Far-from-equilibrium distribution from near-steady-state work fluctuations

Robert Marsland III and Jeremy England

Physics of Living Systems Group, Massachusetts Institute of Technology, 400 Technology Square, Cambridge, Massachusetts 02139, USA

A long-standing goal of nonequilibrium statistical mechanics has been to extend the conceptual power of the Boltzmann distribution to driven systems. We report some new progress towards this goal. Instead of writing the nonequilibrium steady-state distribution in terms of perturbations around thermal equilibrium, we start from the linearized driven dynamics of observables about their stable fixed point, and expand in the strength of the nonlinearities encountered during typical fluctuations away from the fixed point. The first terms in this expansion retain the simplicity of known expansions about equilibrium, but can correctly describe the statistics of a certain class of systems even under strong driving. We illustrate this approach by comparison with a numerical simulation of a sheared Brownian colloid, where we find that the first two terms in our expansion are sufficient to account for the shear thinning behavior at high shear rates.

I. INTRODUCTION

For over a century, the formalism of equilibrium statistical mechanics has provided a powerful means to explain how the macroscopic properties of many-body systems at thermal equilibrium arise from the microscopic interactions that occur among their constituent parts. The centerpiece in this approach is the Boltzmann distribution, which posits that the probability of observing an equilibrated system with energy ϵ in microstate x at temperature $k_B T = 1/\beta$ is proportional to the so-called "Boltzmann weight" $p_{\text{bz}}(x) \propto \exp[-\beta\epsilon(x)]$. The key assumption used in deriving the Boltzmann distribution is that the system has spent an "ergodically" long time in contact with its surrounding heat bath, so that the combined setup of bath and system is equally likely to be in any arrangement that is allowed by conservation of energy. As a result, a quantity evaluated for the system at one instant in time [namely, $\epsilon(x)$] can immediately be translated into a probability of occurrence for the state x. This microscopic result can be coarse-grained to yield the probability of observing the system in a given macroscopic state X, defined as a set of microstates that share the same values of some observable properties. The coarse-grained probability $p_{\text{bz}}(X) \propto e^{-\beta F(X)}$ can then be written in terms of a free energy $F(X) = -k_B T \ln\left(\sum_{x \in X} e^{-\beta\epsilon(x)}\right)$.

Once time-varying fields drive the system from equilibrium by changing the energies $\epsilon(x,t)$ on time scales comparable to the system's relaxation time, the story must necessarily become more complicated. In the arbitrary nonequilibrium scenario, the probability of being at a given location in phase space at time t clearly can depend strongly on where the system was at some earlier moment. There is, however, a tempting special case to consider even when the Boltzmann distribution does not apply: in circumstances where $\epsilon(x,t)$ is periodic, the system may still ergodically lose its memory of initial conditions after enough time in contact with the fluctuating bath. In such a case, it is reasonable to consider whether the Boltzmann distribution admits a generalization, in which the probability of observing the system in a particular state after the memory of the initial state is lost can still be related exactly to some function of thermodynamic observables.

Yamada and Kawasaki answered this question in the affirmative almost fifty years ago, when they derived an

effective partition function for a generic nonequilibrium steady state in terms of correlations in the currents of conserved quantities passing through the system [1], launching a fruitful field of research on the exact microscopic distribution in driven steady states [2–5]. It is now well known that the simplicity of the Boltzmann weight cannot be reproduced in the microscopic probability distribution for an arbitrary driven steady state, which depends in general on all orders of the time correlations in the currents over the system's past history.

Thus any attempt to uncover simple principles beneath the statistics of nonequilibrium steady states must begin by specifying a particular regime of applicability, where certain simplifying approximations become valid. The near-equilibrium regime was the first to receive careful study, leading to an elegant representation of the steady-state distribution in terms of the dissipation due to externally imposed thermal gradients, chemical potential gradients, and velocity fields [6]. This "McLennan ensemble" applies to a wide range of near-equilibrium systems, and can be obtained through a variety of independent routes (cf. Ref. [7]). Most recently, it has been shown that this form can be derived in an especially transparent way from the assumption of "microscopic reversibility," which holds for a wide class of physical systems [5,8,9].

When external drives become arbitrarily strong, the steady-state distribution of observables in a generic physical system no longer follows this form. But some intuition about this regime can be built up around a special case, where a simple expression for the distribution does exist for arbitrarily strong driving, with the general expression written in terms of perturbations about this case. In this paper, we derive such an expansion about the case where fluctuations in the instantaneous dissipative current in the driven system obey a linear overdamped Langevin equation with additive white noise. In this scenario, a form essentially equivalent to the McLennan ensemble can be shown to hold regardless of the drive strength, and to be compatible with large departures from the predictions of linear-response theory.

This approach bears some resemblances to macroscopic fluctuation theory, which derives various statistical properties of far-from-equilibrium macroscopic systems from some minimal restrictions on the form of the macroscopic dynamics and the character of the current fluctuations [10]. In particular,

a simple form for the driven steady-state distribution is obtained that looks very similar to the McLennan ensemble, but holds for arbitrarily strong driving with arbitrary nonlinearities [11]. This result is obtained through a decomposition of the dissipative current into "symmetric" and "asymmetric" parts, which purchases a broader range of applicability by sacrificing the immediate physical meaning of the terms in the original McLennan form. We do not take advantage of this decomposition, but instead seek a representation of the distribution entirely in terms of the equilibrium free energy and the statistics of the bare externally applied work.

We start in Sec. II by deriving an exact expression for the steady-state distribution of an arbitrary observable in a system with microscopic reversibility whose distribution relaxes exponentially to a unique stationary state. Then in Sec. III we evaluate this expression in our special case, and compute the first correction in the perturbation expansion. In Sec. IV we apply this expansion to a sheared colloid, as a specific example of a strongly driven system. Section IV E gives a quantitative comparison with a numerical simulation of the sheared colloid, showing that the McLennan-like part is sufficient to reproduce the qualitative nonlinear response behavior, and that the first correction term gives good quantitative agreement deep into the shear thinning regime. Finally, in Sec. V we describe the thermodynamic intuition that can be extracted from this new form for the distribution, and define some avenues for further investigation.

II. DERIVATION OF DISTRIBUTION

A. Microscopic reversibility

Consider a generic physical system coupled to a large "heat bath" of inverse temperature $\beta = 1/k_B T$, but otherwise isolated, so that the total energy of the system plus bath can only be changed by external manipulation of a specified set of "control parameters" λ. These control parameters directly affect only the system energy $\epsilon(x, \lambda)$ and not the bath. As usually assumed in statistical mechanics, the potential energy of the interactions of system components with bath components is taken to be small compared with the energy in the system, so that the total energy can be cleanly divided between the energy "in the system" and the energy "in the bath." The bath volume is also held fixed, and the system volume is only allowed to vary if it is chosen as one of the control parameters λ.

The microstate x of the system evolves according to a stochastic process, due to its interactions with the fluctuating heat bath. If the combined setup is modeled with classical mechanics, x specifies generalized positions and momenta of all degrees of freedom in the system (not including the bath degrees of freedom). x can also be taken to be a discrete microstate label, with jumps between microstates governed by a rate matrix $W_{xx'}(\lambda(t))$. In this case, thermodynamic consistency is imposed by requiring that the rate matrix should eventually equilibrate the system to the Boltzmann distribution if the λ's are held fixed.

We keep the system out of equilibrium by varying the λ's according to a given protocol, which gives the microstate energies $\epsilon(x, \lambda(t))$ an explicit time dependence. When we

average over trajectories below, we will always assume that the $\lambda(t)$ protocol is held fixed.

Building on the work of Jarzynski [12], Crooks has shown that a number of important results concerning the nonequilibrium behavior of such a system can be derived from what he calls the "microscopic reversibility" condition [5]:

$$\frac{p_R[x^*(\Delta t - t)|x_2^*]}{p_F[x(t)|x_1]} = e^{-\beta Q_F[x(t)]}. \tag{1}$$

Here $x(t)$ is a system trajectory of duration Δt and the heat $Q_F[x(t)]$ is the energy transferred from the system to the bath over the course of that trajectory. The left-hand side contains the probability of taking the time-reversed path $x^*(\Delta t - t)$ given a starting state x_2^*, divided by the probability of taking the forward path $x(t)$ given the starting state x_1. The $*$ indicates time reversal of the microstate (changing the signs of all the momenta in the classical model). The R and F subscripts refer to the driving protocol $\lambda(t)$: the F probabilities and heat are computed with the protocol forward from time $-\Delta t$ to time 0, and the R probabilities have it running in reverse from time 0 to time $-\Delta t$. In the sheared colloid we will analyze below, the R and F quantities are computed with the shear applied in opposite directions.

Crooks showed that condition (1) holds for stochastic dynamics of discrete states under the thermodynamic consistency requirement given above [5]. This relation can also be derived directly from the time reversibility and phase space conservation of Hamiltonian dynamics, if the combined system-plus-bath setup can be treated as a closed Hamiltonian system driven by an explicit time dependence in the system Hamiltonian (cf. Ref. [13] for the basic approach, although a slightly different result is discussed there). The expression can be generalized to allow particle fluxes into and out of chemical baths, in which case an extra term involving chemical potentials must be added to the Q_F in the exponent [14–16].

B. Coarse graining

Now we group the system microstates x according to some observable properties, following the approach detailed in Ref. [17]. We can give each group a unique name, which will be generically represented by the capital letter X. Below, in the analysis of colloidal steady states, X will stand for a single number (the mean shear stress at the wall) characterizing the configuration of all the colloidal particles. In general, X represents a label for some group of microstates, which could be picked out using any definite procedure. X^* will refer to the group consisting of the time-reversed versions x^* of all the microstates in X.

The coarse-grained version of Eq. (1) involves transition probabilities among the different groups labeled by different X values. This probability will in general depend on how the initial state was prepared, since different protocols will give rise to different probability distributions of microstates within the macrostate. In this paper, however, we consider only systems with finite relaxation times, and our goal is to analyze the steady state that is reached after all correlations between current X values and the initial conditions have died out. In such a case, the choice of initial distribution becomes irrelevant to the steady-state statistics, and we can choose the

distribution that gives the most useful form for the final steady-state distribution of X.

We will call the initial distributions for the forward and reverse trajectories $p_1(x)$ and $p_2(x)$, respectively, and choose them to be Boltzmann distributions $p(x) = \exp\{-\beta[\epsilon(x,\lambda(t)) - F(X,\lambda(t))]\}$ over the microstates in the respective macrostates X_1 and X_2^*. $F(X,\lambda(t)) = -k_B T \ln \sum_{x \in X} \exp[-\beta\epsilon(x,\lambda(t))]$ gives the proper normalization for a distribution defined only over microstates x in macrostate X (where the sum becomes an integral in the classical case). To investigate steady-state behavior, we must vary these fields periodically, and consider trajectories whose duration Δt is an integer multiple of the period. This implies that $\epsilon(x,\lambda(-\Delta t)) = \epsilon(x,\lambda(0))$ and $F(x,\lambda(-\Delta t)) = F(x,\lambda(0))$, so for both cases we can drop the time dependence from the notation.

