate mechanics, rheology, and dynamics, *Biophys. J.* **98**, 2215 (2010).
- [21] D. J. Quinn, I. Pivkin, S. Y. Wong, K. H. Chiam, M. Dao, G. E. Karniadakis, and S. Suresh, Combined simulation and experimental study of large deformation of red blood cells in microfluidic systems, *Ann. Biomed. Eng.* **39**, 1041 (2011).
- [22] D. A. Fedosov, M. Peltomäki, and G. Gompper, Deformation and dynamics of red blood cells in flow through cylindrical microchannels, *Soft Matter* **10**, 4258 (2014).
- [23] H. Turlier, D. A. Fedosov, B. Audoly, T. Auth, N. S. Gov, C. Sykes, J. F. Joanny, G. Gompper, and T. Betz, Equilibrium physics breakdown reveals the active nature of red blood cell flickering, *Nat. Phys.* **12**, 513 (2016).
- [24] E. Henry, S. H. Holm, Z. Zhang, J. P. Beech, J. O. Tegenfeldt, D. A. Fedosov, and G. Gompper, Sorting cells by their dynamical properties, *Sci. Rep.* **6**, 34375 (2016).
- [25] L. Lanotte, J. Mauer, S. Mendez, D. A. Fedosov, J. M. Fromental, V. Claveria, F. Nicoud, G. Gompper, and M. Abkarian, Red cells' dynamic morphologies govern blood shear thinning under microcirculatory flow conditions, *PNAS* **113**, 13289 (2016).
- [26] J. Mauer, S. Mendez, L. Lanotte, F. Nicoud, M. Abkarian, G. Gompper, and D. A. Fedosov, Flow-Induced Transitions of Red Blood Cell Shapes under Shear, *Phys. Rev. Lett.* **121**, 118103 (2018).
- [27] D. A. Fedosov, B. Caswell, A. S. Popel, and G. E. Karniadakis, Blood flow and cell-free layer in microvessels, *Microcirculation* **17**, 615 (2010).
- [28] D. A. Fedosov, B. Caswell, S. Suresh, and G. E. Karniadakis, Quantifying the biophysical characteristics of Plasmodium-falciparum-parasitized red blood cells in microcirculation, *PNAS* **108**, 35 (2011).

- [29] D. A. Fedosov, W. Pan, B. Caswell, G. Gompper, and G. E. Karniadakis, Predicting human blood viscosity in silico, *PNAS* **108**, 7 (2011).
- [30] A. Yazdani and G. E. Karniadakis, Sub-cellular modeling of platelet transport in blood flow through microchannels with constriction, *Soft Matter* **12**, 4339 (2016).
- [31] P. Dimitrakopoulos, Analysis of the variation in the determination of the shear modulus of the erythrocyte membrane: Effects of the constitutive law and membrane modeling, *Phys. Rev. E* **85**, 1 (2012).
- [32] K. Sinha and M. D. Graham, Dynamics of a single red blood cell in simple shear flow, *Phys. Rev. E* **92**, 1 (2015).
- [33] A. Z. K. Yazdani, R. M. Kalluri, and P. Bagchi, Tank-treading and tumbling frequencies of capsules and red blood cells, *Phys. Rev. E* **83**, 046305 (2011).
- [34] J. Sigüenza, S. Mendez, and F. Nicoud, How should the optical tweezers experiment be used to characterize the red blood cell membrane mechanics?, *Biomech. Model. Mechanobiol.* **16**, 1645 (2017).
- [35] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis* (CRC Press, Boca Raton, Florida, 2013).
- [36] P. D. Congdon, *Applied Bayesian Hierarchical Methods* (CRC Press, Boca Raton, Florida, 2010).
- [37] M. C. Kennedy and A. O'Hagan, Bayesian calibration of computer models, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63**, 425 (2001).
- [38] J. T. Oden, Adaptive multiscale predictive modelling, *Acta Numer.* **27**, 353 (2018).
- [39] P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework, *J. Chem. Phys.* **137**, 144103 (2012).
- [40] K. Farrell, J. T. Oden, and D. Faghihi, A Bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems, *J. Comput. Phys.* **295**, 189 (2015).
