

Irreversibility in Active Matter Systems: Fluctuation Theorem and Mutual Information

Lennart Dabelow

Fakultät für Physik, Universität Bielefeld, 33615 Bielefeld, Germany

Stefano Bo and Ralf Eichhorn

*Nordita, Royal Institute of Technology and Stockholm University,
Roslagstullsbacken 23, SE-106 91 Stockholm, Sweden* (Received 13 June 2018; revised manuscript received 22 February 2019; published 15 April 2019)

We consider a Brownian particle which, in addition to being in contact with a thermal bath, is driven by fluctuating forces which stem from active processes in the system, such as self-propulsion or collisions with other active particles. These active fluctuations do not fulfill a fluctuation-dissipation relation and therefore play the role of a nonequilibrium environment, which keeps the system permanently out of thermal equilibrium even in the absence of external forces. We investigate how the out-of-equilibrium character of the active matter system and the associated irreversibility is reflected in the trajectories of the Brownian particle. Specifically, we analyze the log ratio of path probabilities for observing a certain particle trajectory forward in time versus observing its time-reversed twin trajectory. For passive Brownian motion, it is well known that this path probability ratio quantifies irreversibility in terms of entropy production. For active Brownian motion, we show that in addition to the usual entropy produced in the thermal environment, the path probability ratio contains a contribution to irreversibility from mutual information “production” between the particle trajectory and the history of the nonequilibrium environment. The resulting irreversibility measure fulfills an integral fluctuation theorem and a second-law-like relation. When deriving and discussing these relations, we keep in mind that the active fluctuations can occur either due to a suspension of active particles pushing around a passive colloid or due to active self-propulsion of the particle itself; we point out the similarities and differences between these two situations. We obtain explicit expressions for active fluctuations modeled by an Ornstein-Uhlenbeck process. Finally, we illustrate our general results by analyzing a Brownian particle which is trapped in a static or moving harmonic potential.

DOI: [10.1103/PhysRevX.9.021009](https://doi.org/10.1103/PhysRevX.9.021009)

Subject Areas: Statistical Physics

I. INTRODUCTION AND MAIN RESULTS**A. Introduction**

Active particle systems consist of individual entities (“particles”) which have the ability to perform motion by consuming energy from the environment and converting it into a self-propulsion drive [1–7]. Prototypical examples are collections of macroorganisms, such as animal herds, schools of fish, flocks of birds, or ant colonies [8–10], and suspensions of biological microorganisms or artificial microswimmers, such as bacteria and colloidal particles with catalytic surfaces [2,3,5,11,12]. Systems of this kind exhibit a variety of intriguing properties, e.g., clustering

and swarming [13–18], bacterial turbulence [19], or motility-induced phase separation [20,21] to name but a few.

Microorganisms and microswimmers are usually dispersed in an aqueous solution at room temperature and therefore experience thermal fluctuations which give rise to a diffusive component in their self-propelled swimming motion. In addition, the self-propulsion mechanism is typically noisy in itself [2], for instance, due to environmental factors or intrinsic stochasticity of the mechanisms creating self-propulsion. These “active fluctuations” exhibit two essential features. First, a certain persistence in the direction of driving over length- and timescales comparable to observational scales. Second, an inherent nonequilibrium character as a consequence of permanently converting and dissipating energy in order to fuel self-propulsion. Interestingly, similar active fluctuations with the same characteristics can be observed in a complementary class of active matter systems, namely, a passive colloidal “tracer” particle which is suspended in an aqueous solution of active swimmers. The collisions with the active

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

particles in the environment entail directional persistence and nonequilibrium features in the motion of the passive tracer particle [22–24]. The sketch in Fig. 1 illustrates these two types of active particle systems.

Despite the inherent nonequilibrium properties of active matter systems, they appear to bear striking similarities to equilibrium systems [21,25–27], for instance, the dynamics of individual active particles at large scales often looks like passive Brownian diffusion [3] (i.e., at scales beyond at least the “persistence length” of the particle motion). In connection with the ongoing attempts to describe active matter systems by thermodynamic (-like) theories, this observation raises the important question, “How far from equilibrium is active matter?” [28,29]. More specifically, the questions are (i) in which respects do emergent properties of active systems resemble the thermodynamics of thermal equilibrium systems, (ii) in which respects do they not, and (iii) how do these deviations manifest themselves in observables describing the thermodynamic character of the active matter system, like, e.g., the entropy production?

Such questions should be answered *independently* of the specific microscopic processes creating the active driving because, while these processes maintain the active particle system out of equilibrium, they contain little or no information on whether the emergent system behavior appears equilibriumlike or not. It is, for instance, of little relevance for the system properties emerging on the coarse-grained scale of particle trajectories (e.g., the “diffusion” of active particles) and their similarities or dissimilarities to thermal equilibrium, if an active particle is self-propelled by chemical processes occurring at its surface or by a beating flagellum. In this spirit, here we address the above questions from a fundamental nonequilibrium statistical

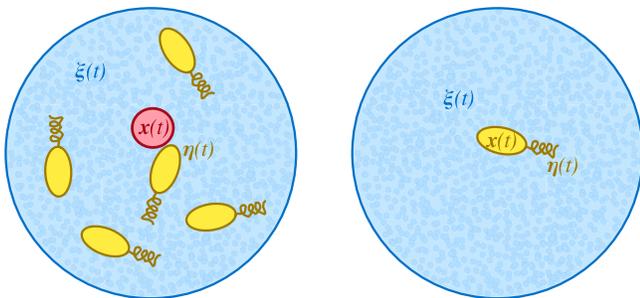


FIG. 1. Illustration of the two different kinds of active matter systems. Left: Passive Brownian particle (red) in a thermal bath (blue) with a suspension of active self-propelled particles (yellow); collisions of the active particles with the passive one are modeled by a fluctuating force $\eta(t)$. Right: Self-propelled particle (yellow) in a thermal bath (blue); the fluctuating self-propulsion velocity is modeled to result from a fluctuation force $\eta(t)$. In both cases, thermal fluctuations are described by unbiased δ -correlated Gaussian white-noise sources $\xi(t)$, and the position of the particle of interest is denoted by $x(t)$; see Eqs. (5)–(7) in the main text.

mechanics viewpoint by developing a trajectorywise thermodynamic description of active particle systems as a natural generalization of stochastic energetics [30,31] and thermodynamics [32–36] of passive particles in a purely thermal environment. In particular, we study the breakdown of time-reversal symmetry in the motion of particles which are driven by thermal and active fluctuations. We quantify the associated irreversibility of individual particle trajectories as a measure for nonequilibrium by comparing the probability of a specific particle trajectory to occur forward in time to that of observing its time-reversed counterpart. We link the corresponding probability ratio to functionals over the forward trajectory and provide a thermodynamic interpretation of the resulting extensive quantities in terms of entropy production and mutual information.

In the case of passive particles in contact with a single purely thermal bath, the path probability ratio is a well-established fundamental concept in stochastic thermodynamics for assessing irreversibility [32–35]. It is known to provide relations between thermodynamic quantities, such as entropy production, work, or heat, and dynamical properties encoded in path probability ratios. Many such relations have been found over the last two decades [33,34,37–39], which, in their integrated forms, typically yield refinements of the second law of thermodynamics [33,38,40]. In particular, the fluctuation theorem for the total entropy production of a passive particle in a thermal environment reinforces the fundamental interpretation of entropy as a measure of irreversibility (we give a brief summary of the results most relevant to our present work in Sec. III). In the presence of active fluctuations, however, the identification of dissipated heat and entropy production is less straightforward and is connected to the problem of how to calculate and interpret the path probability ratio in a physically meaningful way [41,42].

In the present work, we analyze the path probability ratio based on particle trajectories, i.e., from the evolution of the particle positions in time, without explicitly modeling the (system-specific) microscopic processes behind the active fluctuations. This approach allows for deriving general results on the nonequilibrium character and thermodynamic content of the dynamical behavior emerging in active matter systems independent of the specific processes generating the active fluctuations but with the caveat that dissipation and irreversibility occurring in these processes cannot be assessed [41]. The derived quantitative irreversibility and thermodynamic measures are easily accessible in standard experiments, e.g., from recording particle trajectories using video microscopy. In order to account for the active fluctuations, which either stem from an active bath the (passive) particle is dispersed in or from active self-propulsion (see Fig. 1 and Sec. IV A for specific examples of both cases), without resolving the microscopic processes that govern the interactions between active and passive particles or drive self-propulsion, we follow the common

approach to include stochastic “active forces” in the equations of motion of the individual particle of interest (see, e.g., Refs. [2,5] and references therein). When calculating the path probabilities, we treat these active nonequilibrium forces in the same way as thermal equilibrium noise, namely, as a “bath” the particle is exposed to with unknown microscopic details but known statistical properties (which are correlated in time and break detailed balance [43]). Our main results following this approach and their implications are briefly summarized in the next section.

B. Brief summary of the main results

Our central findings are the following:

- (i) Using a path-integral approach, we calculate the probability density $\mathfrak{p}[\underline{x}|\underline{x}_0]$ of particle trajectories $\bar{\mathbf{x}} = \{\mathbf{x}(t)\}_{t=0}^\tau = \mathbf{x}_0 \cup \{\mathbf{x}(t)\}_{t>0}^\tau = \mathbf{x}_0 \cup \underline{\mathbf{x}}$ as a result of the statistical properties of thermal and active fluctuations simultaneously affecting the dynamics of the particle [Sec. IV B, Eqs. (38) and (40)]. Here, the coordinate $\mathbf{x}(t)$ may either refer to the position of a passive tracer particle in an active bath or to the position of an active self-propelled particle. When calculating these path weights, we consider the general situation in which the particle is also subject to deterministic external (or interaction) forces $\mathbf{f}_t = \mathbf{f}(\mathbf{x}(t), t)$, which may explicitly depend on time via externally controlled driving protocols. We summarize the mathematical details of the derivation as well as various known limiting cases in Appendixes A–C.
- (ii) We then find [Sec. IV C, Eqs. (42)] that the log ratio of the probabilities for particle trajectories to occur forward in time versus backward in time (time-backward trajectories are marked by a tilde symbol) can be expressed as a functional along the forward path with a nonlocal “memory kernel” $\Gamma_\tau(t, t')$:

$$\frac{\tilde{\mathfrak{p}}[\bar{\mathbf{x}}|\bar{\mathbf{x}}_0]}{\mathfrak{p}[\underline{\mathbf{x}}|\underline{\mathbf{x}}_0]} = e^{-\Delta\Sigma[\bar{\mathbf{x}}]/k_B}, \quad (1a)$$

with

$$\Delta\Sigma[\bar{\mathbf{x}}] = \frac{1}{T} \int_0^\tau dt \int_0^\tau dt' \dot{\mathbf{x}}_t^\top \mathbf{f}_{t'} \left[\delta(t-t') - \frac{D_a}{D} \Gamma_\tau(t, t') \right]. \quad (1b)$$

The functional $\Delta\Sigma$ therefore quantifies irreversibility in our active matter systems. With the explicit form of $\Gamma_\tau(t, t')$ given in Eq. (40), it can be calculated for any particle trajectory $\mathbf{x}(t)$. Information on the thermal and active baths enters into $\Delta\Sigma$ and $\Gamma_\tau(t, t')$ only via their statistical parameters, like the thermal diffusion coefficient D , thermal bath temperature T ,

active diffusion coefficient D_a , and active correlation time τ_a (implicit in Γ_τ). It is, in particular, not necessary to make any assumptions on the behavior of the active fluctuations under time reversal to arrive at the above result for $\Delta\Sigma$.

Combined with the change in system entropy $\Delta S_{\text{sys}} = -k_B \ln[p(\mathbf{x}_\tau, \tau)/p(\mathbf{x}_0, 0)]$, which compares system probabilities at the beginning ($t = 0$) and end ($t = \tau$) of the trajectory [34,38], $\Delta\Sigma + \Delta S_{\text{sys}}$ fulfills an integral fluctuation theorem [Sec. IV C, Eq. (44)] valid for any duration τ of the particle trajectory,

$$\langle e^{-(\Delta\Sigma + \Delta S_{\text{sys}})/k_B} \rangle = 1; \quad (2)$$

the average $\langle \cdot \rangle$ is over all possible trajectories advancing from $t = 0$ to $t = \tau$. As a consequence, they obey a second-law-like relation [Eq. (45)],

$$\langle \Delta\Sigma[\bar{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau) \rangle \geq 0. \quad (3)$$

In case of the equal sign, the active matter system appears reversible on the level of particle trajectories $\mathbf{x}(t)$. From Eq. (1b), we see that in the stationary state (with $\Delta S_{\text{sys}} = 0$), the trajectories for “free active diffusion” without external force ($\mathbf{f}_t \equiv 0$) always look reversible and thus equilibriumlike. In the presence of external forces, we have $\langle \Delta\Sigma + \Delta S_{\text{sys}} \rangle > 0$ in general, where $\langle \Delta\Sigma + \Delta S_{\text{sys}} \rangle$ becomes larger for more irreversible particle dynamics, indicating that the system appears to be further away from an equilibriumlike behavior.

The path probability ratio $\tilde{\mathfrak{p}}[\bar{\mathbf{x}}|\bar{\mathbf{x}}_0]/\mathfrak{p}[\underline{\mathbf{x}}|\underline{\mathbf{x}}_0]$ is an important fundamental concept in stochastic thermodynamics because it contains information on the thermodynamics of the system, such as second-law-like bounds on entropy production as a consequence of the fluctuation theorems, features of heat, work, and efficiency distributions in systems that represent heat engines, or violations of the fluctuation-response relation due to nonequilibrium drivings. In the present work, we derive its most immediate implications, namely, fluctuation theorems and second-law-like relations. Further insights into our active matter system that may be derived in future work from the explicit knowledge of the path probability ratio are briefly discussed in the Conclusions, Sec. VII.

- (iii) By keeping track of the specific realization of the active fluctuations which participated in generating the particle trajectory, we can identify the two individual contributions to $\Delta\Sigma$ stemming from the thermal environment and the active bath. The thermal part is the usual entropy $\Delta S_{\text{tot}}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}]$ produced in the thermal environment and in the system along the particle trajectory (for the given active noise

realization, which we denote $\bar{\boldsymbol{\eta}} = \{\boldsymbol{\eta}(t)\}_{t=0}^{\tau}$. The part associated with the active bath is given by the difference $\Delta\mathcal{I}[\bar{\boldsymbol{x}}, \bar{\boldsymbol{\eta}}]$ in the amount of correlations (measured in terms of pathwise mutual information) that are built up between the particle trajectory and the active noise in the time-forward direction as compared to the time-backward direction (Sec. V). This splitting provides the thermodynamic interpretation

$$\Delta\Sigma[\bar{\boldsymbol{x}}] + \Delta S_{\text{sys}}(\boldsymbol{x}_0, \boldsymbol{x}_{\tau}) = \Delta S_{\text{tot}}[\bar{\boldsymbol{x}}|\bar{\boldsymbol{\eta}}] - k_B \Delta\mathcal{I}[\bar{\boldsymbol{x}}, \bar{\boldsymbol{\eta}}] \quad (4)$$

for our irreversibility measure $\Delta\Sigma$ derived from the path probability ratio of particle trajectories; see Eq. (1). It is valid for any active matter model which includes the active fluctuations as forces in the particle equations of motion, no matter if we assume the active fluctuations to be even [Eq. (58)] or odd [Eq. (64)] under time reversal (denoted later in the text by \pm sub- or superscripts). The interpretation (4) generalizes the trajectorywise stochastic thermodynamics of passive Brownian particles to active matter systems. It shows that entropy production in the thermal bath, which fully quantifies the path probability ratio for passive Brownian motion, is complemented by the difference in mutual information accumulated with the active (nonequilibrium) bath in the time-forward versus the time-backward direction. From the fluctuation theorem for $\Delta\Sigma$, we directly obtain an integral fluctuation theorem and a second-law relation for the mutual information difference $\Delta\mathcal{I}$ [Eqs. (59), (65) and (60), (66)].