With these distributions in hand, we multiply Eq. (1) by the denominator of the left-hand side and by $p_2(x_2^*)$, then integrate over all trajectories $x(t)$ connecting states x_1 in a given macrostate X_1 to states x_2 in another macrostate X_2:

$$\int_{X_1 \to X_2} \mathcal{D}[x(t)] p_2(x_2^*) p_R[x^*(\Delta t - t)|x_2^*]$$
$$= \int_{X_1 \to X_2} \mathcal{D}[x(t)] e^{-\beta Q_F[x(t)]} \frac{p_2(x_2^*)}{p_1(x_1)} p_1(x_1) p_F[x(t)|x_1]. \tag{2}$$

We can simplify this expression by introducing the macroscopic transition probabilities

$$\pi_F(X_1 \to X_2) \equiv \int_{X_1 \to X_2} \mathcal{D}[x(t)] p_1(x_1) p_F[x(t)|x_1], \tag{3}$$

$$\pi_R(X_2^* \to X_1^*) \equiv \int_{X_1 \to X_2} \mathcal{D}[x(t)] p_2(x_2^*) p_R[x^*(\Delta t - t)|x_2^*], \tag{4}$$

which are defined as the sums of the probabilities of all microtrajectories $x(t)$ that accomplish the indicated macroscopic transition, given that the system begins in the indicated probability distribution over microstates (x_1 or x_2^*) in the starting macrostate (X_1 or X_2^*). Using these definitions to normalize the distributions over trajectories in the integrands in Eq. (2), we find

$$\pi_R(X_2 \to X_1) = \left\langle e^{-\beta Q_F[x(t)] + \ln \frac{p_2(x_2^*)}{p_1(x_1)}} \right\rangle_{X_1 \to X_2}$$
$$\times \pi_F(X_1 \to X_2). \tag{5}$$

The average $\langle \cdot \rangle_{X_1 \to X_2}$ is over all trajectories $x(t)$ connecting some microstate x_1 in X_1 (chosen from distribution p_1) to some other microstate x_2 in X_2.

Now we insert the explicit expressions for p_1 and p_2, to find

$$\frac{\pi_R(X_2^* \to X_1^*)}{\pi_F(X_1 \to X_2)} = e^{\beta[F(X_2) - F(X_1)]}$$
$$\times \langle e^{-\beta\{Q_F[x(t)] - \epsilon(x_1) + \epsilon(x_2)\}} \rangle_{X_1 \to X_2}. \tag{6}$$

We have dropped the $*$ in the arguments of ϵ and F, because the energy is symmetric under reversal of the signs of the momentum coordinates. Now we note that by conservation of energy, the work done on the system by the variation of

the control parameters $\lambda(t)$ over a trajectory $x(t)$ from x_1 to x_2 satisfies $W_F = Q_F[x(t)] + \epsilon(x_2) - \epsilon(x_1)$. We can thus rewrite the above expression as

$$\frac{\pi_R(X_2^* \to X_1^*)}{\pi_F(X_1 \to X_2)} = e^{\beta[F(X_2) - F(X_1)]}$$
$$\times \langle e^{-\beta W_F[x(t)]} \rangle_{X_1 \to X_2}. \tag{7}$$

The motivation for our choice of the Boltzmann distribution for p_1 and p_2 is precisely because it replaces the heat in the exponent with the work. Since the work is zero in the undriven case, where the control parameters are fixed, this choice splits the right-hand side cleanly into two factors, an equilibrium contribution and a nonequilibrium correction. Other choices could be made for these distributions, and they would generate valid alternative forms of the steady-state distribution.

We can use this expression to compare the probabilities of forward transitions from one state X_0 to two different states X_1 and X_2:

$$\ln \frac{\pi_F(X_0 \to X_1)}{\pi_F(X_0 \to X_2)}$$
$$= \beta[F(X_2) - F(X_1)] - \ln \frac{\langle e^{-\beta W_F[x(t)]} \rangle_{X_0 \to X_1}}{\langle e^{-\beta W_F[x(t)]} \rangle_{X_0 \to X_2}}$$
$$+ \ln \frac{\pi_R(X_1^* \to X_0^*)}{\pi_R(X_2^* \to X_0^*)}. \tag{8}$$

This expression is interesting in its own right, showing how the relative probabilities of different future possibilities depend not only on the free energies of the possible future states, but also on the work done on the way there and on the "durability" measure contained in the reverse probabilities (see Ref. [18] for a detailed analysis of the physical implications of the corresponding terms in a closely related expression). We now specialize to a system in which temporal correlations decay exponentially (or faster) with a finite relaxation time τ. In this case, $\pi(X_i \to X_f)$ must become independent of X_i for $\Delta t \gg \tau$, and simply be equal to the steady-state probability $p_{ss}(X_f)$ of being in the final state:

$$\ln \frac{p_{ss}^F(X_1)}{p_{ss}^F(X_2)} = \beta[F(X_2) - F(X_1)]$$
$$- \lim_{\Delta t \to \infty} \ln \frac{\langle e^{-\beta W_F[x(t)]} \rangle_{X_0 \to X_1}}{\langle e^{-\beta W_F[x(t)]} \rangle_{X_0 \to X_2}} + \ln \frac{p_{ss}^R(X_0^*)}{p_{ss}^R(X_0^*)}. \tag{9}$$

From this equation we can directly extract $p_{ss}^F(X)$ up to an overall constant $\mathcal{N}$ that is independent of X:

$$\ln p_{ss}(X) = -\beta F(X) - \lim_{\Delta t \to \infty} \ln \langle e^{-\beta W} \rangle_{X_0 \to X} + \mathcal{N}, \tag{10}$$

where we have dropped the F's from p_{ss} and W because we will only be considering "forward" quantities from now on.

Even though the work W in the exponent increases without bound as $\Delta t \to \infty$, this expression is well defined if one takes care to perform the average before taking the limit. For exponentially relaxing systems, $\langle e^{-\beta W} \rangle_{X_0 \to X}$ converges to a finite value as Δt increases beyond the relaxation time τ. This can be seen by considering the behavior of the left-hand side of Eq. (7) under these conditions: $\pi_F(X_1 \to X_2)$ and

$\pi_R(X_2^* \to X_1^*)$ approach the finite limiting values of $p_{ss}^F(X_2)$ and $p_{ss}^R(X_1^*)$, respectively, for $\Delta t \gg \tau$.

The individual cumulants of W do in general become infinite as $\Delta t \to \infty$, however. To obtain a useful series representation of the exponential average term, we can add the constant, finite quantity $\lim_{\Delta t \to \infty} \ln\langle e^{-\beta W}\rangle_{X_0 \to X_0}$ to the right-hand side, and compensate by adjusting the normalization constant $\mathcal{N}$. Representing both exponential average terms by their cumulant expansions, we obtain

$$\ln p_{ss}(X) = -\beta F(X) - \lim_{\Delta t \to \infty} \left(\sum_{n=1}^{\infty} \frac{(-\beta)^n}{n!} \langle W^n \rangle_{X_0 \to X}^c \right.$$
$$\left. - \sum_{n=1}^{\infty} \frac{(-\beta)^n}{n!} \langle W^n \rangle_{X_0 \to X_0}^c \right) + \mathcal{N}. \quad (11)$$

We can now rewrite this expression in terms of the finite differences between the cumulants for the trajectories ending in X and the corresponding cumulants for the trajectories ending in X_0:

$$\Delta\langle W^n \rangle^c(X) \equiv \lim_{\Delta t \to \infty} \left(\langle W^n \rangle_{X_0 \to X}^c - \langle W^n \rangle_{X_0 \to X_0}^c \right)$$
$$= \lim_{\Delta t \to \infty} \left(\langle W^n \rangle_{ss \to X}^c - \langle W^n \rangle_{ss \to X_0}^c \right). \quad (12)$$

Since we have assumed that the system's state becomes decorrelated from its past history after a finite time τ, the expression in parentheses becomes independent of Δt and remains finite as we take $\Delta t \to \infty$. In the second line, the ss $\to X$ averages are over the trajectories whose initial conditions are sampled from the steady-state distribution, and that end in state X. The equality follows from the fact that the contribution of the initial relaxation of the system from X_0 to the steady state is the same for both terms on the right-hand side of the first line.

Thus we obtain

$$\ln p_{ss}(X) = -\beta F(X) - \sum_{n=1}^{\infty} \frac{(-\beta)^n}{n!} \Delta\langle W^n \rangle^c(X) + \mathcal{N}. \quad (13)$$

Note that if X is an observable quantified by a continuous parameter, $p_{ss}(X)$ can be regarded as a probability density function for X. In this case, the left-hand side should strictly be written as $\ln[p_{ss}(X)\delta X]$, where a microstate counts as part of macrostate X if the value of the observable lies within δX of X. But the δX can be chosen to be the same for all X, and rolled into the overall constant $\mathcal{N}$.

Aside from the coarse-graining step, the general procedure we have followed thus far for extracting a steady-state distribution from microscopic reversibility and expanding in cumulants resembles the derivation by Komatsu *et al.* of an expansion about equilibrium in the strength of the driving field [8]. Expressions such as (13) obtained in this way contain cumulants of all orders, and only provide new physical insight if the series converges rapidly. Komatsu *et al.* show how a different way of writing the microscopic reversibility assumption leads to an expansion about equilibrium that converges particularly rapidly, and is accurate to second order in the strength of the drive without including any cumulants of higher order than 1 [8]. We are taking a different approach, treating Eq. (13) as an expansion in the size of the

nonlinearities in the coarse-grained equations of motion near the steady state. The coarse graining allows the parameters of the linearized dynamics to have a nontrivial dependence on the strength of the drive, so that the predictions of near-equilibrium linear response theory can break down while our expansion parameter is still near zero. In the next section, we lay out the details of our proposed expansion in terms of a Langevin model for the coarse-grained dynamics. Then we will use a simulated sheared colloid to illustrate a concrete case where the new expansion converges rapidly in a strongly driven system.

Writing the distribution in terms of the $\Delta\langle W^n \rangle^c$'s is convenient for the initial presentation of the theory, but for comparison with the colloid simulation, we need to work with finite Δt's. Thus we define

$$\langle W^n \rangle_{\Delta t}^c(X) \equiv \langle W^n \rangle_{ss \to X}^c$$
$$\approx \Delta\langle W^n \rangle^c(X) + \langle W^n \rangle_{ss \to X_0}^c, \quad (14)$$

where the averages are all over trajectories of length Δt, and the approximation holds for $\Delta t \gg \tau$. Since the second term in the second line is independent of X, replacing $\Delta\langle W^n \rangle^c(X)$ with $\langle W^n \rangle_{\Delta t}^c(X)$ in any expression for the steady-state distribution only affects the normalization. In terms of this new quantity, we thus obtain the alternative form

$$\ln p_{ss}(X) \approx -\beta F(X) - \sum_{n=1}^{\infty} \frac{(-\beta)^n}{n!} \langle W^n \rangle_{\Delta t}^c(X) + \mathcal{N}. \quad (15)$$

III. PERTURBATION ANALYSIS

Non-equilibrium steady states are distinguished from equilibrium states by the existence of nonzero mean currents J_{ss} of some conserved quantities (mass density, momentum, charge, etc.), so that the external field E conjugate to J does work on a system of volume V at a mean rate $\dot{W} = VEJ_{ss}$. In this section, we show that the cumulant differences $\Delta\langle W^n \rangle^c$ for $n > 1$ all vanish in the special case where a set of coarse-grained variables including the relevant J's can be found whose combined dynamics are described by a linear overdamped Langevin equation with additive white noise. This assumption of linearity does not imply a restriction to the linear response regime of near-equilibrium thermodynamics, as we show in the following subsection by computing the deviation of $J_{ss}(E)$ from its linear-response form in terms of the parameters of the linear model. In the final subsection we will examine small perturbations around this regime to see how the cumulant differences grow as nonlinearities are introduced.