- [41] J. T. Oden, K. Farrell, and D. Faghihi, Estimation of error in observables of coarse-grained models of atomic systems, *Adv. Model. Simul. Eng. Sci.* **2**, 1 (2015).
- [42] J. Zavadlav, G. Arampatzis, and P. Koumoutsakos, Bayesian selection for coarse-grained models of liquid water, *Sci. Rep.* **9**, 1 (2019).
- [43] F. Cailliez, P. Pernot, F. Rizzi, R. Jones, O. Knio, G. Arampatzis, and P. Koumoutsakos, in *Uncertainty Quantification in Multiscale Materials Modeling* (Woodhead Publishing, 2020), p. 169.
- [44] J. M. Pereira, J. P. Serra e Moura, A. R. Ervilha, and J. C. Pereira, On the uncertainty quantification of blood flow viscosity models, *Chem. Eng. Sci.* **101**, 253 (2013).
- [45] S. Sankaran, H. J. Kim, G. Choi, and C. A. Taylor, Uncertainty quantification in coronary blood flow simulations: Impact of geometry, boundary conditions and blood viscosity, *J. Biomech.* **49**, 2540 (2016).
- [46] I. G. Johnston, B. C. Rickett, and N. S. Jones, Explicit tracking of uncertainty increases the power of quantitative rule-of-thumb reasoning in cell biology, *Biophys. J.* **107**, 2612 (2015).
- [47] S. M. Martin, D. Wälchli, G. Arampatzis, and P. Koumoutsakos, Korali: A high-performance computing framework for stochastic optimization and Bayesian uncertainty quantification, [arXiv:2005.13457](https://arxiv.org/abs/2005.13457).
- [48] K. Dalbey, M. S. Eldred, G. Geraci, J. D. Jakeman, K. A. Maupin, J. A. Monschke, D. T. Seidl, L. P. Swiler, A. Tran, F. Menhorn, and X. Zeng, Dakota A Multilevel Parallel Object-Oriented Framework for Design Optimization Parameter Estimation Uncertainty Quantification and Sensitivity Analysis: Version 6.12 Theory Manual, Tech. Rep. (Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2020).
- [49] S. Wu, P. Angelikopoulos, G. Tauriello, C. Papadimitriou, and P. Koumoutsakos, Fusing heterogeneous data for the calibration of molecular dynamics force fields using hierarchical Bayesian models, *J. Chem. Phys.* **145**, 244112 (2016).
- [50] J. Weber, Fluctuation dissipation theorem, *Phys. Rev.* **101**, 1620 (1956).
- [51] T. M. Fischer, Shape memory of human red blood cells, *Biophys. J.* **86**, 3304 (2004).
- [52] M. Levant and V. Steinberg, Intermediate regime and a phase diagram of red blood cell dynamics in a linear flow, *Phys. Rev. E* **94**, 1 (2016).
- [53] D. Cordasco, A. Yazdani, and P. Bagchi, Comparison of erythrocyte dynamics in shear flow under different stress-free configurations, *Phys. Fluids* **26**, 041902 (2014).
- [54] Z. Peng, A. Mashayekh, and Q. Zhu, Erythrocyte responses in low-shear-rate flows: Effects of non-biconcave stress-free state in the cytoskeleton, *J. Fluid Mech.* **742**, 96 (2014).
- [55] M. Hoore, F. Yaya, T. Podgorski, C. Wagner, G. Gompper, and D. A. Fedosov, Effect of spectrin network elasticity on the shapes of erythrocyte doublets, *Soft Matter* **14**, 6278 (2018).
- [56] E. Evans and Y.-C. Fung, Improved measurements of the erythrocyte geometry, *Microvasc. Res.* **4**, 335 (1972).
- [57] R. D. Groot and P. B. Warren, Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation, *J. Chem. Phys.* **107**, 4423 (1997).
- [58] P. Espanol, Hydrodynamics from dissipative particle dynamics, *Phys. Rev. E* **52**, 1734 (1995).
- [59] J. P. Morris, P. J. Fox, and Y. Zhu, Modeling low Reynolds number incompressible flows using SPH, *J. Comput. Phys.* **136**, 214 (1997).