We illustrate all these findings in Sec. VI by discussing the example of a colloidal particle subject to thermal and active fluctuations, which is trapped in a (static or moving) harmonic potential. For this simple linear system, all relevant quantities can be calculated explicitly. We give the most important mathematical details in Appendix D.

We finally remark that the explicit expression obtained under (i) and (ii) for $\Gamma_{\tau}(t, t')$ [see Eq. (40)] is valid for a Gaussian Ornstein-Uhlenbeck process [47,48] as a model for the active fluctuations, while result (iii) is valid for general types of active fluctuating forces $\boldsymbol{\eta}$ with arbitrary (but well-defined) statistical properties. The Ornstein-Uhlenbeck process has become quite popular and successful in describing active fluctuations [7,15–18,21–24,28,42,49–61] because it constitutes a minimal model for persistency in the active forcings due to its exponentially decaying correlations with finite correlation time, and it can easily be set up to break detailed balance by a “mismatched” damping term which does not validate the fluctuation-dissipation relation [62,63]. Moreover, it is able to describe both situations of interest mentioned above, namely, passive motion in a bath of active swimmers

[22–24,49] and active motion driven by self-propulsion [7,15–18,21,28,42,50–61]. The details of the model class we consider are given in the next section.

II. MODEL

We consider a colloidal particle at position \boldsymbol{x} in $d = 1, 2,$ or 3 dimensions, which is suspended in an aqueous solution at thermal equilibrium with temperature T . The particle diffuses under the influence of deterministic external forces, in general consisting of potential forces $-\nabla U(\boldsymbol{x}, t)$ [with the potential $U(\boldsymbol{x}, t)$] and nonconservative force components $\boldsymbol{F}(\boldsymbol{x}, t)$. In addition, it experiences fluctuating driving forces due to permanent energy conversion from active processes in the environment or the particle itself. The specific examples we have in mind are a passive tracer in an active nonthermal bath (composed, e.g., of bacteria in aqueous solution; see left panel of Fig. 1) [22–24,49] or a self-propelled particle (e.g., a colloidal microswimmer or bacterium; see right panel of Fig. 1) [7,15–18,21,28,42,50–61].

Neglecting inertia effects [64], we model the overdamped Brownian motion of the colloidal particle by the Langevin equation [47,48,65,66]

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{v}(\boldsymbol{x}(t), t) + \sqrt{2D_a}\boldsymbol{\eta}(t) + \sqrt{2D}\boldsymbol{\xi}(t). \quad (5)$$

The deterministic forces are collected in

$$\boldsymbol{v}(\boldsymbol{x}, t) = \frac{1}{\gamma}\boldsymbol{f}(\boldsymbol{x}, t), \quad (6a)$$

where γ is the hydrodynamic friction coefficient of the particle and

$$\boldsymbol{f}(\boldsymbol{x}, t) = -\nabla U(\boldsymbol{x}, t) + \boldsymbol{F}(\boldsymbol{x}, t). \quad (6b)$$

Thermal fluctuations are described by unbiased Gaussian white-noise sources $\boldsymbol{\xi}(t)$ with mutually independent δ -correlated components $\xi_i(t)$, i.e., $\langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij}\delta(t-t')$, where the angular brackets denote the average over many realizations of the noise and $i, j \in \{1, \dots, d\}$. The strength of the thermal fluctuations is given by the particle’s diffusion coefficient D , which is connected to the temperature T and the friction γ by the fluctuation-dissipation relation $D = k_B T / \gamma$ [62,67–69] (k_B is Boltzmann’s constant), as a consequence of the equilibrium properties of the thermal bath.

The term $\sqrt{2D_a}\boldsymbol{\eta}(t)$ in Eq. (5) represents the active force components with $\boldsymbol{\eta}(t) = [\eta_1(t), \dots, \eta_d(t)]$ being unbiased mutually independent noise processes, and D_a being an effective “active diffusion” characterizing the strength of the active fluctuations. The model (5) does not contain “active friction,” i.e., an integral term over $\dot{\boldsymbol{x}}(t)$ with a friction kernel modeling the damping effects associated with the active forcing. It thus represents the situation of

low activity or particle density, where active fluctuations do not dominate over thermal ones and such friction effects are negligibly small [24,70]. Despite this approximation, the model (5) has been applied successfully to describe various active particle systems [15,17,21–24,49,61] and has become quite popular in an even more simplified variant which neglects thermal fluctuations ($D = 0$) [7,16,18,28,42,50–59]. However, for thermodynamic consistency, it is necessary to consider both noise sources [42], especially when assessing the flow of heat and entropy, which is in general produced in the thermal environment as well as in the processes fueling the active fluctuations [41,71,72].

As active fluctuations are the result of perpetual energy conversion (e.g., by the bacteria in the bath or by the propulsion mechanism of the particle), their salient feature is that they do not fulfill a fluctuation-dissipation relation. In our model (5), this violation is particularly obvious, as active friction effects are absent by the assumption of being negligibly small; in general, the active friction kernel would not match the active noise correlations [73,74]. In that sense, the active fluctuations $\boldsymbol{\eta}(t)$ characterize a nonthermal environment or bath for the Brownian particle with intrinsic nonequilibrium properties. We emphasize that $\boldsymbol{\eta}(t)$ is not a quantity directly measurable, but rather it embodies the net effects of the active components in the environment or the self-propulsion mechanism of the particle, similar to the white noise $\boldsymbol{\xi}(t)$ representing an effective description of the innumerable collisions with the fluid molecules in the aqueous solution.

Although several results in the present paper are generally valid for any noise process $\boldsymbol{\eta}(t)$ (with finite moments), our main aim is to study the specific situation of exponentially correlated Gaussian nonequilibrium fluctuations [28,42,50–53,55,75] generated from an explicitly solvable Ornstein-Uhlenbeck process,

$$\dot{\boldsymbol{\eta}}(t) = -\frac{1}{\tau_a} \boldsymbol{\eta}(t) + \frac{1}{\tau_a} \boldsymbol{\zeta}(t), \quad (7)$$

i.e., the so-called Gaussian colored noise [76]. Here, $\boldsymbol{\zeta}(t) = [\zeta_1(t), \dots, \zeta_d(t)]$ are mutually independent, unbiased, δ -correlated Gaussian noise sources, just like $\boldsymbol{\xi}(t)$, but completely unrelated to them. The characteristic time τ_a quantifies the correlation time of the process ($i, j \in \{1, \dots, d\}$),

$$\langle \eta_i(t) \eta_j(t') \rangle = \frac{\delta_{ij}}{2\tau_a} e^{-|t-t'|/\tau_a}, \quad (8)$$

as can be verified easily from the explicit steady-state solution of Eq. (7). It is thus a measure for the persistence of the active fluctuations.

The model (5), (6) describes a single Brownian particle in simultaneous contact with a thermal bath and an active

environment as a source of nonequilibrium fluctuations $\boldsymbol{\eta}(t)$. Our main aim in this paper is to investigate in detail the role these nonequilibrium fluctuations play for the stochastic energetics [31] and thermodynamics [34] of the Brownian particle. While we adopt the single-particle picture for simplicity, all our results hold for multiple interacting Brownian particles as well. In this case, the symbol \boldsymbol{x} in Eqs. (5) and (6) denotes a supervector collecting the positions \boldsymbol{x}_i ($i = 1, 2, \dots, N$) of all N particles, i.e., $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N)$, and similar for the forces, velocities, and so on. The only requirement is that the particles are identical in the sense that they have the same coupling coefficients D and D_a , and that the active fluctuations $\boldsymbol{\eta}_i(t)$ of the individual constituents are independent but share identical statistical properties.

III. THE IDEAL THERMAL BATH: $D_a = 0$

In order to set up the framework and further establish notation, we start with briefly recalling the well-known case of a Brownian particle in sole contact with a thermal equilibrium reservoir,

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{v}(\boldsymbol{x}(t), t) + \sqrt{2D} \boldsymbol{\xi}(t), \quad (9)$$

where the deterministic driving $\boldsymbol{v}(\boldsymbol{x}(t), t)$ is defined as in Eq. (6). The Brownian particle can be prevented from equilibrating with the thermal bath by time variations in the potential $U(\boldsymbol{x}, t)$ or by nonconservative external forces $\boldsymbol{F}(\boldsymbol{x}, t)$.

A. Energetics

Following Sekimoto [30,31], the heat δQ that the particle exchanges with the thermal bath while moving over an infinitesimal distance $d\boldsymbol{x}(t)$ during a time step dt from t to $t + dt$ is quantified as the energy the thermal bath transfers to the particle along this displacement due to friction $-\gamma \dot{\boldsymbol{x}}(t)$ and fluctuations $\sqrt{2k_B T} \gamma \boldsymbol{\xi}(t)$, i.e.,

$$\delta Q(t) = (-\gamma \dot{\boldsymbol{x}}(t) + \sqrt{2k_B T} \gamma \boldsymbol{\xi}(t)) \cdot d\boldsymbol{x}(t), \quad (10)$$

where the product needs to be interpreted in the Stratonovich sense [77]. With the definition (10), heat is counted as positive if received by the particle and as negative when dumped into the environment (this sign convention is thus the same as Sekimoto's original one [30,31]). From the equation of motion (9), we immediately see that the heat exchange can be equivalently written as

$$\delta Q(t) = -\{-\nabla U(\boldsymbol{x}(t), t) + \boldsymbol{F}(\boldsymbol{x}(t), t)\} \cdot d\boldsymbol{x}(t). \quad (11)$$

The change of the particle's "internal" energy $dU(t)$ over the same displacement $d\boldsymbol{x}(t)$ is given by the total differential

$$dU(t) = \nabla U(\mathbf{x}(t), t) \cdot d\mathbf{x}(t) + \frac{\partial U(\mathbf{x}(t), t)}{\partial t} dt. \quad (12)$$

Hence, if the potential does not vary over time, the only contribution to $dU(t)$ comes from the change in the potential energy associated with the displacement $d\mathbf{x}$.

On the other hand, even if the particle does not move within dt , its internal energy can still change due to variations of the potential landscape by an externally applied time-dependent protocol. Prototype examples are intensity or position variations of optical tweezers, which serve as a trap for the colloidal particle [23,78–81]. Being imposed and controlled externally, these contributions are interpreted as work performed on the particle. A second source of external forces which may contribute to such work are the nonconservative components $\mathbf{F}(\mathbf{x}, t)$ in Eq. (9) (originally not considered by Sekimoto [30] but systematically analyzed later by Speck *et al.* [82]). The total work applied on the particle by external forces is therefore given by

$$\delta W(t) = \frac{\partial U(\mathbf{x}(t), t)}{\partial t} dt + \mathbf{F}(\mathbf{x}(t), t) \cdot d\mathbf{x}(t). \quad (13)$$

Combining Eqs. (11)–(13), we obtain the first law

$$dU = \delta Q + \delta W \quad (14)$$

for the energy balance over infinitesimal displacements $d\mathbf{x}(t)$ valid at any point in time t . After integration along a specific trajectory $\{\mathbf{x}(t)\}_{t=0}^{\tau}$ of duration τ , which starts at $\mathbf{x}(0) = \mathbf{x}_0$ and ends at some $\mathbf{x}(\tau) = \mathbf{x}_\tau$ [which is different for every realization of the thermal noise $\boldsymbol{\xi}(t)$ in Eq. (9), even if $\mathbf{x}(0) = \mathbf{x}_0$ is kept fixed], we find the first law at the trajectory level [35]

$$\Delta U[\bar{\mathbf{x}}] = U(\mathbf{x}_\tau, \tau) - U(\mathbf{x}_0, 0) = Q[\bar{\mathbf{x}}] + W[\bar{\mathbf{x}}]. \quad (15)$$

Here, $\Delta U[\bar{\mathbf{x}}]$, $Q[\bar{\mathbf{x}}]$, and $W[\bar{\mathbf{x}}]$ denote the integrals of the expressions (11) [or, equivalently, Eq. (10)], (12), and (13), respectively, along the trajectory $\bar{\mathbf{x}} = \{\mathbf{x}(t)\}_{t=0}^{\tau}$; the notation $[\bar{\mathbf{x}}]$ explicitly indicates the dependence on that trajectory.

B. Path probability and entropy

The next step towards a thermodynamics characterization of the Brownian motion (9) is to introduce an entropy change or entropy production associated with individual trajectories [38]. From the viewpoint of irreversibility, such an entropy concept has been defined as the log ratio of the probability densities for observing a certain particle trajectory and its time-reversed twin [83]. We express the probability density of a particle trajectory by the standard Onsager-Machlup path integral [84–87],

$$\mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0] \propto \exp \left\{ - \int_0^\tau dt \left[\frac{(\dot{\mathbf{x}}_t - \mathbf{v}_t)^2}{4D} + \frac{\nabla \cdot \mathbf{v}_t}{2} \right] \right\}, \quad (16)$$

where we condition on a fixed starting position \mathbf{x}_0 , and, accordingly, introduce the notation $\underline{\mathbf{x}} = \{\mathbf{x}(t)\}_{t>0}^\tau$ to denote a trajectory which starts at fixed position $\mathbf{x}(0) = \mathbf{x}_0$. In contrast to $\bar{\mathbf{x}}$ from above, the set of points $\underline{\mathbf{x}}$ does therefore not contain the initial point \mathbf{x}_0 (i.e., $\bar{\mathbf{x}} = \mathbf{x}_0 \cup \underline{\mathbf{x}}$). The path probability (16) is to be understood as a product of transition probabilities in the limit of infinitesimal time-step size using a midpoint discretization rule. The divergence of \mathbf{v} in the second term represents the path-dependent part of normalization [86,87], while all remaining normalization factors are path-independent constants and thus omitted. For convenience, we introduce the shorthand notation $\mathbf{x}_t = \mathbf{x}(t)$ and $\mathbf{v}_t = \mathbf{v}(\mathbf{x}(t), t)$.

We now consider the time-reversed version $\tilde{\mathbf{x}}(t)$ of this trajectory, which is traced out backward from the final point $\mathbf{x}(\tau) = \mathbf{x}_\tau$ to the initial point \mathbf{x}_0 when advancing time,

$$\tilde{\mathbf{x}}(t) = \mathbf{x}(\tau - t), \quad (17)$$

and ask how likely it is that $\tilde{\mathbf{x}} = \{\tilde{\mathbf{x}}(t)\}_{t>0}^\tau$ is generated by the same Langevin equation (9) as $\underline{\mathbf{x}}$, with the same deterministic forces \mathbf{v} acting at identical positions along the path. The latter requirement implies that in case of explicitly time-dependent external forces, the force protocol has to be time inverted; i.e., $\mathbf{v}(\mathbf{x}, t)$ is replaced by $\mathbf{v}(\mathbf{x}, \tau - t)$ in Eq. (9) to construct the Langevin equation for the time-reversed path [88]. From that Langevin equation, we can deduce the probability $\tilde{\mathfrak{p}}[\tilde{\mathbf{x}}|\tilde{\mathbf{x}}_0]$ for observing the backward trajectory $\tilde{\mathbf{x}}$ conditioned on its initial position $\tilde{\mathbf{x}}_0 = \mathbf{x}(\tau) = \mathbf{x}_\tau$, in analogy to Eq. (16). Using Eq. (17), we can then express $\tilde{\mathfrak{p}}[\tilde{\mathbf{x}}|\tilde{\mathbf{x}}_0]$ in terms of the forward path $\underline{\mathbf{x}}$ so that we find for the path probability ratio

$$\frac{\tilde{\mathfrak{p}}[\tilde{\mathbf{x}}|\tilde{\mathbf{x}}_0]}{\mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0]} = e^{-\Delta S[\bar{\mathbf{x}}]/k_B}, \quad (18)$$

with the quantity $\Delta S[\bar{\mathbf{x}}]$ being a functional of the forward path only,

$$\begin{aligned} \Delta S[\bar{\mathbf{x}}] &= \frac{1}{T} \int_0^\tau dt \dot{\mathbf{x}}(t) \cdot \mathbf{f}(\mathbf{x}(t), t) \\ &= \frac{1}{T} \int_0^\tau \mathbf{f}(\mathbf{x}(t), t) \cdot d\mathbf{x}(t). \end{aligned} \quad (19)$$

As a quantitative measure of irreversibility, $\Delta S[\bar{\mathbf{x}}]$ is identified with the entropy production along the path $\underline{\mathbf{x}}$ with given initial position \mathbf{x}_0 .