For some of the calculations, we will assume that the observable X whose probability is being computed is one of the currents J. This assumption can be relaxed by a simple change of variables, whose Jacobian will be absorbed into the equilibrium term $F(X)$, as long as the dynamics of the current still satisfy the given requirements.

A. Linear regime

To compute the conditional averages $\langle \cdot \rangle_{ss \to X}$, we start by writing down an equation of motion for the observables in our system, such that the trajectories of the observables for a given realization of the noise can be found by solving the equation and applying the *final* condition that the trajectory *ends* in X at

$t = 0$. We will allow for an arbitrary number of observables, contained in a vector $\mathbf{X}$, but will allow for only one current J with steady-state value J_{ss} that is responsible for the steady-state work. This restriction can be relaxed without affecting the final result, but it simplifies the intermediate notation.

We start with the case of a linear Langevin equation for the fluctuation dynamics:

$$\dot{\mathbf{X}} = A(E)\mathbf{X} + B(E)\boldsymbol{\xi}(t), \tag{16}$$

where we have defined $\mathbf{X}$ such that $\mathbf{X} = 0$ is the most probable value in the steady state. We will choose the first element of $\mathbf{X}$ to contain the current, so that $X_1 = J - J_{\text{ss}}$. The matrices A and B are constant in time, but depend on the strength of the external field E. $\boldsymbol{\xi}(t)$ is a vector of Gaussian white noise, characterized by its mean $\langle \xi_i(t) \rangle = 0$ and two-point function $\langle \xi_i(t)\xi_j(t') \rangle = \delta(t - t')\delta_{ij}$.

With the ansatz $\mathbf{X} = e^{At}[\mathbf{X}(0) + \mathbf{f}(t)]$, we find

$$A\mathbf{X} + e^{At}\frac{d\mathbf{f}}{dt} = A\mathbf{X} + B\boldsymbol{\xi}(t),$$

$$\frac{d\mathbf{f}}{dt} = e^{-At}B\boldsymbol{\xi}(t),$$

$$\mathbf{f}(t) = -\int_t^0 dt' e^{-At'}B\boldsymbol{\xi}(t'), \tag{17}$$

so that $\mathbf{X} = e^{At}\mathbf{X}(0) - \int_t^0 dt'\, e^{A(t-t')}B\boldsymbol{\xi}(t')$ [keeping in mind that $t < 0$, since we are specifying the *final* condition $\mathbf{X}(0)$].

Since the first observable X_1 is equal to the deviation $J - J_{\text{ss}}$ from the mean steady state current, the work done over a given trajectory $\mathbf{X}(t)$ is

$$W = VE\int_{-\Delta t}^0 dt[X_1(t) + J_{\text{ss}}]$$

$$= VE\int_{-\Delta t}^0 dt[(e^{At})_{1j}X_j(0) + (e^{At})_{1j}f_j(t) + J_{\text{ss}}], \tag{18}$$

where we are implicitly summing over all the j's, using the Einstein summation convention.

We can now compute $\Delta\langle W\rangle[X_i(0)]$ as a function of any one of the parameters $X_i(0)$, by averaging W over paths that start in the steady state at time $t = -\Delta t \to -\infty$, and end at specified values of $X_i(0)$ at time $t = 0$:

$$\Delta\langle W\rangle[X_i(0)] = \langle W\rangle_{\text{ss}\to X_i(0)} - \langle W\rangle_{\text{ss}\to 0}$$

$$= VE\int_{-\infty}^0 dt(e^{A(E)t})_{1j}\langle X_j(0)\rangle_{X_i(0)}. \tag{19}$$

To obtain the higher cumulants, we first use the fact that the solution $\mathbf{X}(t)$ obtained above is a sum of independent Gaussian random variables, to conclude that the steady-state distribution for this linear relaxation is itself a Gaussian, which can be written as

$$p_{\text{ss}}(\mathbf{X}) \propto \exp\left(-\frac{1}{2}\sum C_{jk}^{-1}X_jX_k\right), \tag{20}$$

where C is the covariance matrix.

We can similarly show that the work distribution is Gaussian. Since all cumulants beyond the second are zero for a Gaussian distribution, we need only compute the contribution from the variance:

$$\Delta\langle W^2\rangle^c = \langle W^2\rangle_{\text{ss}\to X_i}^c - \langle W^2\rangle_{\text{ss}\to 0}^c$$

$$= V^2E^2\int_{-\infty}^0 dt'\int_{-\infty}^0 dt\,(e^{At'})_{1j}(e^{At})_{1k}$$

$$\times\left(\langle X_j(0)X_k(0)\rangle_{X_i}^c - \langle X_j(0)X_k(0)\rangle_0^c\right). \tag{21}$$

To compute the covariances in the above expression, we need to take a slice through the Gaussian steady-state distribution at the indicated fixed values of X_i. The distribution of the remaining variables $\mathbf{X}'$ in this slice is

$$p_{\text{ss}}(\mathbf{X}'|X_i) \propto \exp\left(-\frac{1}{2}\sum_{j,k\neq i}C_{jk}^{-1}X_jX_k - \sum_{j\neq i}C_{ij}^{-1}X_iX_j\right)$$

$$= \exp\left(-\frac{1}{2}\sum_{jk}C_{jk}'^{-1}X_j'X_k' - \sum_j\mu_jX_j'\right). \tag{22}$$

The mean of the new distribution depends on the vector $\mu_j = C_{ij}^{-1}X_i$ and the new covariance matrix $C' = \langle X_j(0)X_k(0)\rangle_{X_i}^c$ is found by removing the ith row and column from C^{-1} and then inverting it. The key property of this distribution for our purposes is that C' is independent of the value of X_i. It does depend in general on our choice of which row and column to remove, but it does not change when we vary X_i from 0 to some other value. Applying this fact to Eq. (21), we see that the term in parentheses equals zero for all values of X_i, and so Eq. (13) becomes

$$\ln p_{\text{ss}}(X_i) = -\beta[F(X_i) - \Delta\langle W\rangle(X_i)] + \mathcal{N}. \tag{23}$$

For this strictly linear case, then, the $\Delta\langle W^n\rangle^c$ vanish for all $n > 1$. This remains true even if the dynamics are not Markovian for the current J on its own [as is clearly the case in panel (a) of Fig. 2 below], as long as there exists a set of observables $\mathbf{X}$ with sufficiently many additional degrees of freedom that dynamics of the form given in Eq. (16) apply.

Equation (23) is closely related to the McLennan ensemble as presented in Eq. (3.13) of Ref. [8] [our averages are evaluated under the driven dynamics, instead of the undriven dynamics, but that change is $O(\epsilon^2)$ in their notation]. But as we will show below, the derivation we have provided can extend the range of validity for this representation beyond the linear-response regime to which it has been traditionally restricted.

This equation also looks similar to the result of macroscopic fluctuation theory that demonstrates the equality of the log of the steady-state distribution for the densities and the "excess work" [11]. There is no obvious relationship with this result, however, because we have defined $\Delta\langle W\rangle(X_i)$ by subtracting a constant from the bare mean work, rather than decomposing the current into symmetric and asymmetric parts.

B. Departure from linear response $J_{ss}(E)$

In the one-dimensional case, where $X = J - J_{ss}$, the dependence of $J_{ss}(E)$ on E is fully determined by the microscopic reversibility assumption underlying (23) combined with the linear equation of motion (16) and the definition of work (18). The result is

$$J_{ss}(E) = \frac{\beta V E}{A(E)} \langle J^2 \rangle_{eq}. \tag{24}$$

We can compare this to the prediction of the Green-Kubo formula of near-equilibrium linear response theory (cf. Ref. [4]):

$$J_{ss}^{LR}(E) = \beta V E \int_0^\infty dt \langle J(t) J(0) \rangle_{eq}. \tag{25}$$

Under the linear overdamped Langevin dynamics analyzed above, this becomes

$$J_{ss}^{LR}(E) = \frac{\beta V E}{A(0)} \langle J^2 \rangle_{eq}; \tag{26}$$

we see that the system departs from the linear-response regime when the damping rate A begins to depart from its equilibrium value $A(0)$. The size of the departure is given by

$$\frac{J_{ss} - J_{ss}^{LR}}{J_{ss}^{LR}} = \frac{A(0)}{A(E)} - 1. \tag{27}$$

Equations (23), (16), and (18) also impose a definite relationship between A and B:

$$\frac{B(E)^2}{2A(E)} = \langle J^2 \rangle_{eq}. \tag{28}$$

When $E = 0$, this is an analog to the Einstein relation, constraining the relationship between the effective "mobility" and effective "diffusion coefficient" for the fluctuations of the current in equilibrium. Equation (28) says that this relationship continues to hold beyond equilibrium, even where Eq. (27) indicates a large difference between the actual current and the predictions of linear response theory, as long as the equation of motion remains linear. This relationship guarantees that the variance of the steady-state distribution remains equal to its equilibrium value in this regime, even as the external drive alters the rate of relaxation. A related constraint also exists in the multidimensional case, where more parameters come into play, and the whole covariance matrix of the steady-state distribution must remain equal to the equilibrium matrix.

C. First correction

We now allow nonlinear terms to be introduced into the coarse-grained equations of motion. Our first goal is to obtain an expression in terms of work and temperature for the dimensionless expansion parameter that measures the significance of the nonlinearities in the computation of the steady-state distribution via Eq. (13). We also seek to determine which terms from the cumulant expansion are present at first order in this parameter. For this calculation, we again focus on the 1-D case, where the dynamics of the current J are Markovian. We begin by adding a small nonlinear term added to the linear dynamics for $X = J - J_{ss}(E)$:

$$\dot{X} = A(E)X + \frac{\mu}{2}X^2 + B(E)\xi(t), \tag{29}$$

where $\xi(t)$ is again a Gaussian white noise term with mean 0 and autocorrelation function $\langle \xi(0)\xi(t) \rangle = \delta(t)$. Note that the sign on A is positive, because we need to average over trajectories whose final conditions are specified, as opposed to the usual initial conditions. The coefficient μ has dimensions of 1/[time][current], and part of our task is to convert it into a dimensionless expansion parameter.