- [60] M. Revenga, I. Zuniga, and P. Espanol, Boundary conditions in dissipative particle dynamics, *Comput. Phys. Commun.* **121-122**, 309 (1999).
- [61] D. Alexeev, L. Amoudruz, S. Litvinov, and P. Koumoutsakos, Mirheo: High-performance mesoscale simulations for microfluidics, *Comput. Phys. Commun.* **254**, 107298 (2020).
- [62] A random variable ε following a normal distribution with mean μ and variance σ^2 is denoted as $\varepsilon \sim \mathcal{N}(\mu, \sigma^2)$. The corresponding probability density function evaluated at $z \in \mathbb{R}$ is denoted as $\mathcal{N}(z | \mu, \sigma^2)$.
- [63] S. Wu, P. Angelikopoulos, J. L. Beck, and P. Koumoutsakos, Hierarchical stochastic model in Bayesian inference for engineering applications: Theoretical implications and efficient approximation, *ASCE-ASME J. Risk Uncertain. Eng. Syst. B Mech. Eng.* **5**, 011006 (2019).

- [64] S. Wu, P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, Bayesian annealed sequential importance sampling: An unbiased version of transitional Markov chain Monte Carlo, *ASCE-ASME J. Risk Uncertain. Eng. Syst. B Mech. Eng.* **4**, 011008 (2018).
- [65] J. Ching and Y. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, *J. Eng. Mech.* **133**, 816 (2007).
- [66] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, Singapore, 2016).
- [67] J. E. Oakley and A. O'Hagan, Probabilistic sensitivity analysis of complex models: A Bayesian approach, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66**, 751 (2004).
- [68] L. S. Bastos and A. O'hagan, Diagnostics for Gaussian process emulators, *Technometrics* **51**, 425 (2009).
- [69] A. Marrel, B. Iooss, B. Laurent, and O. Roustant, Calculations of Sobol indices for the Gaussian process meta-model, *Reliab. Eng. Syst. Saf.* **94**, 742 (2009).
- [70] E. T. Chang, M. Strong, and R. H. Clayton, Bayesian sensitivity analysis of a cardiac cell model using a Gaussian process emulator, *PLoS ONE* **10**, 1 (2015).
- [71] A. Melis, R. H. Clayton, and A. Marzo, Bayesian sensitivity analysis of a 1D vascular model with Gaussian process emulators, *Int. J. Numer. Method. Biomed. Eng.* **33**, 1 (2017).
- [72] R. Kohavi, in *IJCAI'95: Proceedings of the 14th International Joint Conference on Artificial Intelligence* (Montreal, Canada, 1995), Vol. 14, p. 1137.
- [73] M. Kuhn and K. Johnson, *Applied Predictive Modeling* (Springer, New York, 2013), Vol. 26.
- [74] S. Suresh, J. Spatz, J. P. Mills, A. Micoulet, M. Dao, C. Lim, M. Beil, and T. Seufferlein, Connections between single-cell biomechanics and human disease states: Gastrointestinal cancer and malaria, *Acta Biomater.* **1**, 15 (2005).
- [75] T. M. Fischer, M. Stohr-Lissen, and H. Schmid-Schönbein, The red cell as a fluid droplet: Tank tread-like motion of the human erythrocyte membrane in shear flow, *Science* **202**, 894 (1978).
- [76] T. M. Fischer, On the energy dissipation in a tank-treading human red blood cell, *Biophys. J.* **32**, 863 (1980).
- [77] R. Tran-Son-Tay, Ph.D. thesis, Washington University, St. Louis, MO, 1983.
- [78] R. Tran-Son-Tay, S. P. Sutera, and P. R. Rao, Determination of red blood cell membrane viscosity from rheoscopic observations of tank-treading motion, *Biophys. J.* **46**, 65 (1984).
- [79] T. M. Fischer, Tank-tread frequency of the red cell membrane: Dependence on the viscosity of the suspending medium, *Biophys. J.* **93**, 2553 (2007).
- [80] T. M. Fischer and R. Korzeniewski, Angle of inclination of tank-treading red cells: Dependence on shear rate and suspending medium, *Biophys. J.* **108**, 1352 (2015).