For an infinitesimal displacement $d\mathbf{x}(t)$, the corresponding entropy change reads

$$\delta S(t) = \frac{1}{T} \mathbf{f}(\mathbf{x}(t), t) \cdot d\mathbf{x}(t) = -\frac{\delta Q(t)}{T}. \quad (20)$$

The last equality follows from comparison with Eq. (11) [see also Eq. (6)] and states that the entropy production along $d\mathbf{x}(t)$ is given by the heat $-\delta Q(t)$ dissipated into the environment during that step divided by the bath temperature. For that reason, $\delta S(t)$ and $\Delta S[\bar{\mathbf{x}}]$ are more accurately called entropy production in the environment.

The entropy of the Brownian particle itself (i.e., the entropy associated with the system degrees of freedom (d.o.f.) \mathbf{x}) is defined as the state function [34,38]

$$S_{\text{sys}}(\mathbf{x}, t) = -k_B \ln p(\mathbf{x}, t), \quad (21)$$

where $p(\mathbf{x}, t)$ is the time-dependent solution of the Fokker-Planck equation [47,48,66] associated with Eq. (9), for the same initial distribution $p_0(\mathbf{x}_0) = p(\mathbf{x}_0, 0)$ from which the initial value \mathbf{x}_0 for the path $\underline{\mathbf{x}}$ is drawn. The change in system entropy along the trajectory is therefore given by

$$\Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau) = -k_B \ln p(\mathbf{x}_\tau, \tau) + k_B \ln p(\mathbf{x}_0, 0). \quad (22)$$

Combining $\Delta S[\bar{\mathbf{x}}]$ and $\Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau)$, we define the total entropy production along a trajectory $\bar{\mathbf{x}}$ [38],

$$\Delta S_{\text{tot}}[\bar{\mathbf{x}}] = \Delta S[\bar{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau). \quad (23)$$

It fulfills the integral fluctuation theorem [32,34]

$$\langle e^{-\Delta S_{\text{tot}}/k_B} \rangle = 1 \quad (24)$$

as a direct consequence of Eq. (18) [see also Eq. (22)]. The average in Eq. (24) is over all trajectories with a given but arbitrary distribution $p_0(\mathbf{x}_0)$ of initial values \mathbf{x}_0 [38].

IV. THE NONEQUILIBRIUM ENVIRONMENT

We now focus on the full model (5)–(7) for Brownian motion subject to active fluctuations $\boldsymbol{\eta}(t)$. Our main goal is to develop a trajectorywise thermodynamic description as a natural generalization of stochastic energetics and thermodynamics in a purely thermal environment (see previous section). We thus treat the active forces $\boldsymbol{\eta}(t)$ in the same way as the thermal noise $\boldsymbol{\xi}(t)$, namely, as a source of fluctuations whose specific realizations are not accessible but whose statistical properties determine the probability $\mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0]$ for observing a certain particle trajectory $\underline{\mathbf{x}}$. We are interested in how the nonequilibrium characteristics of the active fluctuations affect the irreversibility measure encoded in the ratio between forward and backward path probabilities and how this measure is connected to the energetics of the active Brownian motion.

A. Energetics and entropy

Comparing Eq. (5) with Eq. (9), we may conclude that the stochastic energetics associated with Eq. (5) can be obtained from the energetics for Eq. (9) by adjusting the

total forces acting on the Brownian particle, i.e., by replacing $\mathbf{v}(\mathbf{x}, t)$ with $\mathbf{v}(\mathbf{x}, t) + \sqrt{2D_a}\boldsymbol{\eta}(t)$. However, there is another, maybe less obvious way of turning Eq. (9) into the model (5) with active fluctuations, namely, by substituting $\dot{\mathbf{x}}$ with $\dot{\mathbf{x}} - \sqrt{2D_a}\boldsymbol{\eta}(t)$. In the following, we argue that these two approaches correspond to two different physical situations (depicted in Fig. 1) with different trajectorywise energy balances.

1. Active bath

Examples for an “active bath” are the intracellular matrix of a living cell [89–95] or an aqueous suspension containing swimming bacteria [5,23]. These active environments then drive a passive Brownian particle, e.g., a colloid, via interactions or “collisions,” which are included in the equation of motion for the passive tracer as effective nonequilibrium fluctuations $\boldsymbol{\eta}(t)$. In this case, the fluctuations $\boldsymbol{\eta}(t)$ can indeed be interpreted as additional time-dependent forces from sources external to the Brownian particle. Hence, the total external force acting on the Brownian particle at time t is given by $\gamma\{\mathbf{v}(\mathbf{x}(t), t) + \sqrt{2D_a}\boldsymbol{\eta}(t)\}$. This modification of the external force obviously does not affect the basic definition (10) of heat exchanged with the thermal bath. Using the force balance expressed in the Langevin equation (5) to replace $-\gamma\dot{\mathbf{x}}(t) + \gamma\sqrt{2D}\boldsymbol{\xi}(t)$, we obtain

$$\delta Q_+(t) = -[\mathbf{f}(\mathbf{x}(t), t) + \sqrt{2D_a}\boldsymbol{\eta}(t)] \cdot d\mathbf{x}(t). \quad (25)$$

The active fluctuations $\boldsymbol{\eta}(t)$ formally play the role of additional nonconservative force components and thus affect the heat which is dissipated into the thermal environment in order to balance all acting external forces. However, they cannot be controlled to perform work on the particle due to their inherent fluctuating character as an active bath, such that the definition (13) of the work remains unchanged. Indeed, this standard definition of work from stochastic thermodynamics has been used to analyze a microscopic heat engine operating between two active baths [96]. Finally, the change of internal energy (12) over a time interval dt is determined by the potential $U(\mathbf{x}, t)$ only and thus is not altered by the presence of the active fluctuations $\boldsymbol{\eta}(t)$ either.

Combining Eqs. (12), (13), and (25), we find the energy balance (first law)

$$dU = \delta Q_+ + \delta W + \delta A_+ \quad (26)$$

for Brownian motion in an active bath. Here, we introduce the energy exchanged with the active bath,

$$\delta A_+ = \sqrt{2D_a}\boldsymbol{\eta}(t) \cdot d\mathbf{x}(t). \quad (27)$$

It might be best interpreted as “heat” in the sense of Sekimoto’s general definition, that any energy exchange

with unknown or inaccessible d.o.f. may be identified as heat [31]. In our setup, the active fluctuations $\boldsymbol{\eta}(t)$ represent an effective description of the forces from the active environment and are thus in general not directly measurable in an experiment.

Based on the heat the Brownian particle exchanges with the thermal environment, we can identify the entropy production in the thermal environment as

$$\begin{aligned}\delta S_+(t) &= -\frac{\delta Q_+(t)}{T} \\ &= \frac{1}{T}[\mathbf{f}(\mathbf{x}(t), t) + \sqrt{2D_a}\boldsymbol{\eta}(t)] \cdot d\mathbf{x}(t),\end{aligned}\quad (28)$$

in analogy to Eq. (20). We refrain, however, from defining an entropy production in the active bath. Such an environment, itself being in a nonequilibrium state due to continuous dissipation of energy, does not possess a well-defined entropy so that the heat dissipated into the active bath cannot be associated with a change of bath entropy.

2. Self-propulsion

Self-propulsion occurs if a particle is able to locally convert external fuel on its own into directed motion; such a fuel may, e.g., be nutrients for a bacterium or some kind of chemical components like hydrogen peroxide for a colloidal Janus particle; see Fig. 1 in the review [5] for an extensive collection of self-propelled (Brownian) particles. This self-propulsion entails that incessant particle motion can occur already without any external forces being applied. It is represented by the active fluctuations $\boldsymbol{\eta}(t)$ in the equations of motion (5) for the self-propelled Brownian particle. For $\mathbf{f}(\mathbf{x}, t) \equiv 0$ and without thermal fluctuations $\sqrt{2D}\boldsymbol{\xi}(t) \equiv 0$, the momentary particle velocity $\dot{\mathbf{x}}(t)$ is exactly equal to the active velocity $\sqrt{2D_a}\boldsymbol{\eta}(t)$. In that sense, self-propulsion is force-free. More precisely, the driving force, which is created locally by the particle for self-propulsion, is compensated according to *actio est reactio*, such that the total force acting on a fluid volume comprising the particle and its active self-propulsion mechanism is zero. The corresponding dissipation (and entropy production) inside such a fluid volume which results from the conversion of energy or fuel to generate the self-propulsion drive cannot be quantified by our effective description of the active propulsion as fluctuating forces $\boldsymbol{\eta}$ because it does not contain any information on the underlying microscopic processes. In order to quantify such an entropy production, a specific model for the self-propulsion mechanism is required [41,71,72]. In other words, our ‘‘coarse-grained’’ description (7) does not allow us to assess how much energy the conversion process behind the self-propulsion drive dissipates.

On the other hand, if the force-free motion of the self-propelled particle is disturbed by the presence of external forces $\mathbf{f}(\mathbf{x}, t) \neq 0$ or thermal fluctuations $\sqrt{2D}\boldsymbol{\xi}(t)$, its

actual velocity deviates from the self-propulsion velocity $\sqrt{2D_a}\boldsymbol{\eta}(t)$, resulting in hydrodynamic friction $-\gamma[\dot{\mathbf{x}}(t) - \sqrt{2D_a}\boldsymbol{\eta}(t)]$ between the fluid volume comprising the particle and its active self-propulsion mechanism on the one hand and the surrounding fluid environment on the other hand. Together with the thermal fluctuations $\sqrt{2D}\boldsymbol{\xi}(t)$, these friction forces are the forces exchanged with the thermal environment, entailing an energy transfer between particle and thermal bath. It is this dissipation associated with the deviations of the particle trajectory from the self-propulsion path which we can quantify. The corresponding heat exchange with the thermal bath for a displacement $d\mathbf{x}(t)$ taking place over a time interval dt at time t is given by $(-\gamma[\dot{\mathbf{x}}(t) - \sqrt{2D_a}\boldsymbol{\eta}(t)] + \sqrt{2D}\boldsymbol{\xi}(t)) \cdot [d\mathbf{x}(t) - \sqrt{2D_a}\boldsymbol{\eta}(t)dt]$, or, using that hydrodynamic friction and thermal fluctuations are exactly balanced by the external force \mathbf{f} [see Eq. (5)],

$$\delta Q_-(t) = -\mathbf{f}(\mathbf{x}(t), t) \cdot (d\mathbf{x}(t) - \sqrt{2D_a}\boldsymbol{\eta}(t)dt). \quad (29)$$

This definition of heat is equivalent to the one suggested in Ref. [72] for active particles which are self-propelled by chemical conversions.

The work $\delta W(t)$, on the other hand, performed on or by the particle during the time step dt , which can be controlled or harvested by an external agent, is exactly the same as without active propulsion given in Eq. (13) because from the operational viewpoint of the external agent, it is irrelevant what kind of mechanisms propel the particle. Accordingly, it has been used in Ref. [97] for connecting work in active particle systems to pressure and in Ref. [98] for measuring the work performed by a colloidal heat engine with an active particle as the working medium. Finally, the definition (12) for the change in internal energy $dU(t)$ is independent of how particle motion is driven as well and thus remains unaffected by our interpretation of $\sqrt{2D_a}\boldsymbol{\eta}(t)$ as active propulsion.

Combining Eqs. (12), (13), and (29), we find the first-law-like relation

$$dU = \delta Q_- + \delta W + \delta A_- \quad (30)$$

for active self-propelled Brownian motion described by the Langevin equation (5). Here, we balance the different energetic contributions by introducing

$$\delta A_-(t) = -\mathbf{f}(\mathbf{x}(t), t) \cdot \sqrt{2D_a}\boldsymbol{\eta}(t)dt. \quad (31)$$

This quantity represents those contributions to the heat exchange which are contained only in the active component $\sqrt{2D_a}\boldsymbol{\eta}(t)dt$ of the full particle displacement $d\mathbf{x}(t)$; see Eq. (29). We can therefore interpret it as the heat transferred from the active fluctuations to the thermal bath via the Brownian particle. Again, $\delta A_-(t)$ is a quantity which, in

general, cannot be measured in an experiment, since it is produced by the inaccessible active fluctuations $\boldsymbol{\eta}(t)$.

Finally, we can relate the heat exchange δQ_- with the thermal bath to dissipation and define a corresponding entropy production in the environment,

$$\begin{aligned} \delta S_-(t) &= -\frac{\delta Q_-(t)}{T} \\ &= \frac{1}{T} \mathbf{f}(\mathbf{x}(t), t) \cdot (d\mathbf{x}(t) - \sqrt{2D_a} \boldsymbol{\eta}(t) dt). \end{aligned} \quad (32)$$

B. Path probability

We now calculate the probability $\mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0]$ for observing a certain path $\underline{\mathbf{x}} = \{\mathbf{x}(t)\}_{t>0}^\tau$ starting at \mathbf{x}_0 generated under the combined influence of thermal and active fluctuations. We treat these two noise sources on equal terms, namely, as fluctuating forces with unknown specific realizations but known statistical properties. Because of the memory of the active noise $\boldsymbol{\eta}(t)$, the system (5) is non-Markovian. Therefore, the standard Onsager-Machlup path integral [84–86] cannot be applied directly to obtain $\mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0]$. However, for the Ornstein-Uhlenbeck model (7) of $\boldsymbol{\eta}(t)$, the combined set of variables $(\mathbf{x}, \boldsymbol{\eta})$ is Markovian, and we can easily write down the path probability $\mathfrak{p}[\underline{\mathbf{x}}, \underline{\boldsymbol{\eta}}|\mathbf{x}_0, \boldsymbol{\eta}_0]$ for the joint trajectory $(\underline{\mathbf{x}}, \underline{\boldsymbol{\eta}}) = \{\{\mathbf{x}(t), \boldsymbol{\eta}(t)\}\}_{t>0}^\tau$ conditioned on an initial configuration $(\mathbf{x}_0, \boldsymbol{\eta}_0)$:

$$\begin{aligned} \mathfrak{p}[\underline{\mathbf{x}}, \underline{\boldsymbol{\eta}}|\mathbf{x}_0, \boldsymbol{\eta}_0] &\propto \exp \left\{ -\int_0^\tau dt \left[\frac{(\dot{\mathbf{x}}_t - \mathbf{v}_t - \sqrt{2D_a} \boldsymbol{\eta}_t)^2}{4D} \right. \right. \\ &\quad \left. \left. + \frac{(\tau_a \dot{\boldsymbol{\eta}}_t + \boldsymbol{\eta}_t)^2}{2} + \frac{\mathbf{\nabla} \cdot \mathbf{v}_t}{2} \right] \right\}. \end{aligned} \quad (33)$$

To calculate the path weight for the particle trajectories $\mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0]$, we have to integrate out the active noise history,

$$\mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0] = \int \mathcal{D}\underline{\boldsymbol{\eta}} d\boldsymbol{\eta}_0 \mathfrak{p}[\underline{\mathbf{x}}, \underline{\boldsymbol{\eta}}|\mathbf{x}_0, \boldsymbol{\eta}_0] p_0(\boldsymbol{\eta}_0|\mathbf{x}_0), \quad (34)$$

where $p_0(\boldsymbol{\eta}_0|\mathbf{x}_0)$ characterizes the initial distribution of the active fluctuations at $t = 0$.

Since the variables $\boldsymbol{\eta}(t)$ represent an effective description of the active fluctuations which are not experimentally accessible, in general it is not possible to set up a specific initial state for $\boldsymbol{\eta}(t)$. As a consequence, there are basically two physically reasonable choices for $p_0(\boldsymbol{\eta}_0|\mathbf{x}_0)$.