We can write the solution as a power series in μ:

$$X(t) = X^{(0)}(t) + \mu X^{(1)}(t) + \mu^2 X^{(2)}(t) + \cdots. \tag{30}$$

Plugging this into the equation of motion and collecting terms in powers of μ, we find

$$\dot{X}^{(0)} = AX^{(0)} + B\xi(t), \tag{31}$$

$$\dot{X}^{(1)} = AX^{(1)} + \tfrac{1}{2}(X^{(0)})^2. \tag{32}$$

The first is the same as the equation we solved above, giving

$$X^{(0)}(t) = e^{At}[X(0) + f(t)] \tag{33}$$

with $f(t) = -\int_t^0 e^{-At'} B\xi(t')dt'$. We can use this result in the second equation and solve in a similar way to obtain

$$X^{(1)}(t) = -\int_t^0 dt' \, e^{A(t-t')} \frac{[X^{(0)}(t')]^2}{2}. \tag{34}$$

Integrating the perturbative solution $X(t) = X^{(0)}(t) + \mu X^{(1)}(t) + O(\mu^2)$, we obtain the work done for a given realization of the noise:

$$
\begin{aligned}
W &= VE \int_{-\infty}^0 dt \left[e^{At} X(0) + e^{At} f(t) - \mu \int_t^0 dt' \, e^{A(t-t')} \frac{[e^{At'} X(0) + e^{At'} f(t')]^2}{2} + O(\mu^2) + J_{ss} \right] \\
&= \frac{VE}{A} X(0) - \frac{VE\mu}{4A^2} X(0)^2 + VE \int_{-\infty}^0 dt \left\{ e^{At} f(t) - \int_t^0 dt' \, e^{A(t+t')} \left[\mu X(0) f(t') + \frac{\mu}{2} f(t')^2 \right] + O(\mu^2) + J_{ss} \right\}.
\end{aligned}
\tag{35}
$$

Now we need to use this expression to compute the $\Delta \langle W^n \rangle^c(X)$. Since $f(t)$ is independent of the choice of ending state $X(0)$, we find

$$\Delta \langle W \rangle(X) = \frac{VE}{A(E)} X - \mu \frac{VE}{4A(E)^2} X^2 + O(\mu^2) = \Delta \langle W \rangle^{(0)}(X) - \mu \frac{VE}{4A(E)^2} X^2 + O(\mu^2), \tag{36}$$

where $\Delta\langle W\rangle^{(0)}(X)$ is computed under the linear dynamics alone. The quadratic term in Eq. (36) provides a correction to the equilibrium variance at order μ, which also breaks the relationship (28) between $A(E)$ and $B(E)$.

To compute the higher cumulants, we first note that the first two terms in Eq. (35) are not random variables, and serve as a constant offset that makes no contribution to the cumulants of order 2 and higher. Focusing on the term in curly brackets, then, we see that the $\mu f(t')^2$ term can only be part of an $X(0)$-dependent term in $\Delta\langle W^n\rangle^c$ when it is multiplied by some nonzero power of the $\mu X(0)f(t')$ term. It therefore contributes only to the $O(\mu^2)$ part of the expression and to the overall normalization. The remaining part of the work can be expressed as a sum of independent Gaussian random variables $\xi(t)$, so it is itself a Gaussian random variable. This implies that it has no nonzero cumulants beyond the variance, so that all higher cumulants are $O(\mu^2)$. The variance is given by

$$
\begin{aligned}
\langle W^2\rangle^c_{\text{ss}\to X} &= \langle W^2\rangle_{\text{ss}\to X} - \langle W\rangle^2_{\text{ss}\to X} \\
&= -2V^2E^2\mu \int_{-\infty}^0 dt \int_{-\infty}^0 dt' \int_{t'}^0 dt''\, e^{A(t+t'+t'')}X \\
&\quad \times \langle f(t)f(t'')\rangle + O(\mu^2) + \mathcal{N},
\end{aligned}
\tag{37}
$$

where we have absorbed the X-independent terms into $\mathcal{N}$.

To simplify this, we must compute the autocorrelation function of $f(t)$. We first consider the case $|t''| \geqslant |t|$:

$$
\begin{aligned}
\langle f(t)f(t'')\rangle &= B^2 \int_t^0 ds \int_{t''}^0 du\, e^{-A(s+u)}\langle \xi(s)\xi(u)\rangle \\
&= B^2 \int_t^0 ds \int_{t''}^t du\, e^{-A(s+u)}\delta(s-u) \\
&\quad + B^2 \int_t^0 ds \int_t^0 du\, e^{-A(s+u)}\delta(s-u) \\
&= B^2 \int_t^0 ds\, e^{-2As} = \frac{B^2}{2A}(e^{-2At}-1).
\end{aligned}
\tag{38}
$$

If $|t''| \leqslant |t|$, this becomes

$$
\langle f(t)f(t'')\rangle = \frac{B^2}{2A}(e^{-2At''}-1).
\tag{39}
$$

Combining the two answers, we find

$$
\langle f(t)f(t'')\rangle = \frac{B^2}{2A}(e^{-2At_m}-1),
\tag{40}
$$

where t_m is equal to whichever of t,t'' has the smaller absolute value.

Plugging this into Eq. (37), we find

$$
\begin{aligned}
\langle W^2\rangle^c_X &= -2V^2E^2\mu X \frac{B^2}{2A} \\
&\quad \times \int_{-\infty}^0 dt \int_{-\infty}^0 dt' \int_{t'}^0 dt''\, (e^{A(t'-|t-t''|)} \\
&\quad - e^{A(t+t'+t'')}) + O(\mu^2) + \mathcal{N} \\
&= -\mu\frac{V^2E^2XB^2}{A^4} + O(\mu^2) + \mathcal{N}.
\end{aligned}
\tag{41}
$$

Thus we obtain the next term in the cumulant expansion in Eq. (13):

$$
\begin{aligned}
\frac{\beta^2}{2}\Delta\langle W^2\rangle^c(X) &= \frac{\beta^2}{2}\left(\langle W^2\rangle^c_{\text{ss}\to X} - \langle W^2\rangle^c_{\text{ss}\to 0}\right) \\
&= -\mu\frac{\beta^2 V^2 E^2 B^2}{2A^4}X + O(\mu^2).
\end{aligned}
\tag{42}
$$

By the argument given above Eq. (37), the rest of the terms in the cumulant expansion are guaranteed to be $O(\mu^2)$. To obtain the correct dimensionless form of μ, we need to compare this term to the $\beta\Delta\langle W\rangle$ term, and see what combination of parameters controls their relative size. Using Eq. (36), we find

$$
\frac{\beta^2}{2}\Delta\langle W^2\rangle^c(X) = -\mu\frac{\beta VEB^2}{2A^3}\beta\Delta\langle W\rangle(X) + O(\mu^2),
\tag{43}
$$

so that $\tilde{\mu} = -\mu\frac{\beta VEB^2}{2A^3}$ is the appropriate dimensionless version of μ that controls how quickly the expansion converges.

We can interpret $\tilde{\mu}$ thermodynamically by comparing to the quadratic $O(\mu)$ term in (36) that corrects the variance of the distribution. Using the fact that the variance in the unperturbed steady-state distribution is $\sigma_X^2 = B^2/2A$ (cf. Eqs. (3.8.74) and (4.3.23) in Ref. [19]), we can write

$$
\tilde{\mu} = 4\beta[\Delta\langle W\rangle(\sigma_X) - \Delta\langle W\rangle^{(0)}(\sigma_X)],
\tag{44}
$$

which is four times the typical extra mean work difference due to the nonlinear term in the dynamics, divided by k_BT.

Thus we conclude that the first terms in the expansion of $p_{\text{ss}}(X)$ about the linearized dynamics in this 1-D case are given by

$$
\begin{aligned}
\ln p_{\text{ss}}(X) &= -\beta[F(X) - \Delta\langle W\rangle(X)] - \frac{\beta^2}{2}\Delta\langle W^2\rangle^c(X) \\
&\quad + O(\tilde{\mu}^2) + \mathcal{N}
\end{aligned}
\tag{45}
$$

with $\tilde{\mu}$ as defined above.

In the multidimensional case, defining the $\tilde{\mu}$ explicitly in terms of the model parameters is more challenging, but it is reasonable to suppose that the thermodynamic expression (44) should still give a good estimate of its size. In the next section, we will apply this result to our numerical simulation of a sheared colloid, which will require the finite-time approximate form

$$
\begin{aligned}
\ln p_{\text{ss}}(X) &\approx -\beta[F(X) - \langle W\rangle_{\Delta t}(X)] - \frac{\beta^2}{2}\langle W^2\rangle^c_{\Delta t}(X) \\
&\quad + O(\tilde{\mu}^2) + \mathcal{N},
\end{aligned}
\tag{46}
$$

which holds when $\Delta t \gg \tau = 1/A$, with the finite-time cumulants defined in Eq. (14).

IV. SHEAR THINNING EXAMPLE

Strongly driven colloids are commonly used as examples of far-from-equilibrium steady-state systems where the relationship between currents and applied fields can violate the predictions of linear response theory (cf. [20–24]). Using the shear stress σ_{xy} as the current and minus the shear rate $-\dot{\gamma}$ as the conjugate field, one can describe departures from linear response in terms of the dependence of the viscosity $\eta = -\sigma_{xy}/\dot{\gamma}$ on $\dot{\gamma}$. In the linear response regime, σ_{xy} is proportional

to $\dot{\gamma}$, and η is constant. As the shear rate is increased in a typical colloid, $\sigma_{xy}(\dot{\gamma})$ becomes sublinear, so η decreases and the suspension is said to "shear thin" (cf. Ref. [25]). In this section, we describe how to measure $\langle W \rangle_{\Delta t}(\sigma_{xy})$ and $\frac{\beta^2}{2}\langle W^2 \rangle_{\Delta t}^c(\sigma_{xy})$ in a computer model of a colloid that exhibits shear thinning. We will show that the above expansion in the degree of nonlinearity converges quickly in this case, so that the $\tilde{\mu} \to 0$ form given in Eq. (23) generates a qualitatively correct description of the shear thinning phenomenon, while the $O(\tilde{\mu})$ correction in Eq. (46) is sufficient to maintain quantitative agreement with the actual distribution well into the thinning regime.

A. Setup

Consider a suspension of small identical spheres in a Newtonian solvent at a fixed temperature. The particles are small enough that Brownian motion can equilibrate their spatial configuration rapidly compared with the duration of a typical experiment, producing a steady state independent of initial conditions. Electrostatic repulsion keeps the spheres far enough apart that the disturbance each particle creates in the flow field has no effect on the trajectories of the other particles, while ions in the solvent screen the charges and exponentially suppress the interaction at large separations.

A nonequilibrium steady state can be created by moving one wall of the chamber containing the suspension at a constant velocity v while keeping the opposite wall fixed, thus setting up a steady shear flow in the gap of width d between the walls. [The wall velocity is ultimately due to some time-varying fields $\lambda(t)$—such as the fields inside an electric motor.] The strength of the shear flow can be quantified in a form independent of the system dimensions as the "shear rate" $\dot{\gamma} = v/d$. A constant shear rate can be maintained by using periodic boundary conditions in the flow direction (which can be approximated in experiment by using a cylindrical geometry). As indicated in Fig. 1, we will define coordinates such that the moving wall travels in the $+x$ direction and the y axis points from the stationary wall to the moving wall.