- [81] V. R. Joseph, E. Gul, and S. Ba, Designing computer experiments with multiple types of factors: The MaxPro approach, *J. Qual. Technol.* **52**, 343 (2020).
- [82] S. Hénon, G. Lenormand, A. Richert, and F. Gallet, A new determination of the shear modulus of the human erythrocyte membrane using optical tweezers, *Biophys. J.* **76**, 1145 (1999).
- [83] M. M. Dupin, I. Halliday, C. M. Care, L. Alboul, and L. L. Munn, Modeling the flow of dense suspensions of deformable particles in three dimensions, *Phys. Rev. E* **75**, 1 (2007).
- [84] D. Salac and M. J. Miksis, Reynolds number effects on lipid vesicles, *J. Fluid Mech.* **711**, 122 (2012).
- [85] A. Yazdani and P. Bagchi, Influence of membrane viscosity on capsule dynamics in shear flow, *J. Fluid Mech.* **718**, 569 (2013).
- [86] A. Z. K. Yazdani and P. Bagchi, Phase diagram and breathing dynamics of a single red blood cell and a biconcave capsule in dilute shear flow, *Phys. Rev. E* **84**, 1 (2011).
- [87] P. M. Vlahovska, Y. N. Young, G. Danker, and C. Misbah, Dynamics of a non-spherical microcapsule with incompressible interface in shear flow, *J. Fluid Mech.* **678**, 221 (2011).
- [88] E. Foessel, J. Walter, A. V. Salsac, and D. Barthès-Biesel, Influence of internal viscosity on the large deformation and buckling of a spherical capsule in a simple shear flow, *J. Fluid Mech.* **672**, 477 (2011).
- [89] C. Dupont, F. Delahaye, D. Barthès-Biesel, and A. V. Salsac, Stable equilibrium configurations of an oblate capsule in simple shear flow, *J. Fluid Mech.* **791**, 738 (2016).
- [90] D. Cordasco and P. Bagchi, Orbital drift of capsules and red blood cells in shear flow, *Phys. Fluids* **25**, 091902 (2013).
- [91] Z. Peng, S. Salehyar, and Q. Zhu, Stability of the tank treading modes of erythrocytes and its dependence on cytoskeleton reference states, *J. Fluid Mech.* **771**, 449 (2015).
- [92] R. M. Hochmuth, P. R. Worthy, and E. A. Evans, Red cell extensional recovery and the determination of membrane viscosity, *Biophys. J.* **26**, 101 (1979).
- [93] X. Bian, S. Litvinov, R. Qian, M. Ellero, and N. A. Adams, Multiscale modeling of particle in suspension with smoothed dissipative particle dynamics, *Phys. Fluids* **24**, 012002 (2012).
- [94] J. A. Backer, C. P. Lowe, H. C. J. Hoefsloot, and P. D. Iedema, Poiseuille flow to measure the viscosity of particle model fluids, *J. Chem. Phys.* **122**, 154503 (2005).
- [95] J. M. Kim and R. J. Phillips, Dissipative particle dynamics simulation of flow around spheres and cylinders at finite Reynolds numbers, *Chem. Eng. Sci.* **59**, 4155 (2004).
- [96] Z. Zhang, E. Henry, G. Gompper, and D. A. Fedosov, Behavior of rigid and deformable particles in deterministic lateral displacement devices with different post shapes, *J. Chem. Phys.* **143**, 243145 (2015).
- [97] A. Yazdani and P. Bagchi, Three-dimensional numerical simulation of vesicle dynamics using a front-tracking method, *Phys. Rev. E* **85**, 1 (2012).
- [98] R. M. Hochmuth, K. L. Buxbaum, and E. A. Evans, Temperature dependence of the viscoelastic recovery of red cell membrane, *Biophys. J.* **29**, 177 (1980).
- [99] G. R. Cokelet and H. J. Meiselman, Rheological comparison of hemoglobin solutions and erythrocyte suspensions, *Science* **162**, 275 (1968).
- [100] libgp, <https://github.com/mblum/libgp> (retrieved June 2019).