On the one hand, we can assume that the active bath has reached its stationary state before we immerse the Brownian particle, such that the bath's initial distribution is *independent* of \mathbf{x}_0 and given by the stationary distribution of $\boldsymbol{\eta}(t)$, i.e., $p_0(\boldsymbol{\eta}_0|\mathbf{x}_0) = p_s(\boldsymbol{\eta}_0)$. For the Ornstein-Uhlenbeck process (7), the stationary distribution reads [47,99]

$$p_s(\boldsymbol{\eta}_0) = \sqrt{\frac{\tau_a}{\pi}} e^{-\tau_a \boldsymbol{\eta}_0^2}. \quad (35)$$

At $t = 0$, we place the Brownian particle into the fluid with an initial distribution $p_0(\mathbf{x}_0)$ of particle positions which can be prepared arbitrarily and start measuring immediately.

On the other hand, we may let the Brownian particle adapt to the active and thermal environments before performing measurements and assume that the system $(\mathbf{x}_0, \boldsymbol{\eta}_0)$ is in a *joint* steady state $p_s(\mathbf{x}_0, \boldsymbol{\eta}_0) = p_0(\boldsymbol{\eta}_0|\mathbf{x}_0) p_0(\mathbf{x}_0)$ at $t = 0$. In that case, control over the distribution $p_0(\mathbf{x}_0)$ of initial particle positions is limited, as it is influenced by the active fluctuations. The form of $p_s(\mathbf{x}_0, \boldsymbol{\eta}_0)$ depends on the particular setup, i.e., the specific choices for $U(\mathbf{x}, t)$ and $\mathbf{F}(\mathbf{x}, t)$ in Eq. (6).

In the following, we perform the path integration (34) for the first option, starting from independent initial conditions $p_0(\boldsymbol{\eta}_0|\mathbf{x}_0) = p_s(\boldsymbol{\eta}_0)$; the calculation for the second option can be carried out along the same lines if the joint steady state $p_s(\mathbf{x}_0, \boldsymbol{\eta}_0)$ is Gaussian in the active noise $\boldsymbol{\eta}_0$ [100]. Plugging Eqs. (33) and (35) into Eq. (34) and performing a partial integration of the term proportional to $\dot{\boldsymbol{\eta}}_t^2$ in the exponent, we obtain

$$\begin{aligned} \mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0] &\propto \exp \left\{ \int_0^\tau dt \left[-\frac{(\dot{\mathbf{x}}_t - \mathbf{v}_t)^2}{4D} - \frac{\mathbf{\nabla} \cdot \mathbf{v}_t}{2} \right] \right\} \\ &\quad \times \int \mathcal{D}\bar{\boldsymbol{\eta}} \exp \left\{ \int_0^\tau dt \frac{\sqrt{2D_a}}{2D} \bar{\boldsymbol{\eta}}_t^\top (\dot{\mathbf{x}}_t - \mathbf{v}_t) \right. \\ &\quad \left. - \frac{1}{2} \int_0^\tau dt \int_0^\tau dt' \bar{\boldsymbol{\eta}}_t^\top \hat{V}_\tau(t, t') \bar{\boldsymbol{\eta}}_{t'} \right\}, \end{aligned} \quad (36)$$

where we use the abbreviation $\bar{\boldsymbol{\eta}} = \boldsymbol{\eta}_0 \cup \underline{\boldsymbol{\eta}}$ for the full path including the initial point $\boldsymbol{\eta}_0$ in order to write $\mathcal{D}\underline{\boldsymbol{\eta}} d\boldsymbol{\eta}_0 = \mathcal{D}\bar{\boldsymbol{\eta}}$. The differential operator

$$\hat{V}_\tau(t, t') = \delta(t - t') [\hat{V}(t) + \hat{V}_0(t) + \hat{V}_\tau(t)] \quad (37a)$$

consists of an ordinary component

$$\hat{V}(t) = -\tau_a^2 \partial_t^2 + (1 + D_a/D) \quad (37b)$$

and two boundary components

$$\hat{V}_0(t) = \delta(t) (\tau_a - \tau_a^2 \partial_t), \quad (37c)$$

$$\hat{V}_\tau(t) = \delta(t - \tau) (\tau_a + \tau_a^2 \partial_t), \quad (37d)$$

which include the boundary terms picked up by the partial integration of $\dot{\boldsymbol{\eta}}_t^2$ and from the initial distribution (35) of $\boldsymbol{\eta}_0$. The subscript τ in Eq. (37a) indicates that the operator \hat{V}_τ is acting on trajectories of duration τ .

Since the path integral over the active noise histories $\boldsymbol{\eta}(t)$ is Gaussian, we can perform it exactly [102]. We find

$$\begin{aligned} \mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0] \propto \exp \left\{ -\frac{1}{4D} \int_0^\tau dt \int_0^\tau dt' (\dot{\mathbf{x}}_t - \mathbf{v}_t)^\top \right. \\ \times \left[\delta(t-t') - \frac{D_a}{D} \Gamma_\tau(t, t') \right] (\dot{\mathbf{x}}_{t'} - \mathbf{v}_{t'}) \\ \left. - \frac{1}{2} \int_0^\tau dt \nabla \cdot \mathbf{v}_t \right\}, \end{aligned} \quad (38)$$

where $\Gamma_\tau(t, t')$ denotes the operator inverse or Green's function of $\hat{V}_\tau(t, t')$ in the sense that

$$\int_0^\tau dt' \hat{V}_\tau(t, t') \Gamma_\tau(t', t'') = \delta(t-t''). \quad (39)$$

Roughly speaking, this Gaussian integration can be understood by thinking of \hat{V}_τ and Γ_τ as matrices with continuous

indices t, t' . The path integral (36) is then a continuum generalization of an ordinary Gaussian integral for finite-dimensional matrices and can be performed by “completing the square.” We provide a rigorous derivation of Eqs. (38) and (39) in Appendix A.

In order to obtain the explicit form of the Green's function $\Gamma_\tau(t, t')$, we need to solve the integrodifferential equation (39). In our case, the operator $\hat{V}_\tau(t, t')$ is proportional to $\delta(t-t')$ [see Eq. (37)] and thus has a “diagonal” structure, such that Eq. (39) turns into an ordinary linear differential equation. We can solve it by following standard methods [103,104]; details are given in Appendix B. We obtain

$$\Gamma_\tau(t, t') = \left(\frac{1}{2\tau_a^2 \lambda} \right) \frac{\kappa_+^2 e^{-\lambda|t-t'|} + \kappa_-^2 e^{-\lambda(2\tau-|t-t'|)} - \kappa_+ \kappa_- [e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}]}{\kappa_+^2 - \kappa_-^2 e^{-2\lambda\tau}}, \quad (40a)$$

with

$$\lambda = \frac{1}{\tau_a} \sqrt{1 + D_a/D}, \quad (40b)$$

and

$$\kappa_\pm = 1 \pm \lambda \tau_a = 1 \pm \sqrt{1 + D_a/D}. \quad (40c)$$

With this expression for $\Gamma_\tau(t, t')$, Eq. (38) represents the exact path probability density for the dynamics of the Brownian particle (5) under the influence of active Ornstein-Uhlenbeck fluctuations (7). This is our first main result. We see that the active fluctuations with their colored noise character lead to correlations in the path weight via the memory kernel $\Gamma_\tau(t, t')$. They relate trajectory points at different times by an exponential weight factor similar to the active noise correlation function (8) but with a correlation time which is a factor of $(1 + D_a/D)^{-1/2}$ smaller. We emphasize again that we assumed independent initial conditions for the particle's position \mathbf{x}_0 and the active bath variables $\boldsymbol{\eta}_0$, $p_0(\boldsymbol{\eta}_0|\mathbf{x}_0) = p_s(\boldsymbol{\eta}_0)$ [see also the discussion around Eq. (35)]. A different choice for the initial distribution, for instance, the joint stationary state for \mathbf{x}_0 and $\boldsymbol{\eta}_0$, would result in a modified $\Gamma_\tau(t, t')$, whose precise form in general depends on the specific implementation of the deterministic forces \mathbf{f} . The correlations in the path weight we measure via the memory kernel $\Gamma_\tau(t, t')$ are thus influenced by our choice of the time instance at which we start observing the particle trajectory. In that sense, the system “remembers its past” even prior to the initial time point $t=0$ because of the finite correlation time in the active fluctuations.

We finally remark that our general expression (38) with (40a) reduces to known results in the three limiting cases $D_a \rightarrow 0$ (passive particle), $\tau_a \rightarrow 0$ (white active noise), and $D \rightarrow 0$ (no thermal bath); details of the calculations can be found in Appendix C. Without active fluctuations ($D_a \rightarrow 0$), we trivially recover the standard Onsager-Machlup expression (16) for passive Brownian motion. In the white-noise limit for $\boldsymbol{\eta}(t)$ ($\tau_a \rightarrow 0$), the equation of motion (5) involves two independent Gaussian white-noise processes with vanishing means and variances $2D$ and $2D_a$, respectively. Their sum is itself a white-noise source with zero mean but variance $2(D + D_a)$. Accordingly, as $\tau_a \rightarrow 0$, we obtain from Eq. (38) an Onsager-Machlup path weight of the form (16) but with the diffusion coefficient D being replaced by $D + D_a$. In the third limiting case of vanishing thermal fluctuations, we are left with a pure colored noise path weight [105,106],

$$\begin{aligned} \mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0]_{D \rightarrow 0} \propto \exp \left\{ -\int_0^\tau dt \frac{[(1 + \tau_a \partial_t)(\dot{\mathbf{x}}_t - \mathbf{v}_t)]^2}{4D_a} - \frac{\nabla \cdot \mathbf{v}_t}{2} \right\} \\ \times p_s \left(\frac{\dot{\mathbf{x}}_0 - \mathbf{v}_0}{\sqrt{2D_a}} \right), \end{aligned} \quad (41)$$

where p_s denotes the steady-state distribution (35) of the colored noise.

C. Fluctuation theorem

With the explicit form (38) of the path probability, we can now derive an exact expression for the log ratio of path probabilities which relates time-forward and -backward paths. Following exactly the same line of reasoning as described in Sec. III B for the case of a passive Brownian particle, we consider the ratio of probabilities for observing

a specific trajectory \underline{x} and its time-reversed twin $\tilde{\underline{x}}$ created under a time-reversed protocol $\mathbf{v}(\mathbf{x}, \tau - t)$ [88]. With the definition (17) of the time-reversed trajectory, we can express its probability in terms of the forward path. Using the property $\Gamma_\tau(\tau - t, \tau - t') = \Gamma_\tau(t, t')$ of the memory kernel [see Eq. (40a)], we then obtain the path probability ratio

$$\frac{\tilde{\mathbf{p}}[\tilde{\underline{x}}|\tilde{\mathbf{x}}_0]}{\mathbf{p}[\underline{x}|\mathbf{x}_0]} = e^{-\Delta\Sigma[\tilde{\underline{x}}]/k_B}, \quad (42a)$$

with

$$\Delta\Sigma[\tilde{\underline{x}}] = \frac{1}{T} \int_0^\tau dt \int_0^\tau dt' \dot{\mathbf{x}}_t^\top \mathbf{f}_{t'} \left[\delta(t - t') - \frac{D_a}{D} \Gamma_\tau(t, t') \right]. \quad (42b)$$

As a stochastic integral, this expression is to be understood in the Stratonovich sense. Note that the procedure of time inversion does not involve the active fluctuations $\boldsymbol{\eta}(t)$ in any way, and thus does not require any assumptions on their properties under time reversal. In fact, the probability density $\mathbf{p}[\underline{x}|\mathbf{x}_0]$ for the trajectories of the Brownian particle is a result of integrating over *all* possible realizations $\boldsymbol{\eta}(t)$ of the active fluctuations containing any pair of conceivable time-forward and time-backward twins with their natural weight of occurrence [see also Eq. (33)]. Hence, for the probability ratio of particle trajectories (38), the behavior of the active fluctuations $\boldsymbol{\eta}(t)$ under time inversion is irrelevant.

As we describe in Sec. III B, a relation like Eq. (42) based on path probability ratios entails an integral fluctuation theorem if the entropy production in the system $\Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau) = -k_B \ln p(\mathbf{x}_\tau, \tau) + k_B \ln p(\mathbf{x}_0, 0)$ [see Eq. (22)] is taken into consideration. Explicitly, we find

$$\frac{\tilde{\mathbf{p}}[\tilde{\underline{x}}]}{\mathbf{p}[\tilde{\underline{x}}]} = e^{-[\Delta\Sigma[\tilde{\underline{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau)]/k_B}, \quad (43)$$

and therefore,

$$\langle e^{-(\Delta\Sigma + \Delta S_{\text{sys}})/k_B} \rangle = 1. \quad (44)$$

By Jensen's inequality, we conclude

$$\langle \Delta\Sigma + \Delta S_{\text{sys}} \rangle \geq 0. \quad (45)$$

The fluctuation theorem (44) and the bound (45) are a direct consequence of the definition of $\Delta\Sigma + \Delta S_{\text{sys}}$ as the logarithm of a path probability ratio (43). Obviously, formally equivalent relations are valid for any quantity defined in this way. Here these relations (43)–(45) become useful because our analysis equips us with an explicit expression for $\Delta\Sigma$ as a functional of the particle trajectory in forward time.

In Eq. (45), equality is achieved if and only if the dynamics is symmetric under time reversal. The setting considered here, however, is generally not symmetric under time reversal because of our choice of particle position and active fluctuations being independent initially. The approach to a correlated (stationary) state is irreversible, such that we inevitably pick up transient contributions to $\Delta\Sigma + \Delta S_{\text{sys}}$, which are strictly positive on average, even if the external forces \mathbf{f} are time independent and conservative.

For the same reason, namely, the build up of correlations between the particle trajectory $\mathbf{x}(t)$ and the active fluctuation $\boldsymbol{\eta}(t)$, the quantity $\Delta\Sigma$ is nonadditive, i.e., $\Delta\Sigma[\tilde{\underline{x}}] \neq \Delta\Sigma[\{\mathbf{x}(t)\}_{t=0}^{t'}] + \Delta\Sigma[\{\mathbf{x}(t)\}_{t=t'}^\tau]$ in general, for any intermediate time $0 < t' < \tau$. In order to establish such an additivity property, we have to take into account the correlations between $\mathbf{x}(t)$ and $\boldsymbol{\eta}(t)$ in the initial distribution $p_{t'}(\boldsymbol{\eta}_{t'}|\mathbf{x}_{t'})$ for the path integral (34) of the second part of the trajectory $\mathbf{x}(t)$ with $t' < t < \tau$ that have been built up until the time point t' . However, it is not possible to specify this conditional distribution in the general situation considered here allowing for arbitrary forces and force protocols $\mathbf{f} = \mathbf{f}(\mathbf{x}, t)$; see Eq. (6b) and the remark below Eq. (35).

The path probability ratio (42a) relating the time reversibility of trajectories to the extensive quantity $\Delta\Sigma[\tilde{\underline{x}}]$, together with its corresponding integral fluctuation theorem (44), constitute our second main result. The interpretation of $\Delta\Sigma$ and its fluctuation theorems in physical terms is, however, not as straightforward as in the case of a passive Brownian particle in which we could identify the logarithm of the path probability ratio as the heat dissipated into the thermal environment divided by the bath temperature [34,38]; see Eq. (20). Although we cannot make such a simple identification for the Brownian particle driven by active fluctuations, it is still possible to connect $\Delta\Sigma$ to physically meaningful quantities, as we show in the following.

As a first step in this direction, we observe a formal similarity between ΔS given in Eq. (19) and $\Delta\Sigma$ given in Eq. (42b). Defining the nonlocal ‘‘force’’

$$\boldsymbol{\varphi}_\tau[\tilde{\underline{x}}, t] = \mathbf{f}(\mathbf{x}(t), t) - \frac{D_a}{D} \int_0^\tau dt' \mathbf{f}(\mathbf{x}(t'), t') \Gamma_\tau(t, t'), \quad (46)$$

we can bring Eq. (42b) into the form

$$\begin{aligned} \Delta\Sigma[\tilde{\underline{x}}] &= \frac{1}{T} \int_0^\tau dt \dot{\mathbf{x}}(t) \cdot \boldsymbol{\varphi}_\tau[\tilde{\underline{x}}, t] \\ &= \frac{1}{T} \int_0^\tau d\mathbf{x}(t) \cdot \boldsymbol{\varphi}_\tau[\tilde{\underline{x}}, t], \end{aligned} \quad (47)$$

in obvious analogy to Eq. (19) but with the essential difference that the ‘‘force’’ $\boldsymbol{\varphi}_\tau[\tilde{\underline{x}}, t]$ at time $0 \leq t \leq \tau$ depends not only on $\mathbf{x}(t)$ but rather on the full trajectory $\tilde{\underline{x}} = \{\mathbf{x}(t)\}_{t=0}^\tau$ via the memory term $\int_0^\tau dt' \mathbf{f}(\mathbf{x}(t'), t') \Gamma_\tau(t, t')$.