Two important dimensionless parameters for the dynamics of a sheared colloid are the Reynolds number $\mathrm{Re} = \rho \dot{\gamma} a^2 / \eta_0$ and the Peclet number $\mathrm{Pe} = \dot{\gamma} a^2 b / k_B T$. Here ρ is the mass density of the fluid (assumed to be comparable to the density of the particles), a is the radius of a particle,
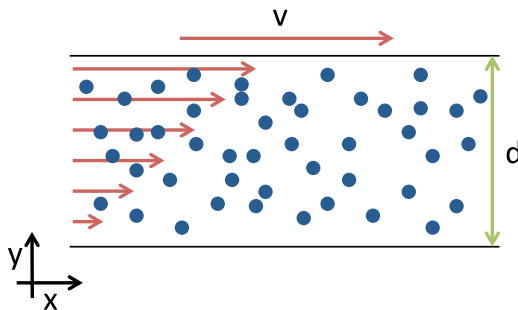

FIG. 1. (Color online) Shear cell with periodic boundary conditions along flow direction. The reflecting walls on the top and bottom are separated by a distance d. The top wall moves at constant speed v in the x direction, while the bottom wall is fixed, causing a linear gradient $\dot{\gamma} = v/d$ in the solvent flow velocity along the y direction.

η_0 is the viscosity of the suspending fluid, b is the drag coefficient of a particle ($=6\pi\eta_0 a$ for a sphere with no-slip boundary conditions), k_B is Boltzmann's constant, and T is the temperature of the heat bath coupled to the fluid. Re measures the importance of inertia relative to viscous drag, and Pe measures the importance of motion by convection in the shear flow relative to diffusive motion in a dilute suspension (when the suspension becomes sufficiently dense, the equilibrium relaxation is slowed down by the interactions among the particles, and the relevant dimensionless parameter becomes the density-dependent Weissenberg number). Pe thus measures the distance from equilibrium, so that $\mathrm{Pe} \ll 1$ gives rise to the linear-response regime, and $\mathrm{Pe} \gg 1$ constitutes the "far from equilibrium" regime where shear thinning occurs.

In the overdamped limit where $\mathrm{Re} \ll 1$, the instantaneous velocity of the particles can be regarded as fully determined by their spatial configuration (up to the rapidly equilibrating contribution from Brownian motion), so the set of particle positions is sufficient to define the full microstate, and the corresponding simulation algorithm becomes very simple. Re can be kept in this regime while sweeping Pe up to any desired maximum value $\mathrm{Pe}_{\max}$ by choosing a viscosity such that $\eta_0 \gg \sqrt{\rho \mathrm{Pe}_{\max} k_B T/a}$.

B. Equations of motion

To describe this system mathematically, we will use the model employed in Refs. [23,24] for the investigation of departures from near-equilibrium linear-response behavior in nonequilibrium steady states. This model can be numerically simulated with the dilute limit of the Brownian dynamics of Ermak and McCammon [26] or of the Stokesian dynamics of Brady and Bossis [27], where hydrodynamic interactions are ignored, and the $\mathrm{Re} \ll 1$ limit is invoked so that the particle inertia becomes irrelevant. These assumptions lead to the following discretized equations of motion for the position (x_i, y_i) of particle i confined to move in two dimensions:

$$x_i(t+\Delta t) = x_i(t) + y_i(t)\dot{\gamma}\Delta t + \frac{1}{b}\sum_j \hat{\mathbf{x}} \cdot \mathbf{F}_{ji}\Delta t + \Delta x_i^r,$$

(47)

$$y_i(t+\Delta t) = y_i(t) + \frac{1}{b}\sum_j \hat{\mathbf{y}} \cdot \mathbf{F}_{ji}\Delta t + \Delta y_i^r,$$ (48)

where b is the drag coefficient for the particles, and $\mathbf{F}_{ji}$ is the force exerted on particle i by particle j. We choose the force to be a screened Coulomb repulsion, with potential energy $U(r) = k_B T e^{-r/\lambda} z l_B / r$ as a function of the distance r separating a pair of particles. λ is the screening length, l_B is the Bjerrum length, and z is the number of elementary charges on each particle. Δx_i^r and Δy_i^r are random displacements due to Brownian motion. Brownian and Stokesian dynamics assume that the solvent degrees of freedom always equilibrate before the spatial configuration of the colloidal particles can change appreciably, so that the Brownian displacements can be chosen from a Gaussian distribution whose variance $(2k_B T/b)\Delta t$ is related to the temperature and drag coefficient by the Einstein relation. Equations (47) and (48) are then simply iterated with

a small enough time step that the results are insensitive to variations in time-step size [28].

As mentioned at the beginning of this section, we are considering the case where the particle size is much smaller than λ or zl_B, so that "hydrodynamic interactions" (particle-particle interactions mediated by disturbances in the solvent flow) make a negligible impact on the particle trajectories. This is what allows us to use the "dilute limit" of the Stokesian or Brownian dynamics, where the mobility and resistance tensors are diagonal and independent of particle positions. Another consequence of this limit is that the actual particle radius a does not appear in the equations of motion; we therefore use the screening length λ as the microscopic length scale for computing the Peclet number and measuring distance from equilibrium.

C. Shear stress

The macroscopic viscosity of the whole suspension at equilibrium will be larger than η_0, because both the disturbance of the flow field produced by individual particles and the mutual repulsion between pairs of particles make the suspension harder to shear than the bare fluid. As the suspension is sheared, however, the contribution of the particle repulsion to the viscosity decreases, and the suspension shear thins (cf. Ref. [25]). The particles cause the shear stress to vary with position in the suspension, so we define an overall shear stress for the system by averaging the local shear stress at the moving wall of the system over the whole wall area. This will be convenient for computing the work done by the moving wall later on, and gives us a macroscopic parameter that can be directly observed in experiment via a measurement of the force applied to the wall. As shown in the Appendix, for a suspension of particles in a Newtonian solvent in the limit of zero Re with no hydrodynamic interactions, the instantaneous mean shear stress $\sigma_{xy}^{\text{wall}}$ exerted by the fluid on the moving wall is

$$\sigma_{xy}^{\text{wall}} = \sigma_{xy}^I + \sigma_{xy}^0. \tag{49}$$

The first term depends on the force $\mathbf{F}_{ij}$ exerted by each particle i on the other particles j:

$$\sigma_{xy}^I = \frac{1}{2V} \sum_{i \neq j} \hat{\mathbf{x}} \cdot \mathbf{F}_{ij} \Delta y_{ij}. \tag{50}$$

Here V is the system volume, $\hat{\mathbf{x}}$ is the unit vector in the $+x$ direction, and $\Delta y_{ij} = y_j - y_i$. The right-hand side can be unambiguously determined from the system microstate, which we are taking to be the list of positions of all the particles. We can therefore choose $X = \sigma_{xy}^I$ as our coarse-grained observable and test how well the first terms in our expansion describe its fluctuations at a fixed shear rate $\dot{\gamma}$. The remaining term σ_{xy}^0 is independent of the particle positions, so the work done by the moving wall will depend on the particle configuration through σ_{xy}^I alone.

When the shear rate is small compared to the diffusive relaxation rate, the overall viscosity of the suspension can be computed from the equilibrium fluctuations in σ_{xy} using linear response theory [4]. As the shear rate continues to increase, the viscosity begins to deviate from this value as the suspension shear thins. In the following sections, we will use Eq. (46)

to determine the most probable value of σ_{xy}^I and hence the contribution $-\sigma_{xy}^I/\dot{\gamma}$ to the viscosity in both the linear response regime and the shear thinning regime. We start in Sec. IV D by demonstrating how to extract $\langle W \rangle_{\Delta t}(\sigma_{xy}^I)$ and $\frac{\beta^2}{2}\langle W^2 \rangle_{\Delta t}^c(\sigma_{xy}^I)$ from the simulation in a way that should also be experimentally accessible. Then in Sec. IV E we plot the results over a range of shear rates, and compare the prediction of Eq. (46) with the observed steady-state distribution.

D. Measuring the work cumulants

The rate at which the moving wall does work on the fluid is just the force $-A\sigma_{xy}^{\text{wall}}$ it exerts against the fluid (where A is the surface area of the wall) times the speed of the wall $\dot{\gamma}d$. Using Eq. (49), we thus obtain

$$\dot{W} = -V\dot{\gamma}\sigma_{xy}^I + \dot{W}_0, \tag{51}$$

where $\dot{W}_0$ is the part of the work that does not depend on the configuration of the particles. Since $\dot{W}_0$ only affects the normalization, but not the shape of the distribution, we will set it to zero for the purpose of the calculations in this section. Then we can treat σ_{xy}^I as the current J and minus the shear rate $-\dot{\gamma}$ as its conjugate field E.

Using Eq. (51), we can now compute the work done along any stochastic trajectory $\sigma_{xy}^I(t)$ by simply integrating the trajectory with respect to time. We can then compute the distribution of W for trajectories ending at a given σ_{xy}^I value by letting the system relax to the steady state at some value of $\dot{\gamma}$ and run there for a long time, while continuously recording the fluctuations in σ_{xy}^I (which can in principle be determined directly from the fluctuations in the force applied to the moving plate in an experiment). As shown in Fig. 2, we then choose some time interval Δt longer than the time scale τ of relaxation to the steady state, and compute both the work and the final value of σ_{xy}^I for every segment of length Δt in the whole trajectory. Finally, we bin the work outputs by the corresponding final value of σ_{xy}^I to obtain the distribution of work for each bin, from which we can compute the cumulants $\langle W^n \rangle_{\Delta t}^c(\sigma_{xy}^I)$.

Figure 3 shows the zeroth-order term $\langle W \rangle_{\Delta t}(\sigma_{xy}^I)$ in the expansion (46) and the first-order correction $(\beta/2)\langle W^2 \rangle_{\Delta t}^c(\sigma_{xy}^I)$ as a function of ending value of σ_{xy}^I, at three different values of the dimensionless shear rate Pe.

E. Simulation results

With this procedure in hand for extracting $\langle W \rangle_{\Delta t}$ and $\langle W^2 \rangle_{\Delta t}^c(\sigma_{xy}^I)$, we can use Eq. (46) to compute the steady-state distribution of σ_{xy}^I from the work statistics, and compare it to the directly measured distribution in our simulation.

We simulated a sheared colloidal monolayer of $N = 100$ particles using the equations of motion (47) and (48) [28]. The colloid was confined to a square box of side length 20, with reflecting boundary conditions on the moving wall and the opposite wall, and periodic boundary conditions on the other sides. This concentration is sufficiently dilute that the equilibrium relaxation time does not vary much with changes in concentration, so the Peclet number should still be the relevant parameter for measuring distance from equilibrium.
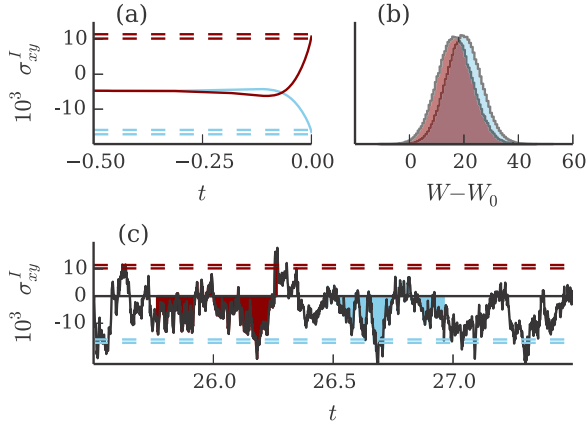
FIG. 2. (Color online) (a) $\sigma_{xy}^{I}(t)$ averaged over all trajectory segments of length $\Delta t = 0.5$ that end at a specified value, from a simulation run with the same parameters as Fig. 4, and Pe = 19. Two arbitrarily chosen ending state restrictions are indicated by dotted lines, with the corresponding average trajectories shown in the same colors. (b) Measured probability distributions of work done over all the trajectories that go into the averages of panel (a), shaded in the same colors. (c) A portion of the raw σ_{xy}^{I} time series from which the other two panels were generated. The ending state restrictions of panel (a) are indicated with dotted lines of the same color, and two typical trajectory segments ending at these values are shaded in these colors. The shaded areas are proportional to the interaction-dependent part of the work, according to Eq. (51).

The other parameters were chosen as $k_B T = b = \lambda = z l_B = 1$. We ran this simulation for 20 different values of Pe, from 0 to 19, generating trajectories with lengths up to $t = 120\,000$ in the given units, with time step size 0.001. The simulations were initialized with uniform random distributions of particle positions, and the initial transients were removed from the time series before analysis.
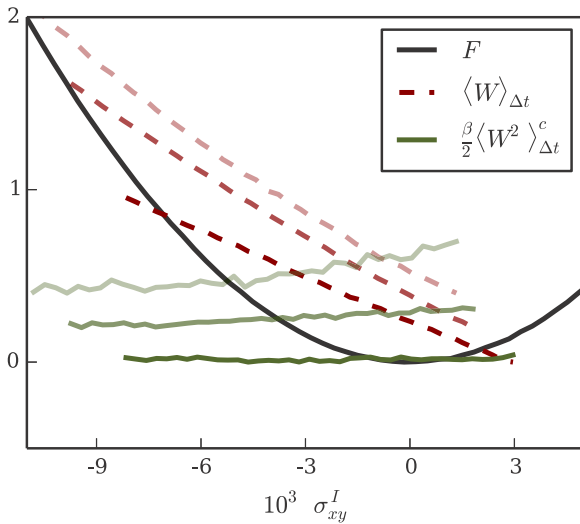

FIG. 3. (Color online) Using the method illustrated in Fig. 2, we compute $\langle W \rangle_{\Delta t}(\sigma_{xy}^{I})$ and $(\beta/2)\langle W^2 \rangle_{\Delta t}^{c}(\sigma_{xy}^{I})$ for a range of values of σ_{xy}^{I}, and plot this data for Pe values 4, 10, and 19, increasing from bottom to top. Also plotted is the free energy F extracted from the Pe = 0 simulation run. Traces have been vertically shifted for easier comparison of slopes.
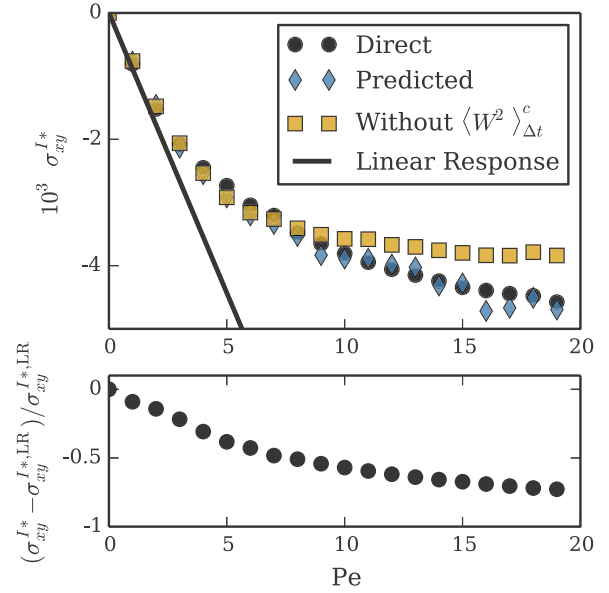

FIG. 4. (Color online) Top: Location σ_{xy}^{I*} of peak of probability distribution, computed via Eq. (46), compared with directly measured most frequent values in the simulation [28]. Also plotted is the prediction using the mean work alone via Eq. (23), without the $O(\tilde{\mu})$ variance term, as well as the linear response prediction $\sigma_{xy}^{I*,\mathrm{LR}}$. Bottom: Relative size of departure from linear-response prediction.

To apply Eq. (46), we first need to find the equilibrium free energy $F(\sigma_{xy}^{I})$ as a function of σ_{xy}^{I}. This can be extracted from the simulation by extracting the distribution of equilibrium fluctuations from a run at $\dot{\gamma} = 0$, taking the natural logarithm, and multiplying by $-k_B T$.

Figure 3 shows how F and the other two terms in Eq. (46) depend on σ_{xy}, with the nonequilibrium terms evaluated at three different values of the shear rate. F is parabolic near $\sigma_{xy}^{I} = 0$, but requires a fourth-order polynomial to fit the far tails. $\langle W \rangle_{\Delta t}$ starts out linear at low shear rates, starts curving slightly by Pe = 10, and becomes noticeably quadratic by Pe = 19, indicating that the $O(\tilde{\mu})$ term has become important. $\langle W^2 \rangle_{\Delta t}^{c}$ is independent of σ_{xy}^{I} at low shear rates, but starts becoming σ_{xy}^{I}-dependent at about the same shear rate at which $\langle W \rangle_{\Delta t}$ begins to deviate from linearity, as expected from our analysis in Sec. III C.

Figure 4 shows the location of the peak of the steady-state distribution as a function of Pe computed using Eq. (46), compared with the distribution directly sampled from the simulation. We have also plotted the result ignoring the $\Delta \langle W^2 \rangle^{c}$ term, to show the size of the impact of the $O(\tilde{\mu})$ correction compared to the zeroth-order expression (23). When we compare both curves to the most probable values of σ_{xy}^{I} actually measured in the simulations, we see that the zeroth-order expression correctly captures the qualitative shear thinning behavior: σ_{xy}^{I} departs from its linear-response dependence on $\dot{\gamma}$ and eventually saturates, causing the interparticle force contribution $-\sigma_{xy}^{I}/\dot{\gamma}$ to the viscosity to fall off as $1/\dot{\gamma}$. This approximation predicts that the saturation occurs sooner than it actually does, but the first-order term appears to entirely compensate for the discrepancy. The straight line is the linear-response prediction for the mean shear stress,
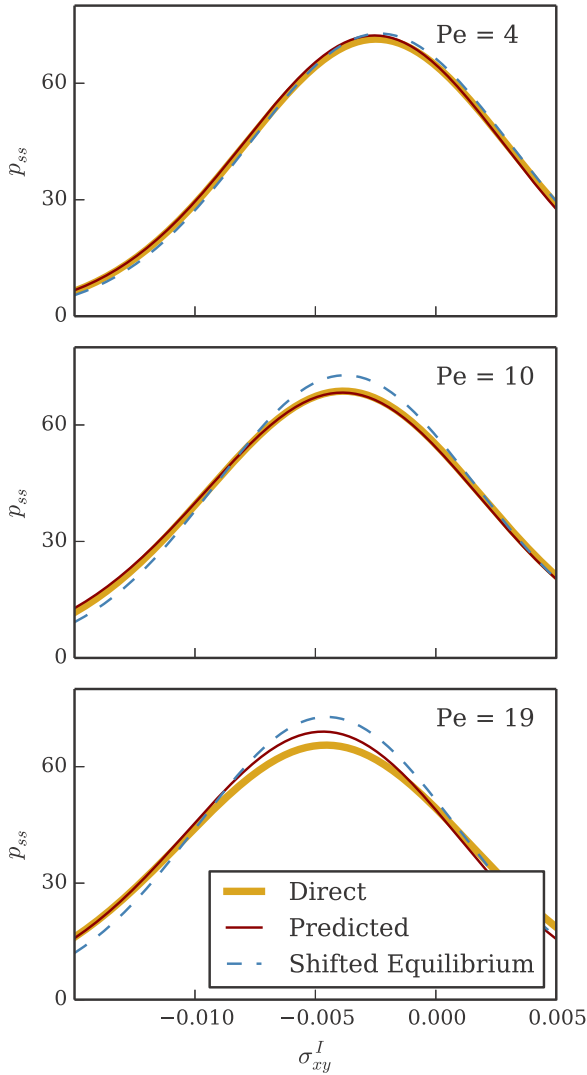
FIG. 5. (Color online) Full probability distribution from Eq. (46), compared with shifted equilibrium distribution and with distribution directly sampled from simulation at Pe = 4, Pe = 10, and Pe = 19.

computed from the equilibrium fluctuations using the Green-Kubo formula given in Eq. (25). The bottom panel shows the relative difference between the linear-response prediction and the location of the observed peak of the probability distribution, as defined in Eq. (27) of Sec. III.

In Fig. 5 we have plotted the normalized probability distributions at Pe = 4, 10, and 19, showing the direct measurement and the prediction of Eq. (46), smoothed with polynomial fits to the data in Fig. 3. Also plotted is the equilibrium distribution shifted to the location of the new peak, to better visualize the change in the shape of the distribution at high shear rates. The variance is still within 2% of the equilibrium value at Pe = 4, even though the relative deviation of the mean from the linear-response prediction has already reached 30%. This is an example of a case where the constraint of Eq. (28) derived from the linear approximation remains in effect beyond the linear-response regime. Equation (46) successfully accounts for the change in variance visible at Pe = 10, although it fails to fully capture the asymmetry that

enters the distribution at Pe = 19, which should depend on higher-order terms according to the analysis of Sec. III C.

V. DISCUSSION

We have shown that the probability distribution of a macroscopic observable in a driven steady state can be written as a perturbation expansion in the nonlinearities of the coarse-grained dynamics, which can converge quickly in some systems even under strong driving. The approximate formulas (23) and (45) give a good description of the statistics of the steady state under the following conditions:

(1) The transition rates between system states must satisfy the "microscopic reversibility" condition given in Eq. (1).

(2) The system exponentially loses memory of its initial conditions, with a decay time that is short compared to the duration of the experiment or simulation.

(3) The contribution to the mean work difference $\Delta \langle W \rangle$ [defined in Eq. (12)] due to nonlinear terms in the coarse-grained fluctuation dynamics is small compared to the thermal energy scale $k_B T$.

The simulation results presented in Sec. IV D and Sec. IV E illustrate the physical meaning of the terms in the expansion, and confirm that it can converge rapidly even when external drives are strong enough to make the peak of the steady-state distribution depart significantly from the linear-response prediction.

Evaluating the distribution by measuring $\langle W \rangle_{\Delta t}$ and $\langle W^2 \rangle^c_{\Delta t}$ numerically, as done in Sec. IV E, is much more computationally expensive than a direct measurement of the distribution from the simulation data. This computation was done in order to demonstrate that the statistics of the simulated system can be captured by the first terms in the perturbation expansion, but the real significance of our result lies in the physical intuition that can be obtained for the steady-state behavior of systems that reside in this regime.