Similar “memory forces” have already been found to affect irreversibility by contributing to dissipation in Langevin systems with colored noise, which does not obey the fluctuation-dissipation relation [107].

From Eq. (47), we can read off a production rate for Σ ,

$$\sigma_\tau(t) = \dot{\mathbf{x}}(t) \cdot \frac{1}{T} \boldsymbol{\varphi}_\tau[\bar{\mathbf{x}}, t] \quad (48a)$$

or

$$\delta\Sigma_\tau(t) = d\mathbf{x}(t) \cdot \frac{1}{T} \boldsymbol{\varphi}_\tau[\bar{\mathbf{x}}, t]. \quad (48b)$$

Even though this expression for the Σ production is analogous to the actual entropy productions in the environment as identified in Eqs. (28) or (32) [see also Eq. (20)] and even contains a term $\{d\mathbf{x}(t) \cdot \mathbf{f}(\mathbf{x}(t), t)\}/T$, which quantifies dissipation due to the external force \mathbf{f} , its physical meaning beyond formally defined memory forces is unclear. In the following, we argue that a valid interpretation is provided by the mutual information between the active fluctuations $\boldsymbol{\eta}(t)$ and the particle trajectory $\mathbf{x}(t)$.

V. MUTUAL INFORMATION

The pathwise mutual information [108] between the particle trajectory $\bar{\mathbf{x}}$ (starting at \mathbf{x}_0) and a realization $\bar{\boldsymbol{\eta}}$ of the active fluctuations [with $\boldsymbol{\eta}(0) = \boldsymbol{\eta}_0$] is given as

$$\mathcal{I}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] = \ln \frac{\mathfrak{p}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}]}{\mathfrak{p}[\bar{\mathbf{x}}]\mathfrak{p}[\bar{\boldsymbol{\eta}}]} = \ln \frac{\mathfrak{p}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}]}{\mathfrak{p}[\bar{\mathbf{x}}]}, \quad (49)$$

where $\mathfrak{p}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] = \mathfrak{p}[\mathbf{x}, \boldsymbol{\eta}|\mathbf{x}_0, \boldsymbol{\eta}_0]p(\mathbf{x}_0, \boldsymbol{\eta}_0)$, $\mathfrak{p}[\bar{\mathbf{x}}] = \mathfrak{p}[\mathbf{x}|\mathbf{x}_0]p(\mathbf{x}_0)$, and $\mathfrak{p}[\bar{\boldsymbol{\eta}}] = \mathfrak{p}[\boldsymbol{\eta}|\boldsymbol{\eta}_0]p(\boldsymbol{\eta}_0)$ include the initial densities $p(\mathbf{x}_0, \boldsymbol{\eta}_0)$, $p(\mathbf{x}_0) = \int d\boldsymbol{\eta}_0 p(\mathbf{x}_0, \boldsymbol{\eta}_0)$, and $p(\boldsymbol{\eta}_0) = \int d\mathbf{x}_0 p(\mathbf{x}_0, \boldsymbol{\eta}_0)$ of the Brownian particle and the active fluctuations (see also the discussion in Sec. IV B). This pathwise mutual information quantifies the reduction in uncertainty about the path $\bar{\mathbf{x}}$ when we know the realization $\bar{\boldsymbol{\eta}}$ and vice versa, and can therefore, loosely speaking, be seen as a measure of the correlations between $\bar{\mathbf{x}}$ and $\bar{\boldsymbol{\eta}}$. Note that it can become negative if $\mathfrak{p}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] < \mathfrak{p}[\bar{\mathbf{x}}]$, while its average is always positive.

Likewise, the pathwise mutual information between the time-reversed trajectory $\tilde{\bar{\mathbf{x}}}$ from Eq. (17) and a suitably chosen time-reversed realization $\tilde{\bar{\boldsymbol{\eta}}}$ of the active fluctuations is

$$\tilde{\mathcal{I}}[\tilde{\bar{\mathbf{x}}}, \tilde{\bar{\boldsymbol{\eta}}}] = \ln \frac{\tilde{\mathfrak{p}}[\tilde{\bar{\mathbf{x}}}, \tilde{\bar{\boldsymbol{\eta}}}]}{\tilde{\mathfrak{p}}[\tilde{\bar{\mathbf{x}}}]}. \quad (50)$$

For their difference

$$\Delta\mathcal{I}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] = \mathcal{I}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] - \tilde{\mathcal{I}}[\tilde{\bar{\mathbf{x}}}, \tilde{\bar{\boldsymbol{\eta}}}], \quad (51a)$$

we thus find

$$\Delta\mathcal{I}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] = \ln \frac{\tilde{\mathfrak{p}}[\bar{\mathbf{x}}]}{\mathfrak{p}[\bar{\mathbf{x}}]} - \ln \frac{\tilde{\mathfrak{p}}[\bar{\mathbf{x}}|\tilde{\bar{\boldsymbol{\eta}}}]}{\mathfrak{p}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}]}. \quad (51b)$$

This expression represents the pathwise mutual information difference between a combined forward process $(\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}})$ and its backward twin. If $\Delta\mathcal{I}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}]$ is positive, the pathwise mutual information $\mathcal{I}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}]$ along the combined forward path is larger than $\tilde{\mathcal{I}}[\tilde{\bar{\mathbf{x}}}, \tilde{\bar{\boldsymbol{\eta}}}]$ for the time-reversed path [see Eq. (51a)], implying that correlations between the particle trajectory $\bar{\mathbf{x}}$ and the active fluctuation $\bar{\boldsymbol{\eta}}$ are stronger in the time-forward direction. Intuitively, we may thus say that they are more likely to occur together than their time-reversed twins, making the combined forward process $(\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}})$ the more “natural” one of the two processes in terms of pathwise mutual information.

To make the connection to irreversibility more rigorous, we rewrite in Eq. (51b) the pathwise mutual information difference $\Delta\mathcal{I}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}]$ explicitly as path probability ratios between time-forward and time-backward processes. We now see that $\Delta\mathcal{I}$ via the term $-\ln(\tilde{\mathfrak{p}}[\bar{\mathbf{x}}]/\mathfrak{p}[\bar{\mathbf{x}}])$ is directly related to the irreversibility measure $\Delta\Sigma$ for the actively driven Brownian particle and its change in system entropy ΔS_{sys} . The additional log ratio $\ln(\tilde{\mathfrak{p}}[\bar{\mathbf{x}}|\tilde{\bar{\boldsymbol{\eta}}}]/\mathfrak{p}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}])$ involves the probability of the forward path $\bar{\mathbf{x}}$ being generated by the specific realization $\bar{\boldsymbol{\eta}}$ of the active fluctuations for which we measure the mutual information content with $\bar{\mathbf{x}}$ in $\Delta\mathcal{I}$, and, likewise, the path probability of the time-reversed twin being generated by the time-reversed fluctuation. We can rewrite this term by splitting off the contributions from the initial densities,

$$\ln \frac{\tilde{\mathfrak{p}}[\bar{\mathbf{x}}|\tilde{\bar{\boldsymbol{\eta}}}]}{\mathfrak{p}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}]} = \ln \frac{\tilde{\mathfrak{p}}[\tilde{\bar{\mathbf{x}}}, \tilde{\bar{\boldsymbol{\eta}}}]}{\mathfrak{p}[\tilde{\bar{\mathbf{x}}}, \tilde{\bar{\boldsymbol{\eta}}}]} + \ln \frac{p(\tilde{\bar{\mathbf{x}}}_0|\tilde{\bar{\boldsymbol{\eta}}})}{p(\mathbf{x}_0|\boldsymbol{\eta}_0)}. \quad (52)$$

Here we keep the possibility that the initial particle position is conditioned on the initial state of the active fluctuations $p(\mathbf{x}_0|\bar{\boldsymbol{\eta}}) = p(\mathbf{x}_0|\boldsymbol{\eta}_0)$, which is more general than the situation $p(\mathbf{x}_0|\bar{\boldsymbol{\eta}}) = p(\mathbf{x}_0)$ for which we calculated $\Delta\Sigma$ in Sec. IV B [see also the discussion around Eq. (35)]. According to Eq. (17), the time-reversed initial position is given by the final point of the forward path, $\tilde{\bar{\mathbf{x}}}_0 = \mathbf{x}_\tau$. It therefore depends on the complete history of the active fluctuations, which is captured equivalently by $\tilde{\bar{\boldsymbol{\eta}}}$ or $\bar{\boldsymbol{\eta}}$, for any reasonable choice of the time-reversed fluctuations in terms of the forward realization [see also Eq. (55) below]. With $p(\tilde{\bar{\mathbf{x}}}_0|\tilde{\bar{\boldsymbol{\eta}}}) = p(\mathbf{x}_\tau|\bar{\boldsymbol{\eta}})$, we can identify the boundary term in Eq. (52) as the change in system entropy

$$\Delta S_{\text{sys}}^{|\bar{\boldsymbol{\eta}}|}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] = -k_B \ln p(\mathbf{x}_\tau|\bar{\boldsymbol{\eta}}) + k_B \ln p(\mathbf{x}_0|\boldsymbol{\eta}_0), \quad (53)$$

which occurs along the trajectory $\bar{\mathbf{x}}$ under the specific realization $\bar{\boldsymbol{\eta}}$ of the active fluctuations (we explicitly indicate the conditioning on the realization of the active fluctuation by adding “ $|\bar{\boldsymbol{\eta}}$ ” as a superscript).

To quantify the first term in Eq. (52), we exploit that a prescribed active fluctuation $\bar{\boldsymbol{\eta}} = \{\boldsymbol{\eta}(t)\}_{t=0}^{\tau}$ acts like an additional driving force in Eq. (5). Without needing to know any further details about $\boldsymbol{\eta}(t)$ (e.g., of how it is generated), we can thus derive the conditioned forward path probability $\mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0, \bar{\boldsymbol{\eta}}]$ directly from Eq. (5) as a standard Onsager-Machlup path integral,

$$\mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0, \bar{\boldsymbol{\eta}}] \propto \exp \left\{ - \int_0^{\tau} dt \frac{(\dot{\mathbf{x}}_t - \mathbf{v}_t - \sqrt{2D_a}\boldsymbol{\eta}_t)^2}{4D} - \frac{\nabla \cdot \mathbf{v}_t}{2} \right\}. \quad (54)$$

In order to evaluate the time-reversed counterpart, we have to specify the behavior of the active fluctuation signal $\boldsymbol{\eta}(t)$ under inversion of the direction of time. There are essentially two choices,

$$\tilde{\boldsymbol{\eta}}_{\pm}(t) = \pm \boldsymbol{\eta}(\tau - t), \quad (55)$$

corresponding to $\boldsymbol{\eta}(t)$ being an even or odd process. While both these options *a priori* appear to be equally valid, they are connected to different interpretations of the active fluctuations. We argue below (beginning of Sec. VA) that for passive tracer particles which are suspended in an active bath, the active fluctuations $\boldsymbol{\eta}(t)$ bear the character of an external force and therefore should be considered even under time reversal. In contrast, for an active particle with self-propulsion (see beginning of Sec. VB), the active fluctuations $\boldsymbol{\eta}(t)$ are internal to the particle and thus should change sign when time is reversed (odd under time reversal), otherwise, the particle trajectory would not be reversed under inversion of the direction of time in a setup in which only self-propulsion is acting (no thermal noise and no external forces) [109].

A. Active bath

If we interpret $\boldsymbol{\eta}(t)$ as fluctuations coming from an active bath, they are external to the particle and as such are similar to externally applied forces. The question of irreversibility under a prescribed realization $\bar{\boldsymbol{\eta}}$ is then related to the question of how likely the backward particle trajectory $\tilde{\mathbf{x}}(t)$ [see Eq. (17)] is to be observed when exactly the same forces act on the particle at identical positions during the forward and backward motion. Hence, this case is described by active fluctuations which are even under time reversal, i.e., by the plus sign in Eq. (55).

Using $\tilde{\boldsymbol{\eta}}_+(t) = +\boldsymbol{\eta}(\tau - t)$ as the time-reversed fluctuation, we find

$$\ln \frac{\tilde{\mathfrak{p}}[\tilde{\mathbf{x}}|\tilde{\mathbf{x}}_0, \tilde{\boldsymbol{\eta}}_+]}{\mathfrak{p}[\underline{\mathbf{x}}|\mathbf{x}_0, \bar{\boldsymbol{\eta}}]} = -\Delta S_+[\underline{\mathbf{x}}|\bar{\boldsymbol{\eta}}]/k_B \quad (56a)$$

for the path probability ratio, where $\Delta S_+[\underline{\mathbf{x}}|\bar{\boldsymbol{\eta}}]$ is given by the entropy production in the thermal environment from Eq. (28),

$$\begin{aligned} \Delta S_+[\underline{\mathbf{x}}|\bar{\boldsymbol{\eta}}] &= \int_0^{\tau} dt \delta S_+(t) \\ &= \frac{1}{T} \int_0^{\tau} d\mathbf{x}(t) \cdot [\mathbf{f}(\mathbf{x}(t), t) + \sqrt{2D_a}\boldsymbol{\eta}(t)]. \end{aligned} \quad (56b)$$

Together with Eq. (53), we thus obtain the total entropy production along the particle trajectory $\underline{\mathbf{x}}$ which is generated by a given realization $\bar{\boldsymbol{\eta}}$ of the active fluctuations assumed to be even under time inversion. It is defined in analogy to Eq. (22) as

$$\Delta S_{\text{tot}}^+[\underline{\mathbf{x}}|\bar{\boldsymbol{\eta}}] = \Delta S_+[\underline{\mathbf{x}}|\bar{\boldsymbol{\eta}}] + \Delta S_{\text{sys}}^{\bar{\boldsymbol{\eta}}}[\underline{\mathbf{x}}|\bar{\boldsymbol{\eta}}]. \quad (57)$$

By construction [see Eqs. (53), (56a), and also Eq. (52)], $\Delta S_{\text{tot}}^+[\underline{\mathbf{x}}|\bar{\boldsymbol{\eta}}]$ is identical to the log ratio $-k_B \ln [\tilde{\mathfrak{p}}[\tilde{\mathbf{x}}|\tilde{\boldsymbol{\eta}}_+]/\mathfrak{p}[\underline{\mathbf{x}}|\bar{\boldsymbol{\eta}}]]$.

Combining this result with Eqs. (52), (43), and (51), we then infer that $\Delta\Sigma[\underline{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_{\tau})$ is a combination of the conditioned entropy production and the difference in mutual information between time-forward and time-backward processes,

$$\Delta\Sigma[\underline{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_{\tau}) = \Delta S_{\text{tot}}^+[\underline{\mathbf{x}}|\bar{\boldsymbol{\eta}}] - k_B \Delta\mathcal{I}_+[\underline{\mathbf{x}}, \bar{\boldsymbol{\eta}}], \quad (58)$$

where $\Delta\mathcal{I}_+[\underline{\mathbf{x}}, \bar{\boldsymbol{\eta}}]$ is given by Eq. (51) with $\bar{\boldsymbol{\eta}} = \bar{\boldsymbol{\eta}}_+ = \{\tilde{\boldsymbol{\eta}}_+(t)\}_{t=0}^{\tau}$. Note that the individual terms on the right-hand side both depend on the active noise $\bar{\boldsymbol{\eta}}$ with a dependence that is compensated exactly because the left-hand side is independent of $\bar{\boldsymbol{\eta}}$.