The zeroth-order expression (23) constitutes a natural extension of the first attempts by Einstein and his contemporaries at understanding macroscopic fluctuations (cf. Ref. [29]). They interpreted the equilibrium macroscopic distribution to mean that the probability of a fluctuation in a given observable is exponentially suppressed in the ratio of the work that must be done by the rest of the degrees of freedom in the system to produce this fluctuation to the thermal energy scale $k_B T$. Equation (23) simply incorporates the fact that in a driven system, some of the work can be done by the external drive instead of by other internal degrees of freedom. By subtracting off the extra work done by the drive on the way to a given fluctuation (compared to the work done on the way to a fixed reference state), we account for the "help" provided by the drive in supplying the work required to reach each state. The first correction term from the perturbation analysis begins to account for the stochasticity in this extra work: even if two possible fluctuations extract the same amount of work on average from the drive, one can be less likely than the other if its distribution of extracted work extends further towards zero, allowing the fluctuation to be reached by paths that receive little help from the drive.

For strictly linear relaxation of a single dissipative current, the mean extra work from the drive can be computed up

to an additive constant C in terms of the relaxation time $\tau(E) = 1/A(E)$ using Eq. (19) from Sec. III: $\Delta\langle W\rangle(J) = VEJ\tau(E) + C$. For systems close to this regime, the nonlinear response of the steady-state distribution to an increase in the drive can thus be estimated from knowledge of the behavior of the relaxation time. In our sheared colloid case, we can understand the shape of the stress vs shear rate curve in Fig. 4 this way. The equilibrium relaxation time $\tau(0)$ is set by the diffusive time scale ba^2/k_BT, but in the $\dot\gamma \to \infty$ limit, $\tau(E)$ should be dominated by convective "stirring" from the shear flow with time scale $1/\dot\gamma$. The shift in the peak of the σ_{xy}^I distribution should thus increase linearly with the strength of the conjugate field $E = -\dot\gamma$ near equilibrium, with the expected linear-response coefficient via Eq. (26); but the peak should stop shifting once $\tau(\dot\gamma)$ reaches its asymptotic $1/\dot\gamma$ behavior, which cancels out all the $\dot\gamma$ dependence.

As we pointed out in Eq. (28) in Sec. III, the linear regime also exhibits the remarkable property that the relaxation time $\tau(E)$ and the coefficient $B(E)$ on the noise term in the Langevin equation are related by a generalized Einstein relation or fluctuation-dissipation theorem. A decrease in $\tau(\dot\gamma)$ must accompanied by an increase in the strength of the noise term in the coarse-grained dynamics, in such a way that the variance of the steady-state distribution remains exactly equal to the variance of the equilibrium distribution. The size of the fluctuations thus remains unchanged in the regime of linear relaxation, even as the dynamics are significantly altered, with a new time scale and (as in the case of this sheared colloid) a new underlying physical mechanism. The first panel of Fig. 5 confirms that this constraint remains applicable even when the linear-response prediction for the mean shear stress is no longer valid.

One could now explore the application of the theory to richer phenomena, such as hydrocluster formation in shear thickening colloids (cf. Refs. [30,31]), where the distribution of some measure of typical cluster size could be computed in terms of the mean work done for trajectories that end at a given value of that parameter. It would also be interesting to investigate whether suspensions of active particles in a quiescent solution can be analyzed in this way, possibly delivering further insight into the "freezing by heating" transition that takes place at a critical propulsion rate in both experiment and simulation [32–34].

Equation (13) should also be readily generalizable to chemical as opposed to mechanical driving, using the extensions of Eq. (1) to chemical reaction networks mentioned in Sec. II [14–16]. Once the quantitative relationship between the bulk quantities of interest and the work rate are understood, this generalized result could shed light on the steady-state properties of biologically relevant systems, such as active actin-myosin networks driven by ATP hydrolysis [35–37].

APPENDIX: DERIVATION OF MEAN WALL STRESS [EQUATIONS (49) AND (50)]

Formulas for determining the particle contribution to the shear stress of a colloidal suspension have been known for a long time, and received an especially careful treatment by Batchelor in the 1970s [38,39]. The established literature mainly deals with the *mean* shear stress, either averaged over an infinite ensemble of systems or over an infinitely large system. The statistical uniformity of the system can then be invoked to argue that the mean stress over a typical 2-D slice through the system is equal to the mean stress averaged over the whole system volume. Although the wall is not a *typical* 2-D slice, because the boundary condition modifies the particle distribution, the fact that there is no mean net force on any part of the system when it is in steady state implies that the mean stress on all parallel 2-D slices must be the same. The average over an infinite system volume must therefore also be equal to the average over an infinite wall [38].

For the purpose of this paper, it is not enough to know the ensemble- or infinite-system-averaged mean. We need to look at the fluctuations about the mean in order to apply our procedure for empirically determining the mean renormalized work and the equilibrium free energy as a function of the shear stress. Therefore we need to go back through the derivation, and examine the *instantaneous* value of the shear stress at the wall in a suspension of a *finite* number of particles.

In this Appendix, we prove that the instantaneous shear stress exerted by the fluid on the moving wall of the shear apparatus described in Sec. III, averaged over the moving wall area, is

$$\sigma_{xy}^{\text{wall}} = \sigma_{xy}^I + \sigma_{xy}^0, \tag{A1}$$

where σ_{xy}^I is defined by

$$\sigma_{xy}^I \equiv \frac{1}{2V}\sum_{i\neq j}\hat{\mathbf{x}}\cdot\mathbf{F}_{ij}\Delta y_{ij}, \tag{A2}$$

and σ_{xy}^0 is independent of the particle positions.

We start by giving some necessary background on the behavior of shear stress in low-Re Newtonian fluids. To make this proof accessible to readers less familiar with hydrodynamics, we then map to a mathematically analogous problem in electrostatics (which turns out to be a homework problem from Griffiths' *Electricity and Magnetism* [40, problem 3.44a]). After presenting the solution to this electrostatics problem, we finally map back to hydrodynamics to obtain our final result.

1. Stress in Newtonian fluids

The shear stress σ_{xy} is an off-diagonal component of the 3×3 stress tensor σ. σ is defined at each point in the fluid such that $\hat{\mathbf{n}}\cdot\sigma$ is the force per unit area exerted from below on a surface element at that location with unit normal vector $\hat{\mathbf{n}}$. By "from below," we mean from the side opposite to the direction of the normal vector. We will focus on the x column $\sigma\cdot\hat{\mathbf{x}}$ to obtain a vectorial quantity that will be easier to visualize.

By the definition of the stress tensor above, the x component of the force on a region Ω of fluid is given by

$$F_x = -\int_{\partial\Omega} d\mathbf{A} \cdot \sigma \cdot \hat{\mathbf{x}} \qquad (A3)$$

$$= -\int_{\Omega} dV \boldsymbol{\nabla} \cdot \sigma \cdot \hat{\mathbf{x}}, \qquad (A4)$$

where $d\mathbf{A}$ is an infinitesimal area element of the boundary $\partial\Omega$ pointing along the outward normal direction, and dV is an infinitesimal volume element. We add the minus sign because we are computing the force on this surface from the outside. The second line results from the divergence theorem. Since this holds for every possible region Ω, we conclude that the integrand of Eq. (A4) is equal to minus the x component of force per unit volume f_x exerted by the surrounding fluid on an infinitesimal volume element:

$$\boldsymbol{\nabla} \cdot \sigma \cdot \hat{\mathbf{x}} = -f_x. \qquad (A5)$$

Finally, we must invoke the assumption that the solvent in which the particles are suspended is a Newtonian fluid, which implies

$$\sigma \cdot \hat{\mathbf{x}} = -\eta_0 \boldsymbol{\nabla} u_x, \qquad (A6)$$

where u_x is the x component of the fluid velocity field, and η_0 is the (constant) viscosity of the solvent. Combining this with the previous equation gives us the set of equations

$$\eta_0 \nabla^2 u_x = f_x, \qquad (A7)$$

$$\sigma \cdot \hat{\mathbf{x}} = -\eta_0 \boldsymbol{\nabla} u_x \qquad (A8)$$

that together fully determine $\sigma \cdot \hat{\mathbf{x}}$ for a given set of boundary conditions.

2. Mapping to electrostatics

Equations (A7) and (A8) suggest a mapping to electrostatics. $\eta_0 u_x$ is the analog to the electric potential ϕ, $\sigma \cdot \hat{\mathbf{x}}$ is the analog to the electric field $\mathbf{E}$, and $-f_x$ is the analog to the charge density ρ. With these mappings, the mathematics of the problem are identical to electrostatics, and we can do everything in terms of $\mathbf{E}$, ϕ, and ρ until we map back at the end.

The only remaining piece of setup is to map the boundary conditions and the "charge distribution." The nonslip boundary condition requires that every part of the fluid in contact with a nonrotating rigid surface must share the same velocity. Since the electric potential ϕ is the analog of the x component of velocity, this implies that nonrotating surfaces behave like perfect conductors—they are always equipotentials. In particular, the constraint that the bottom wall is fixed and the top wall moves at constant velocity v implies that the walls of the shear cell become parallel conducting plates separated by a distance d, with fixed electric potential difference $\Delta\phi$. The problem of determining the total force on the walls is thus equivalent to determining the induced charge on these conducting plates.

The particles, however, are allowed to rotate. Their boundary conditions are therefore more complicated, involving the

other columns of the stress tensor. Specifically, we have

$$\mathbf{u} = \Omega \times \mathbf{r} + \mathbf{u}_{\rm cm} \qquad (A9)$$

for all points on the surface of the sphere, where $\mathbf{r}$ is the vector pointing from the center of the sphere to the surface point, and Ω is the angular velocity vector. Ω and the center-of-mass velocity $\mathbf{u}_{\rm cm}$ are free parameters that must be adjusted so as to be consistent with Eqs. (A7) and (A8). The resulting restriction on σ is

$$\sigma = -\eta_0 \boldsymbol{\nabla}(\Omega \times \mathbf{r} + \mathbf{u}_{\rm cm}). \qquad (A10)$$

To determine the charge distribution, we use our assumption of low Re to require the total force on any volume element to vanish. In the electrostatic analogy, this implies that the solvent is uncharged, and all charge must reside at the walls or on the particles. The interparticle repulsion exerts force on each particle that must be canceled by the friction of the fluid in order to satisfy the requirement of zero total force. This implies that the total "charge" on each particle must be $q_i = \sum_{j \neq i} \mathbf{F}_{ji} \cdot \hat{\mathbf{x}}$, where $\mathbf{F}_{ji}$ is the force exerted on particle i by particle j. The distribution of this total charge over the surface of each sphere is not fixed in advance, however, and must be determined by solving Eqs. (A7) and (A8) (along with the corresponding equations for the other components of the stress tensor) with the boundary conditions just described. The decision to "ignore hydrodynamic interactions" mentioned in the main text allows us to greatly simplify the problem of determining these distributions, by solving the equations for each particle individually, with boundary condition $\mathbf{E} \rightarrow -(\Delta\phi/d)\hat{\mathbf{y}}$ far from the sphere. This approximation ignores the effect of the other particles and of the induced wall charge on the charge distribution over each sphere. The solutions obtained under this approximation are independent of the particle positions, which will be important later on.

3. Obtaining the induced charge on the conducting plate

Our problem is thus reduced to determining the induced charge on a pair of conducting parallel plates at fixed electric potential due to a given charge distribution inside.

We start by splitting the charge on the plates into two parts, following the strategy of Batchelor in his treatment of the effect of particle interactions on mean shear stress [39]. The derivation will resemble Batchelor's in many ways, despite the electrostatic language, but adds a new element by considering the wall stress due to a given *instantaneous* configuration of particles as opposed to an ensemble average of all possible configurations.

The first part of the charge is the part required to maintain the electric potential difference $\Delta\phi$ in the absence of any additional charges between the plates: $Q_0 = A\Delta\phi/d$ on the top and $-Q_0$ on the bottom. To find the remaining charge, we can solve for the case where the two plates are grounded. When we add up the two charge distributions, the resulting field is guaranteed to produce the desired constant electric potential difference. The case of grounded plates is problem 3.44a in Griffiths, as mentioned above, and we will follow his method to solve it [40].

Griffiths starts by having the student derive a relation known as Green's reciprocity theorem. (This theorem is closely

related to a result due to Lorentz in hydrodynamics, which Batchelor employs in his analysis [41].) Consider two distinct charge distributions $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$, which produce electric fields $\mathbf{E}_1(\mathbf{r})$ and $\mathbf{E}_2(\mathbf{r})$, with electric potentials $\phi_1(\mathbf{r})$ and $\phi_2(\mathbf{r})$. Now use the Maxwell equation $\nabla \cdot \mathbf{E} = \rho$ and the definition of electric potential $\mathbf{E} = -\nabla \phi$ to obtain

$$\int dV \mathbf{E}_1 \cdot \mathbf{E}_2 = -\int dV \nabla \phi_1 \cdot \mathbf{E}_2 = \int dV \phi_1 \nabla \cdot \mathbf{E}_2$$

$$= \int dV \phi_1 \rho_2 \tag{A11}$$

$$= -\int dV \nabla \phi_2 \cdot \mathbf{E}_1 = \int dV \phi_2 \nabla \cdot \mathbf{E}_1$$

$$= \int dV \phi_2 \rho_1, \tag{A12}$$

where we are integrating over all space, and have used integration by parts to switch the ∇ from ϕ to $\mathbf{E}$.

We thus obtain Green's reciprocity theorem:

$$\int dV \phi_1 \rho_2 = \int dV \phi_2 \rho_1. \tag{A13}$$

Now we use this relation to compute the induced charge on our plates. We will start by computing the induced charge due to a point charge q at location $\mathbf{r} = (x, y, z)$. We will work in coordinates where the bottom plate is at $y = 0$ and the top is at $y = d$.

To apply the reciprocity theorem, we choose for ρ_1 the actual charge distribution we are analyzing, with the point charge between the grounded parallel plates. We define Q_+ as the total induced charge on the top plate and Q_- as the total induced charge on the bottom plate. For ρ_2, we choose a charge distribution with conducting plates in the same locations, but with the top plate fixed at electric potential ϕ_0 above the bottom one, and with no charge in the space between them. The left-hand side of the reciprocity theorem vanishes, because $\phi_1 = 0$ whenever ρ_2 is nonzero. The right-hand side has a contribution from the charge distribution on the top plate, and a contribution from the particle. If the plates are infinite, then the potential a distance y above the bottom plate in scenario 2 is exactly $(y/d)\phi_0$. This will still be a good approximation in a finite system for charges that are not too close to the edges of the system, which will be true for the charges on the vast majority of the spheres when the number of spheres is large. Thus we obtain

$$0 = \phi_0 Q_+ + \phi_0 \frac{y}{d} q. \tag{A14}$$

Solving for Q_+, we find

$$Q_+ = -\frac{y}{d} q. \tag{A15}$$

Now we again use the linearity of our equations to obtain the total induced charge by summing up the contributions from all the infinitesimal charge elements in the distribution. A convenient way to perform this sum is to split up the charge distribution on each sphere into two parts: a spatially uniform part equal to the mean surface charge on the sphere, and a spatially varying part that integrates to zero over each sphere surface.

a. Contribution of variations about the mean. We start by computing the contribution of the second part of the charge distribution. Since this part of the charge sums to zero on each sphere, every positive charge δq has a corresponding negative charge $-\delta q$ somewhere else on the sphere. The net induced charge from each such pair is

$$\delta Q_+ = \frac{\delta q}{d}(y_- - y_+), \tag{A16}$$

where y_- and y_+ are the coordinates of the $+\delta q$ and $-\delta q$ charges, respectively. Now recall that by ignoring hydrodynamic interactions, we can solve for the charge distribution over each sphere without knowing its position relative to the plates or the other particles. Furthermore, the linearity of the governing equations implies that the variations about the mean charge density are independent of the size of the mean. This implies that the y distance $y_- - y_+$ between any pair of charges on a single sphere is independent of the spatial configuration of the particles and of the total charge q_i of the particle in question.

Summing over all pairs of charges from all the spheres in the sample, we define the quantity

$$Q_H = \sum \delta Q_+ \tag{A17}$$

as the total induced charge due to the variations about the mean charge on the surface of the spheres. This quantity is independent of the particle positions, and just adds a constant offset to the total charge. The H subscript stands for "hydrodynamic," because this contribution comes purely from the friction of the flow field around each particle.

b. Contribution of the mean charge. To complete our calculation, we must compute the charge induced on the plate by a given configuration of uniformly charged spheres. Since the field of a uniformly charged sphere is equivalent to the field of a point charge (for points outside the surface of the sphere), we can simply evaluate the point charge solution derived above for every particle, and add them all up. We thus find

$$Q_I = -\sum_i \frac{y_i}{d} q_i. \tag{A18}$$

Combining the above results, we find that the total induced charge on the top plate is $Q = Q_I + Q_0 + Q_H$, with Q_I the only term that depends on the particle positions.

4. Mapping back to hydrodynamics

We can now map back into the original variables (recalling that charge is equivalent to minus the force exerted by the fluid) in order to obtain the total force exerted by the fluid on the moving wall of the shear apparatus:

$$F_{\text{wall}} = \sum_i \frac{y_i}{d} \left(\sum_{j \neq i} \hat{\mathbf{x}} \cdot \mathbf{F}_{ji} \right) + F_0 + F_H. \tag{A19}$$

We can simplify this expression by using the fact that $\mathbf{F}_{ji} = -\mathbf{F}_{ij}$:

$$F_{\text{wall}} = \frac{1}{2d} \sum_{i \neq j} \hat{\mathbf{x}} \cdot \mathbf{F}_{ij} \Delta y_{ij} + F_0 + F_H. \tag{A20}$$

Finally, we can divide through by the area A of the wall to obtain the mean shear stress exerted on the wall by the fluid:

$$\sigma_{xy}^{\text{wall}} = \sigma_{xy}^{I} + \sigma_{xy}^{0} + \sigma_{xy}^{H}, \tag{A21}$$

where

$$\sigma_{xy}^{I} = \frac{1}{2V} \sum_{i \neq j} \hat{\mathbf{x}} \cdot \mathbf{F}_{ij} \Delta y_{ij} \tag{A22}$$

and the other two terms are independent of the particle positions. For notational simplicity, we combine them into one term in the main text, which we call σ_{xy}^{0}.

[1] T. Yamada and K. Kawasaki, Prog. Theor. Phys. **38**, 1031 (1967).
[2] G. P. Morriss and D. J. Evans, Mol. Phys. **54**, 629 (1985).
[3] G. P. Morriss and D. J. Evans, Phys. Rev. A **37**, 3605 (1988).
[4] D. J. Evans and G. P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids* (Cambridge University Press, Cambridge, 2008).
[5] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[6] J. A. McLennan, Phys. Fluids **3**, 493 (1960).
[7] J. L. Lebowitz, H. L. Frisch, and E. Helfand, Phys. Fluids **3**, 325 (1960).
[8] T. S. Komatsu, N. Nakagawa, S. Sasa, and H. Tasaki, J. Stat. Phys. **134**, 401 (2009).
[9] C. Maes and K. Netočný, J. Math. Phys. **51**, 015219 (2010).
[10] L. Bertini, A. D. Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, Rev. Mod. Phys. **87**, 593 (2015).
[11] L. Bertini, D. Gabrielli, G. Jona-Lasinio, and C. Landim, J. Stat. Phys. **149**, 773 (2012).
[12] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[13] C. Jarzynski, J. Stat. Phys. **98**, 77 (2000).
[14] H. Qian, J. Phys. Chem. B **109**, 23624 (2005).
[15] H. Ge and H. Qian, Phys. Rev. E **87**, 062125 (2013).
[16] P. Gaspard, J. Chem. Phys. **120**, 8898 (2004).
[17] J. England, J. Chem. Phys. **139**, 121923 (2013).
[18] N. Perunov, R. Marsland, and J. England, arXiv:1412.1875.
[19] C. Gardiner, *Stochastic Methods*, 4th ed. (Springer-Verlag, Berlin, 2009).
[20] M. Fuchs and M. E. Cates, J. Phys.: Condens. Matter **17**, S1681 (2005).
[21] T. Harada, Phys. Rev. E **79**, 030106 (2009).
[22] M. Krüger and M. Fuchs, Phys. Rev. Lett. **102**, 135701 (2009).
[23] T. Speck and U. Seifert, Phys. Rev. E **79**, 040102(R) (2009).
[24] M. Zhang and G. Szamel, Phys. Rev. E **83**, 061407 (2011).
[25] J. M. Brader, J. Phys.: Condens. Matter **22**, 363101 (2010).
[26] D. L. Ermak and J. A. McCammon, J. Chem. Phys. **69**, 1352 (1978).
[27] J. F. Brady and G. Bossis, Annu. Rev. Fluid Mech. **20**, 111 (1988).
[28] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.92.052120 for PYTHON scripts.
[29] A. Einstein, Ann. Phys. **338**, 1275 (1910).
[30] N. J. Wagner and J. F. Brady, Phys. Today **62**(10), 27 (2009).
[31] X. Cheng, J. H. McCoy, J. N. Israelachvili, and I. Cohen, Science **333**, 1276 (2011).
[32] G. S. Redner, M. F. Hagan, and A. Baskaran, Phys. Rev. Lett. **110**, 055701 (2013).
[33] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin, Science **339**, 936 (2013).
[34] D. Helbing, I. J. Farkas, and T. Vicsek, Phys. Rev. Lett. **84**, 1240 (2000).
[35] A. C. Martin and B. Goldstein, Development **141**, 1987 (2014).
[36] A. J. Ridley, Cell **145**, 1012 (2011).
[37] M. Guo, A. J. Ehrlicher, M. H. Jensen, M. Renz, J. R. Moore, R. D. Goldman, J. Lippincott-Schwartz, F. C. Mackintosh, and D. A. Weitz, Cell **158**, 822 (2014).
[38] G. K. Batchelor, J. Fluid Mech. **41**, 545 (1970).
[39] G. K. Batchelor, J. Fluid Mech. **83**, 97 (1977).
[40] D. Griffiths, *Introduction to Electrodynamics* (Prentice Hall, Upper Saddle River, NJ, 1999).
[41] H. A. Lorentz, Versl. Konigl. Akad. Wetensch. Amst. **5**, 168 (1896).