As an immediate consequence of Eq. (44), the total conditioned entropy production together with the mutual information difference fulfill the integral fluctuation theorem

$$\langle e^{-(\Delta S_{\text{tot}}^+/k_B - \Delta\mathcal{I}_+)} \rangle_{\underline{\mathbf{x}}} = 1, \quad (59)$$

and the second-law-like relation

$$\langle \Delta S_{\text{tot}}^+[\underline{\mathbf{x}}|\bar{\boldsymbol{\eta}}] \rangle_{\underline{\mathbf{x}}} - k_B \langle \Delta\mathcal{I}_+[\underline{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \rangle_{\underline{\mathbf{x}}} \geq 0. \quad (60a)$$

These results (59) and (60) are valid for any realization $\bar{\boldsymbol{\eta}}$ of the active fluctuations because the original average $\langle \cdot \rangle$ in Eq. (44) is just over the distribution of all particle trajectories $\underline{\mathbf{x}}$. We accentuate this fact here by the subscript $\underline{\mathbf{x}}$ at the averages $\langle \cdot \rangle_{\underline{\mathbf{x}}}$ to distinguish them from the standard meaning of the brackets $\langle \cdot \rangle$ as an average over *all* stochastic quantities present. Since the terms in Eq. (60a) taken together are independent of $\bar{\boldsymbol{\eta}}$ [see Eq. (58)], we can even average over the joint distribution $\mathfrak{p}[\underline{\mathbf{x}}, \bar{\boldsymbol{\eta}}]$ to obtain a further second-law-like relation

$$\langle \Delta S_{\text{tot}}^+ \rangle - k_B \langle \Delta\mathcal{I}_+ \rangle \geq 0. \quad (60b)$$

We emphasize that Eqs. (58)–(60), being a direct consequence of the structure of the equation of motion (5), are completely independent of the specific model behind the

active fluctuations. Moreover, they are formally similar to the fluctuation theorem with information exchange derived by Sagawa and Ueda [110] for a completely different physical setup. Sagawa and Ueda considered the difference between initial and final correlations of a system of interest with an “information reservoir,” which provides correlations as a resource of entropy changes.

For the Ornstein-Uhlenbeck representation (7), we can use our result (42b) for $\Delta\Sigma[\bar{\mathbf{x}}]$ to derive an explicit expression for the mutual information difference,

$$\begin{aligned} \Delta\mathcal{I}_+[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] &+ \left[\ln \frac{p(\mathbf{x}_\tau|\bar{\boldsymbol{\eta}})}{p(\mathbf{x}_\tau)} - \ln \frac{p(\mathbf{x}_0|\boldsymbol{\eta}_0)}{p(\mathbf{x}_0)} \right] \\ &= \frac{1}{k_B T} \left[\sqrt{2D_a} \int_0^\tau dt \dot{\mathbf{x}}_t^\top \boldsymbol{\eta}_t \right. \\ &\quad \left. + \frac{D_a}{D} \int_0^\tau dt \int_0^\tau dt' \dot{\mathbf{x}}_t^\top \mathbf{f}_{t'} \Gamma_\tau(t, t') \right] \\ &= \frac{1}{k_B T} \left[\Delta A_+[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \right. \\ &\quad \left. + \frac{D_a}{D} \int_0^\tau dt \int_0^\tau dt' \dot{\mathbf{x}}_t^\top \mathbf{f}_{t'} \Gamma_\tau(t, t') \right], \end{aligned} \quad (61)$$

where in the last equality we use $\Delta A_+[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] = \int_0^\tau dt \delta A_+(t)$ with the “heat” $\delta A_+(t)$ exchanged with the active bath as identified in Eq. (27).

B. Self-propulsion

If the active fluctuations represent an effective model for a self-propulsion mechanism of the particle, we choose $\boldsymbol{\eta}(t)$ to be odd under time reversal for a proper assessment of irreversibility; i.e., we choose the minus sign in Eq. (55). Otherwise, we would relate irreversibility to the likelihood that thermal fluctuations make the particle move a certain path backward *against* its internal self-propulsion rather than to the probability that a self-propelled particle traces out a given trajectory backward in time driven by the same propulsion forces [109].

Calculating the path probability ratio $\mathfrak{p}[\bar{\mathbf{x}}|\bar{\mathbf{x}}_0, \bar{\boldsymbol{\eta}}_-]/\mathfrak{p}[\mathbf{x}|\mathbf{x}_0, \bar{\boldsymbol{\eta}}]$ for odd active fluctuations $\bar{\boldsymbol{\eta}}(t) = \tilde{\boldsymbol{\eta}}_-(t) = -\boldsymbol{\eta}(\tau - t)$, we obtain

$$\ln \frac{\tilde{\mathfrak{p}}[\bar{\mathbf{x}}|\bar{\mathbf{x}}_0, \tilde{\boldsymbol{\eta}}_-]}{\mathfrak{p}[\mathbf{x}|\mathbf{x}_0, \bar{\boldsymbol{\eta}}]} = -\Delta S_-[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}]/k_B. \quad (62a)$$

The quantity $\Delta S_-[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}]$ is given by

$$\begin{aligned} \Delta S_-[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] &= \int_0^\tau dt \delta S_-(t) \\ &= \frac{1}{T} \int_0^\tau (d\mathbf{x}(t) - \sqrt{2D_a} \boldsymbol{\eta}(t) dt) \cdot \mathbf{f}(\mathbf{x}(t), t). \end{aligned} \quad (62b)$$

It thus represents the entropy production in the environment along the trajectory $\bar{\mathbf{x}}$ because $\delta S_-(t)$ results from the heat

$-\delta Q_-(t)$ dissipated into the thermal bath during a displacement of the active self-propelled particle, which deviates from the “force-free” path that would be carved out by the self-propulsion drive alone; see also Eq. (32) and the discussion above Eq. (29). In analogy to Eq. (22), we finally combine $\Delta S_-[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}]$ with Eq. (53) to define the total entropy production under a given realization $\bar{\boldsymbol{\eta}}$ of the active propulsion as

$$\Delta S_{\text{tot}}^-[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] = \Delta S_-[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] + \Delta S_{\text{sys}}^{\bar{\boldsymbol{\eta}}}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}]. \quad (63)$$

Note that with Eqs. (53) and (62a) [see also Eq. (52)], we can write this total entropy production as the log ratio $-k_B \ln(\tilde{\mathfrak{p}}[\bar{\mathbf{x}}|\tilde{\boldsymbol{\eta}}_-]/\mathfrak{p}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}])$ of path weights.

Hence, we find again that $\Delta\Sigma[\bar{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau)$ consists of conditional total entropy production and mutual information difference [compare Eqs. (51), (43), and (63)],

$$\Delta\Sigma[\bar{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau) = \Delta S_{\text{tot}}^-[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] - k_B \Delta\mathcal{I}_-[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}]. \quad (64)$$

This identity provides exactly the same interpretation of $\Delta\Sigma[\bar{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau)$ as we found earlier for the case of an active bath [cf. Eq. (58)] but with different expressions for the entropy production and the mutual information difference. The mutual information in the backward process is now measured with respect to the sign-inverted active fluctuation process $\tilde{\boldsymbol{\eta}}_-(t) = -\boldsymbol{\eta}(\tau - t)$, i.e., $\Delta\mathcal{I}_-[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}]$ in Eq. (64) is given by Eq. (51) with $\bar{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}_- = \{\tilde{\boldsymbol{\eta}}_-(t)\}_{t=0}^\tau$. As in the case of an active bath (58), the dependence of the individual terms ΔS_{tot}^- and $\Delta\mathcal{I}_-$ in Eq. (64) on the active forcing $\bar{\boldsymbol{\eta}}$ cancel exactly to result in the $\bar{\boldsymbol{\eta}}$ -independent expression $\Delta\Sigma[\bar{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau)$.

From Eq. (44), we immediately obtain the integral fluctuation theorem

$$\langle e^{-(\Delta S_{\text{tot}}^-/k_B - \Delta\mathcal{I}_-)} \rangle_{\bar{\mathbf{x}}} = 1, \quad (65)$$

and the second-law-like relation

$$\langle \Delta S_{\text{tot}}^-[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] \rangle_{\bar{\mathbf{x}}} - k_B \langle \Delta\mathcal{I}_-[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \rangle_{\bar{\mathbf{x}}} \geq 0 \quad (66a)$$

for active self-propulsion. Like in Eq. (59), we use the subscript $\bar{\mathbf{x}}$ to explicitly indicate that the averages are over the distribution $\mathfrak{p}[\bar{\mathbf{x}}]$ alone and that the resulting quantities are still functionals of the active fluctuations and are valid for any realization $\bar{\boldsymbol{\eta}}$ of the self-propelling forces. After extending to the average over $\mathfrak{p}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}]$ to include these active fluctuations, we end up with the bound

$$\langle \Delta S_{\text{tot}}^- \rangle - k_B \langle \Delta\mathcal{I}_- \rangle \geq 0. \quad (66b)$$

Again, these findings (64)–(66) are independent of the specific model for the active fluctuations because they are a direct consequence of the particle’s equation of motion (5). Moreover, they are formally similar to the Sagawa-Ueda

fluctuation theorem with information [110]. If the active propulsion $\boldsymbol{\eta}(t)$ is modeled by Eq. (7), we can write the mutual information difference in the explicit form

$$\begin{aligned} \Delta\mathcal{I}_-[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] &+ \left[\ln \frac{p(\mathbf{x}_\tau|\bar{\boldsymbol{\eta}})}{p(\mathbf{x}_\tau)} - \ln \frac{p(\mathbf{x}_0|\boldsymbol{\eta}_0)}{p(\mathbf{x}_0)} \right] \\ &= \frac{1}{k_B T} \left[-\sqrt{2D_a} \int_0^\tau dt \boldsymbol{\eta}_i^\top \mathbf{f}_i \right. \\ &\quad \left. + \frac{D_a}{D} \int_0^\tau dt \int_0^\tau dt' \dot{\mathbf{x}}_i^\top \mathbf{f}_i \Gamma_\tau(t, t') \right] \\ &= \frac{1}{k_B T} \left[\Delta A_-[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \right. \\ &\quad \left. + \frac{D_a}{D} \int_0^\tau dt \int_0^\tau dt' \dot{\mathbf{x}}_i^\top \mathbf{f}_i \Gamma_\tau(t, t') \right] \end{aligned} \quad (67)$$

by using Eq. (42b) and by defining $\Delta A_-[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] = \int_0^\tau dt \delta A_-(t)$ [see Eq. (31)] as the total ‘‘heat’’ transferred from the active propulsion to the thermal bath along the trajectory $\bar{\mathbf{x}}$.

C. Relation to information theory

The second-law-like relations (60) and (66) involve the averages $\langle \Delta\mathcal{I}_+ \rangle$ and $\langle \Delta\mathcal{I}_- \rangle$ of the mutual information difference. As we show in the following, these averages are closely related to central concepts in information theory entailing additional bounds and an interpretation in terms of hypothesis testing.

Because of the unbiased Gaussian character of the active fluctuations, the probabilities for observing a time-forward and its time-reversed realization are the same, i.e., $\tilde{\mathbf{p}}[\bar{\boldsymbol{\eta}}] = \mathbf{p}[\bar{\boldsymbol{\eta}}]$. We can therefore rewrite Eq. (51b) as

$$\Delta\mathcal{I}_\pm[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] = \ln \frac{\tilde{\mathbf{p}}[\bar{\mathbf{x}}]}{\mathbf{p}[\bar{\mathbf{x}}]} - \ln \frac{\tilde{\mathbf{p}}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}_\pm]}{\mathbf{p}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}]}, \quad (68)$$

where we insert our two options $\bar{\boldsymbol{\eta}} = \bar{\boldsymbol{\eta}}_\pm$ from Eq. (55) for the time-reversed active fluctuation. Performing the average over $(\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}})$ with density $\mathbf{p}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}]$, we then find

$$\langle \Delta\mathcal{I}_\pm[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \rangle = D_{\text{KL}}(\mathbf{p}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \| \tilde{\mathbf{p}}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}_\pm]) - D_{\text{KL}}(\mathbf{p}[\bar{\mathbf{x}}] \| \tilde{\mathbf{p}}[\bar{\mathbf{x}}]). \quad (69)$$

Here, we use the definition

$$D_{\text{KL}}(\mathbf{p}[\bar{\mathbf{y}}] \| \tilde{\mathbf{p}}[\bar{\mathbf{y}}]) = \int \mathcal{D}\bar{\mathbf{y}} \mathbf{p}[\bar{\mathbf{y}}] \ln \frac{\mathbf{p}[\bar{\mathbf{y}}]}{\tilde{\mathbf{p}}[\bar{\mathbf{y}}]} \quad (70)$$

of the Kullback-Leibler divergence between a probability density $\mathbf{p}[\bar{\mathbf{y}}]$ for a process $\bar{\mathbf{y}}$ and another density $\tilde{\mathbf{p}}[\bar{\mathbf{y}}]$ for a second process $\tilde{\bar{\mathbf{y}}}$; in our case, the two processes are related by time reversal. The Kullback-Leibler divergence is a standard concept in information theory to measure how

distinct two probability densities are. It is non-negative and equals zero if and only if the two probabilities are identical [111,112].

The result (69) shows that the average mutual information difference $\langle \Delta\mathcal{I}_\pm[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \rangle$ is the difference between the Kullback-Leibler divergence of the particle trajectory $\bar{\mathbf{x}}$ relative to its time-reversed twin $\tilde{\bar{\mathbf{x}}}$ and the Kullback-Leibler divergence of the combined path $(\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}})$ relative to the time-reversed realization $(\tilde{\bar{\mathbf{x}}}, \tilde{\bar{\boldsymbol{\eta}}}_\pm)$. We can therefore interpret it to measure how much harder it is to discriminate between time-forward and time-backward realizations if only the particle trajectory is known rather than the full dynamics including the active noise realization $\bar{\boldsymbol{\eta}}$. Since $\bar{\mathbf{x}}$ can be seen as a ‘‘coarse-graining projection’’ of $(\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}})$, we expect the discrimination to become harder, i.e., $D_{\text{KL}}(\mathbf{p}[\bar{\mathbf{x}}] \| \tilde{\mathbf{p}}[\tilde{\bar{\mathbf{x}}}])$ to become smaller compared to $D_{\text{KL}}(\mathbf{p}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \| \tilde{\mathbf{p}}[\tilde{\bar{\mathbf{x}}, \tilde{\bar{\boldsymbol{\eta}}}_\pm}])$ (see Ref. [113] for a similar discussion). This intuition is corroborated by the so-called data processing inequality [111,112], which, applied to our situation, proves

$$D_{\text{KL}}(\mathbf{p}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \| \tilde{\mathbf{p}}[\tilde{\bar{\mathbf{x}}, \tilde{\bar{\boldsymbol{\eta}}}_\pm}]) \geq D_{\text{KL}}(\mathbf{p}[\bar{\mathbf{x}}] \| \tilde{\mathbf{p}}[\tilde{\bar{\mathbf{x}}}). \quad (71)$$

As a direct consequence, we find from Eq. (69) the bound

$$\langle \Delta\mathcal{I}_\pm[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \rangle \geq 0 \quad (72)$$

on the average mutual information difference. Exploiting Eqs. (60b) and (66b), respectively, we infer that $\langle \Delta S_{\text{tot}}^\pm \rangle \geq 0$. This follows also from the fact that the $\Delta S_{\text{tot}}^\pm$ are given as log ratios of path probabilities, which are conditioned on $\bar{\boldsymbol{\eta}}$ [see the discussion around Eqs. (56) and (62)]. Accordingly, they obey an integral fluctuation theorem when averaged over the conditioned density $\mathbf{p}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}]$ but then also when averaged over the full density $\mathbf{p}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}]$.

More interestingly, we can also use Eq. (72) to equip the second-law-like relations for the total irreversibility measure $\Delta\Sigma + \Delta S_{\text{sys}}$ with an upper bound. Taking the average of Eqs. (58) and (64) over $(\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}})$, the non-negativity (72) of $\langle \Delta\mathcal{I}_\pm \rangle$ implies

$$\langle \Delta S_{\text{tot}}^\pm[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] \rangle \geq \langle \Delta\Sigma[\bar{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau) \rangle \geq 0. \quad (73)$$

The total average irreversibility of a particle trajectory measured as $\langle \Delta\Sigma + \Delta S_{\text{sys}} \rangle$ is thus always smaller than (or equal to) the mean entropy change $\int \mathcal{D}\bar{\mathbf{x}} \mathbf{p}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] \Delta S_{\text{tot}}^\pm[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}]$, which would occur for a given realization $\bar{\boldsymbol{\eta}}$ of the active fluctuations if treated as a known (and measurable) external force contributing to the dissipation in the thermal environment averaged over the distribution $\mathbf{p}[\bar{\boldsymbol{\eta}}]$ of all possible active fluctuations, i.e., $\int \mathcal{D}\bar{\boldsymbol{\eta}} \mathbf{p}[\bar{\boldsymbol{\eta}}] \int \mathcal{D}\bar{\mathbf{x}} \mathbf{p}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] \Delta S_{\text{tot}}^\pm[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] = \int \mathcal{D}\bar{\boldsymbol{\eta}} \int \mathcal{D}\bar{\mathbf{x}} \mathbf{p}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \Delta S_{\text{tot}}^\pm[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] = \langle \Delta S_{\text{tot}}^\pm[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] \rangle$. For this bound to be valid, it does not matter whether these forcings come from an active bath and thus are considered even under time reversal (‘‘+’’ sign), or from active self-propulsion, which is odd under time reversal (‘‘-’’ sign). The difference between

$\langle \Delta S_{\text{tot}}^{\pm}[\bar{\mathbf{x}}|\bar{\boldsymbol{\eta}}] \rangle$ and $\langle \Delta \Sigma + \Delta S_{\text{sys}} \rangle$ is compensated by the build up of mutual information.

We point out that the upper bound for the total average irreversibility in Eq. (73) is stronger than the usual coarse-graining inequality obtained when integrating out “hidden” d.o.f. (see, e.g., Ref. [114]). In the present notation, such an inequality would read $\langle \Delta S_{\text{tot}}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \rangle \geq \langle \Delta \Sigma[\bar{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau) \rangle$. Here, $\langle \Delta S_{\text{tot}}[\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}}] \rangle$ would denote the total average entropy production in the thermal environment of the combined processes $(\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}})$, a quantity which is not well defined, however, because we cannot quantify the entropy production associated with the auxiliary process $\bar{\boldsymbol{\eta}}$.

Finally, we remark that the inequality (71) has an interesting interpretation in the context of hypothesis testing [112,115] when estimating the direction of the arrow of time in the system [33,116]. It states that the probability of misclassifying a specific path as having been generated by a forward dynamics when, in reality, it was generated by a backward one decreases faster with the number of observations if more detailed information on the dynamics of the system is available by additionally monitoring the realizations $\bar{\boldsymbol{\eta}}$ of the active fluctuations.

D. Discussion

The fluctuation theorems and second-law-like relations (59), (60) and (65), (66) for an active bath and for active self-propulsion, respectively, and their interpretation in terms of mutual information differences are our third main result; see also Eq. (73). For both cases, we find that the total irreversibility measure $\Delta \Sigma[\bar{\mathbf{x}}] + \Delta S_{\text{sys}}(\mathbf{x}_0, \mathbf{x}_\tau)$ consists of two contributions: First, the “usual” entropy production $\Delta S_{\text{tot}}^{\pm}$, which we would measure for the motion of the Brownian particle under a given realization of the active fluctuations (as if the active fluctuations were just some additional known “external” driving force), and second, the difference in mutual information $\Delta \mathcal{I}_{\pm}$ accumulated between the particle trajectory and the active nonequilibrium environment along the forward versus the backward path. We note that even though the individual contributions in $\Delta S_{\text{tot}}^{\pm} - k_B \Delta \mathcal{I}_{\pm}$ depend on the specific realization $\bar{\boldsymbol{\eta}}$ of the active fluctuations, their sum does not [see Eqs. (58) and (64)]; i.e., entropy production and change in mutual information always compensate their dependence on the realization of the active fluctuations. However, if we want to have access to these individual contributions $\Delta S_{\text{tot}}^{\pm}$ and $\Delta \mathcal{I}_{\pm}$ directly, we would have to measure the active fluctuations $\bar{\boldsymbol{\eta}}$, i.e., the direction and magnitude of the forces representing the activity in the system. In general, this may be a challenging task in typical experiments with active Brownian particles, as one would have to separate thermal from active fluctuations. We can imagine, however, that at least partial information about $\bar{\boldsymbol{\eta}}$ could be obtained experimentally. For example, tracking the orientation of an active self-propelled particle (e.g., a bacterium with a

flagellum), it may be possible to infer the direction of $\boldsymbol{\eta}$. Moreover, we could think of an experimental setup in which the active fluctuations $\bar{\boldsymbol{\eta}}$ are realized artificially by an external colored noise source, like in Ref. [117], then all relevant quantities may be accessible.

The entropy productions $\Delta S_{\text{tot}}^{\pm}$ for an active bath and for active self-propulsion, respectively, as obtained from irreversibility arguments in Eqs. (56) and (62) are consistent with the energetics derived in Secs. IVA 1 and IVA 2 [compare with Eqs. (25) and (29)]. They are thus directly related to the heat dissipated into the *thermal* part of the environment. Hence, the appearance of the pathwise mutual information difference in the fluctuation theorem is a consequence of the *active* fluctuations being present as a nonequilibrium bath in addition to the usual thermal bath. Indeed, we can easily see from Eqs. (61) and (67) that $\Delta \mathcal{I}_{\pm}$ vanish identically in the absence of active fluctuations, $D_a = 0$. Accordingly, ΔS_{\pm} reduce to the standard entropy production in the thermal environment in this limit [see Eqs. (56), (62), and (19)].

We emphasize that the pathwise mutual information quantifies how the active fluctuations contribute to the irreversibility of the particle trajectory but does not capture the unavoidable dissipation connected with maintaining the active nature of the nonequilibrium environment itself. In fact, our effective description (7) of the active fluctuations as a time-correlated Gaussian noise source does not have any knowledge about the microscopic details generating this noise so that it obviously cannot assess the associated dissipative processes. On the one hand, we may therefore suspect that the appearance of the pathwise mutual information is a consequence of this coarse-grained description of the active environment and might be replaced by a “finer” measure once all the details are known [41,71,72]. On the other hand, we may argue that these microscopic details behind the active fluctuations are irrelevant if we are interested only in characterizing the irreversibility of the particle trajectory and related thermodynamiclike properties of the system emerging on this level of coarse graining. Then, knowledge of the statistical properties of the active fluctuations, as provided by Eq. (7), is sufficient, and the pathwise mutual information between particle trajectory and active noise realization emerges as a natural and adequate irreversibility measure. This situation may be comparable to the one for entropy production in a thermal bath: We quantify entropy production solely from the statistical properties of the thermal noise without having to rely on a detailed microscopic description of the bath.

VI. EXAMPLE: HARMONIC POTENTIAL

To illustrate our results, we consider the example of a one-dimensional Brownian particle trapped in a harmonic potential $U(x, t) = (k/2)(x - ut)^2$, whose center is either held at fixed position ($u = 0$) or else is displaced at constant velocity $u \neq 0$. Such a setup can be readily

implemented in experiment with state-of-the-art optical tweezers and in fact has been used to study various aspects of stochastic thermodynamics and active matter, for instance, in Refs. [23,49,79,118–120].

The Langevin equation of motion (5) for this specific setup is linear,

$$\dot{x}(t) = -\frac{k}{\gamma}[x(t) - ut] + \sqrt{2D}\xi(t) + \sqrt{2D_a}\eta(t). \quad (74)$$

We can therefore solve the associated Fokker-Planck equation analytically [47,99,121] to obtain the propagator in closed form. From that, we can explicitly compute the averages of the irreversibility measure $\Delta\Sigma$, the mutual information difference $\Delta\mathcal{I}_\pm$, and the change in system entropy ΔS_{sys} for a Gaussian initial distribution $p_0(x_0)$ of particle positions with zero mean and variance c_0^2 . Although the calculations are straightforward (we present some details in Appendix D), the resulting formulas for $\langle\Delta\Sigma\rangle$, $\langle\Delta S_{\text{sys}}\rangle$, and $\langle\Delta\mathcal{I}_\pm\rangle$ are lengthy and bulky so that we discuss them mostly in graphical form (see Figs. 2 and 4) and give explicit expressions only in some limiting cases [see Eqs. (75), (76), and (79) below]. For the initial variance c_0^2 , two specific cases are of particular physical relevance. First, $c_0^2 = k_B T/k$ corresponding to a Gaussian distribution that would be created as an equilibrium state by thermal fluctuations only. Starting from this distribution, the time evolution of our irreversibility measures includes the transient relaxation from the thermal state towards the steady state which develops due to the presence of the active fluctuations. Second, $c_0^2 = (1/k)\{k_B T + [(\gamma D_a)/(1 + (k\tau_a/\gamma))]\}$ corresponding to a Gaussian distribution with a variance which is exactly the same as the one the particle distribution has when particle and active fluctuations are in their joint steady state. In that case, the particle distributions at the beginning and end of the process are identical so that any contributions to the irreversibility measure $\Delta\Sigma$ that are not associated with the displacement u of the trap are solely due to the build up of correlations between particle position and active d.o.f. In the following, we focus on the first alternative and briefly come back to the second alternative at the end of Sec. VIA and in Sec. VIB.

A. Irreversibility

In Fig. 2, we compare $\langle\Delta\Sigma\rangle$, $\langle\Delta S_{\text{sys}}\rangle$, and $\langle\Delta\Sigma + \Delta S_{\text{sys}}\rangle$ for $u = 0$ and $u = 1$ as a function of the duration τ of the trajectories. We find that the *average* change in system entropy $\langle\Delta S_{\text{sys}}\rangle$ is independent of the driving velocity u , because it depends only on the variance of the initial and final Gaussian distributions but not on their centers. For long trajectories, i.e., large τ , it approaches the constant value

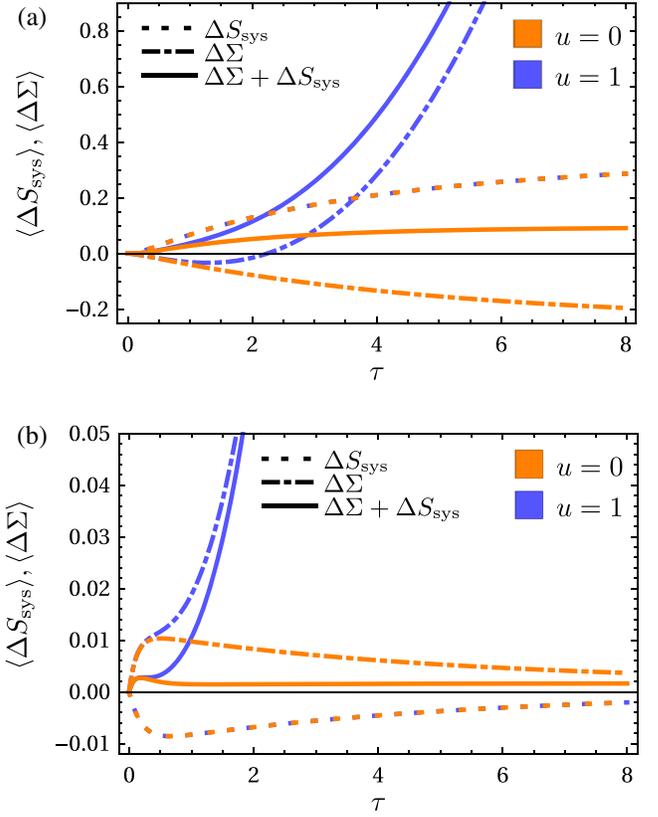


FIG. 2. Average contributions to irreversibility $\langle\Delta\Sigma\rangle$ and average change in system entropy $\langle\Delta S_{\text{sys}}\rangle$ as a function of the observation time τ . We compare the cases of a particle trapped in a static ($u = 0$, orange lines) or moving ($u = 1$, blue lines) harmonic potential having an initial distribution which is Gaussian with variance (a) $c_0^2 = k_B T/k = D\gamma/k$ and (b) $c_0^2 = (1/k)\{k_B T + [(\gamma D_a)/(1 + (k\tau_a/\gamma))]\}$. Parameter values are $k = 0.1$, $\gamma = 1$, $\tau_a = 0.2$, $D = D_a = 0.2$. Note that the blue and orange lines for the change in system entropy are on top of each other in both figures, as $\langle\Delta S_{\text{sys}}\rangle$ does not depend on u .

$$\lim_{\tau \rightarrow \infty} \langle\Delta S_{\text{sys}}\rangle = \frac{k_B}{2} \ln \left[1 + \frac{D_a}{D} \frac{1}{1 + \frac{k\tau_a}{\gamma}} \right]. \quad (75)$$

In contrast, the irreversibility measures $\langle\Delta\Sigma\rangle$ for the static trap $u = 0$ and the moving trap $u = 1$ are similar only during a short transient phase but then become qualitatively different. In the static trap, $\langle\Delta\Sigma\rangle$ becomes constant at large times τ ,

$$\lim_{\tau \rightarrow \infty} \langle\Delta\Sigma\rangle|_{u=0} = -\frac{k_B}{2} \frac{(1 - \frac{k\tau_a}{\gamma}) + 2\frac{D}{D_a}(1 + \frac{k\tau_a}{\gamma}) \left(1 + \sqrt{1 + \frac{D_a}{D}}\right)}{\left(1 + \sqrt{1 + \frac{D_a}{D}}\right)^2 \left(\frac{k\tau_a}{\gamma} + \sqrt{1 + \frac{D_a}{D}}\right)^2} \left(\frac{D_a}{D}\right)^2, \quad (76)$$

in accordance with the system reaching a currentless equilibriumlike steady state. In the moving trap, the particle

is transported continuously by permanent dissipation of energy so that $\langle \Delta \Sigma \rangle$ increases with the length τ of the trajectories. For large τ , the growth rate is given by the ensemble average of the time-averaged production rate $\sigma = \lim_{\tau \rightarrow \infty} (1/\tau) \int_0^\tau dt \sigma_\tau(t)$, where $\sigma_\tau(t)$ is defined in Eq. (48a). Explicitly, we obtain

$$\langle \sigma \rangle = \lim_{\tau \rightarrow \infty} \frac{\langle \Delta \Sigma \rangle}{\tau} = k_B \frac{u^2}{D + D_a}. \quad (77)$$

We thus see that the total “irreversibility production” $\langle \Delta \Sigma + \Delta S_{\text{sys}} \rangle$ in the moving trap grows extensively with the observation time as it would without active fluctuations, too. The growth rate, however, is *reduced* by the additional presence of active fluctuations, as the combined environment can be interpreted to have a higher “effective temperature.” Remarkably, the average growth rate (77) is independent of the relevant system timescales, i.e., the correlation time τ_a of active fluctuations and the relaxation time γ/k in the harmonic potential.

In the static trap $u = 0$, we find that no such extensive growth occurs for $\langle \Delta \Sigma + \Delta S_{\text{sys}} \rangle$. Hence, the system reaches a steady state, which appears equilibriumlike from the viewpoint of particle trajectories without access to the microscopic processes generating the active driving. Nevertheless, we observe $\lim_{\tau \rightarrow \infty} \langle \Delta \Sigma + \Delta S_{\text{sys}} \rangle > 0$ regardless of the initial distribution $p_0(x_0)$. For the case of a Gaussian with “thermal variance” $c_0^2 = k_B T/k$, the relevant results are given in Fig. 2 and Eqs. (75) and (76). Both $\lim_{\tau \rightarrow \infty} \langle \Delta S_{\text{sys}} \rangle$ and $\lim_{\tau \rightarrow \infty} \langle \Delta \Sigma \rangle$ depend on just two dimensionless parameters: the ratio of the two noise intensities D_a/D and the ratio of the two system timescales $[(k\tau_a)/\gamma]$. The limits of large and small correlation time and noise amplitudes can be easily computed and present no difficulties. As an interesting example, we consider the case of vanishing correlation time of the active fluctuations,

$$\lim_{\tau_a \rightarrow 0} \lim_{\tau \rightarrow \infty} \langle \Delta \Sigma + \Delta S_{\text{sys}} \rangle|_{u=0} = \frac{k_B}{2} \ln \left[1 + \frac{D_a}{D} \right] - \frac{k_B}{2} \frac{D_a}{D + D_a}. \quad (78)$$

This expression is exactly the entropy which would be produced by a thermal white-noise process with diffusion constant $D + D_a$ relaxing from an initial Gaussian distribution with variance $D\gamma/k = k_B T/k$ to its equilibrium state, a Gaussian with variance $(D + D_a)\gamma/k$.

It is quite obvious that for a variance $c_0^2 = k_B T/k$ of the initial Gaussian distribution, there must be some change in system entropy and a build up of irreversibility because the Gaussian particle distribution approached at long times has a larger variance $(1/k)\{(k_B T) + [(\gamma D_a)/(1 + k\tau_a/\gamma)]\}$ [see Eq. (D2d)]. But even if we start with the initial variance $c_0^2 = (1/k)\{(k_B T) + [(\gamma D_a)/(1 + (k\tau_a/\gamma))]\}$, we find a positive $\langle \Delta \Sigma \rangle$ while reaching the steady state at large τ ,

$$\lim_{\tau \rightarrow \infty} \langle \Delta \Sigma \rangle|_{u=0} = \frac{k_B \frac{k\tau_a}{\gamma} \left(\frac{D_a}{D}\right)^2}{\left(1 + \sqrt{1 + \frac{D_a}{D}}\right)^2 \left(\frac{k\tau_a}{\gamma} + \sqrt{1 + \frac{D_a}{D}}\right)^2}. \quad (79)$$

The origin of this positive contribution is our choice of *independent* initial conditions in the form of a product density $p_0(x_0, \eta_0) = p_0(x_0)p_s(\eta_0)$ [see also the discussion of Eq. (35)]. Hence, during the initial transient, there must be an irreversible build up of correlations between the particle and the active fluctuations becoming manifest in a nonzero total $\langle \Delta \Sigma + \Delta S_{\text{sys}} \rangle$. It turns out that the associated change in $\langle \Delta \Sigma + \Delta S_{\text{sys}} \rangle$ is nonmonotonic [see Fig. 2(b)], indicating that the variance of the particle distribution departs over some (transient) time period, even though the initial variance $c_0^2 = (1/k)\{(k_B T) + [(\gamma D_a)/(1 + (k\tau_a/\gamma))]\}$ is identical to the final one. In order to obtain a vanishing average $\langle \Delta \Sigma + \Delta S_{\text{sys}} \rangle$, we would have to start from a joint stationary state $p_s(x_0, \eta_0)$ of particle positions and active fluctuations instead of a factorized one. It can even be shown that forward and backward paths are equally likely in that case (see Ref. [100]), implying that $\Delta \Sigma[\bar{x}] + \Delta S_{\text{sys}}(x_0, x_\tau) = 0$ holds already on the level of individual trajectories. It is unsolved, however, if the latter property is generic for stationary states of trapped active particles (in the absence of symmetry-breaking forces), or if it is specific to the harmonic trap, to the Ornstein-Uhlenbeck realization of the active fluctuations, or to the dimensionality one.

We emphasize again that these correlations are the very reason that the irreversibility measure $\Delta \Sigma$ is nonadditive [see also the discussion below Eq. (45)]. The curves in Fig. 2 apply only to trajectories evolving over the complete time interval $[0, \tau]$. We cannot split a trajectory at an intermediate time, calculate the individual $\Delta \Sigma$ for the two parts of the trajectory from Eq. (42b), and then add them up to obtain $\Delta \Sigma$ for the full time interval because the “initial” state for the second part will inevitably depend on correlations between particle and active fluctuations accumulated during the first part. Such correlations are not taken into account in Eq. (42b) which is based on the assumption of an initial product state.

B. Fluctuation theorem

To illustrate the integral fluctuation theorem (44) satisfied by $\Delta \Sigma + \Delta S_{\text{sys}}$, we show in Fig. 3 the probability densities for the path probability ratio (43) for the same two situations of a static and a moving harmonic trapping potential already analyzed in Fig. 2(b). The probability densities are obtained from simulating 10^5 trajectories of length $\tau = 1$, well within the transient regime of the system evolution (see Fig. 2). While the distribution for $\exp[-(\Delta \Sigma + \Delta S_{\text{sys}})/k_B]$ is almost symmetric about 1 in the static trap $u = 0$, the most probable value lies visibly below 1 for the moving trap $u = 1$, indicating that trajectories with a positive $\Delta \Sigma + \Delta S_{\text{sys}}$ are more likely

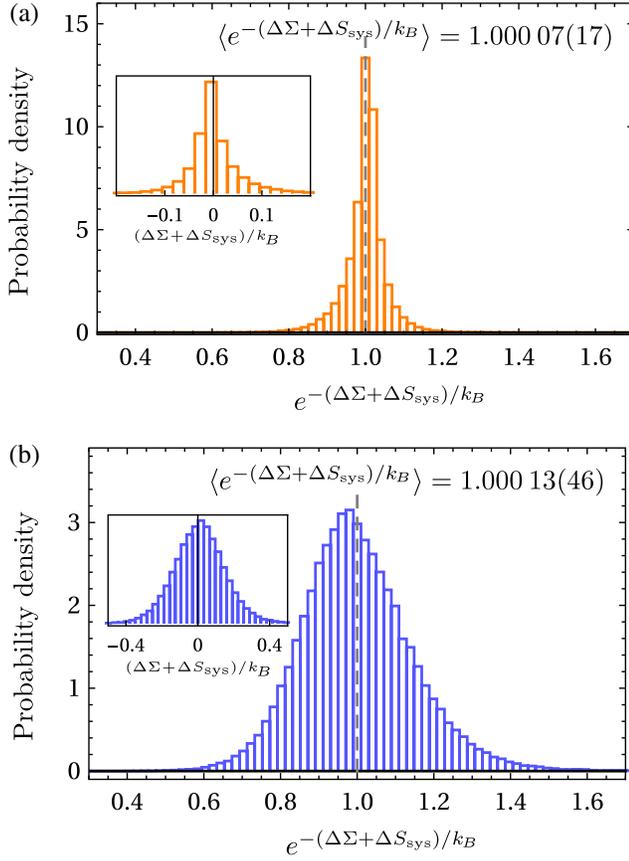


FIG. 3. Probability densities of the path probability ratio (43) for a harmonically trapped particle (a) in a static trap $u = 0$ and (b) in a moving trap $u = 1$. The densities are obtained from numerically simulating 10^5 sample trajectories of duration $\tau = 1$. Parameters are $k = 0.1$, $\gamma = 1$, $\tau_a = 0.2$, $D = D_a = 0.2$, $c_0^2 = (\gamma/k)\{D + [(D_a)/(1 + k\tau_a/\gamma)]\}$. The insets show the corresponding densities of $\Delta S_{\text{tot}} = \Delta\Sigma + \Delta S_{\text{sys}}$. The average values $\langle\Delta\Sigma\rangle$, $\langle\Delta S_{\text{sys}}\rangle$, and $\langle\Delta\Sigma + \Delta S_{\text{sys}}\rangle$ as obtained from the numerical simulations are consistent with the corresponding theoretical predictions. (a) $u = 0$, simulation $\langle\Delta\Sigma\rangle = 0.010(4)$, $\langle\Delta S_{\text{sys}}\rangle = -0.0090(13)$, $\langle\Delta\Sigma + \Delta S_{\text{sys}}\rangle = 0.00139(17)$; $u = 0$, theory $\langle\Delta\Sigma\rangle = 0.00975$, $\langle\Delta S_{\text{sys}}\rangle = -0.00828$, $\langle\Delta\Sigma + \Delta S_{\text{sys}}\rangle = 0.00148$. (b) $u = 1$, simulation $\langle\Delta\Sigma\rangle = 0.017(4)$, $\langle\Delta S_{\text{sys}}\rangle = -0.0073(13)$, $\langle\Delta\Sigma + \Delta S_{\text{sys}}\rangle = 0.0104(5)$; $u = 1$, theory $\langle\Delta\Sigma\rangle = 0.01883$, $\langle\Delta S_{\text{sys}}\rangle = -0.00828$, $\langle\Delta\Sigma + \Delta S_{\text{sys}}\rangle = 0.01056$.

than those with a negative value (see insets in Fig. 2). The sample mean for the path probability ratio lies well within 1 standard deviation of 1 in both cases, in accordance with the exact result (44).

C. Mutual information

The quantities $\Delta\Sigma$ and ΔS_{sys} analyzed in the previous two sections as a measure for the irreversibility in the system evolution depend only on the particle trajectories but not on the realizations of the active fluctuations, and

thus are readily accessible in experiments and simulations. The specific role of the active fluctuations, on the other hand, is nicely captured by the splitting of irreversibility, i.e., the log ratio of path probabilities, into total conditional entropy production $\Delta S_{\text{tot}}^\pm$ and mutual information $\Delta\mathcal{I}_\pm$, as described Sec. VA for an active bath (+ sign) and in Sec. VB for self-propulsion (− sign). Since for the present example of a particle trapped in a harmonic potential we have analytical expressions at hand for the combined propagator of particle position and active fluctuations (see Appendix D), we can calculate the conditional entropy production from Eqs. (56) and (62) and the mutual information from Eqs. (61) and (67) explicitly. Only the quantity $\ln p(\mathbf{x}_\tau|\bar{\boldsymbol{\eta}})$ is not easily accessible because it is conditioned on the full realization $\bar{\boldsymbol{\eta}}$, such that we add it to the change in mutual information in the form of the change in (conditional) system entropy $\Delta S_{\text{sys}}^{\bar{\boldsymbol{\eta}}}$ [see Eq. (53)]. Note that $\Delta S_{\text{sys}}^{\bar{\boldsymbol{\eta}}}$ is nonextensive with τ , and thus only a small correction to $\Delta\mathcal{I}_\pm$ which becomes negligible for long times.

In Fig. 4, we show the average conditional entropy production $\langle\Delta S^\pm\rangle$ and the average mutual information $k_B\langle\Delta\mathcal{I}_\pm\rangle - \langle\Delta S_{\text{sys}}^{\bar{\boldsymbol{\eta}}}\rangle$ for an active bath [Fig. 4(a)] and active self-propulsion [Fig. 4(b)]. The system parameters are the same as before in Figs. 2 and 3; in particular, we again compare a static trap $u = 0$ with a moving trap $u = 1$. We can see that now in all cases, both $\langle\Delta S^\pm\rangle$ and $\langle\Delta\mathcal{I}_\pm\rangle$ grow linearly with time for large τ . This conforms nicely with our previous findings: For the conditional total entropy $\Delta S_{\text{tot}}^\pm$, the active fluctuations are treated like an external time-dependent forcing. Such a force is then naturally expected to produce entropy extensively. For the ensemble- and time-averaged rate of total entropy production $\langle\sigma_{\text{tot}}^\pm\rangle = \lim_{\tau \rightarrow \infty} \langle\Delta S_{\text{tot}}^\pm\rangle/\tau$, we find

$$\langle\sigma_{\text{tot}}^+\rangle = k_B \left[\frac{(1 + \frac{D_a}{D})u^2}{D + D_a} + \frac{D_a}{D} \left(\frac{1}{1 + \frac{k\tau_a}{\gamma}} \right) \frac{1}{\tau_a} \right], \quad (80a)$$

$$\langle\sigma_{\text{tot}}^-\rangle = k_B \left[\frac{(1 + \frac{D_a}{D})u^2}{D + D_a} + \frac{D_a}{D} \left(\frac{1}{1 + \frac{k\tau_a}{\gamma}} \right) \frac{k}{\gamma} \right]. \quad (80b)$$

The first terms in these expressions contain the production rate of $\langle\Delta\Sigma\rangle$ from Eq. (77). The second terms quantify the additional contributions from the active fluctuations. They are balanced by the mutual information production rates $\langle\sigma_{\mathcal{I}}^\pm\rangle = \lim_{\tau \rightarrow \infty} k_B\langle\Delta\mathcal{I}_\pm\rangle/\tau$ (note that we include a factor of k_B in this definition of the rates so that they have units of entropy/time), which explicitly read

$$\langle\sigma_{\mathcal{I}}^+\rangle = k_B \frac{D_a}{D} \left[\frac{u^2}{D + D_a} + \left(\frac{1}{1 + \frac{k\tau_a}{\gamma}} \right) \frac{1}{\tau_a} \right], \quad (81a)$$

$$\langle\sigma_{\mathcal{I}}^-\rangle = k_B \frac{D_a}{D} \left[\frac{u^2}{D + D_a} + \left(\frac{1}{1 + \frac{k\tau_a}{\gamma}} \right) \frac{k}{\gamma} \right]. \quad (81b)$$

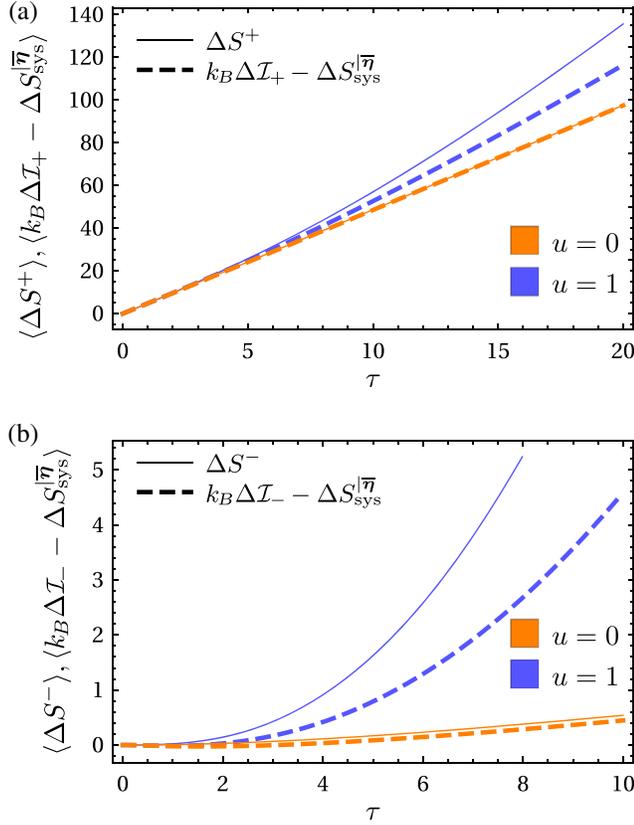


FIG. 4. Average mutual information and conditional entropy production as a function of the observation time τ for a particle trapped in a static ($u = 0$, orange lines) or moving ($u = 1$, blue lines) harmonic potential. Parameters are again $k = 0.1$, $\gamma = 1$, $\tau_a = 0.2$, $D = D_a = 0.2$. (a) Interpretation of the active fluctuations as an active bath; i.e., they are considered being even under time reversal. (b) Interpretation of the active fluctuations as self-propulsion; i.e., they are considered being odd under time reversal. Note that even for $u = 0$ (orange lines), the curves for $\langle \Delta S^+ \rangle$, $k_B \langle \Delta \mathcal{I}_+ \rangle - \langle \Delta S_{\text{sys}}^{\overline{\eta}} \rangle$ and $\langle \Delta S^- \rangle$, $k_B \langle \Delta \mathcal{I}_- \rangle - \langle \Delta S_{\text{sys}}^{\overline{\eta}} \rangle$, respectively, are not identical, but just appear very similar on the shown scale.

In total, we therefore find in both cases that $\langle \sigma \rangle = \langle \sigma_{\text{tot}}^{\pm} \rangle - \langle \sigma_{\mathcal{I}}^{\pm} \rangle$ holds, where $\langle \sigma \rangle$ is given in Eq. (77).

As we discuss in Sec. IV A, the two interpretations of the active fluctuations as active bath or as self-propulsion mechanism correspond to measuring the mutual information with respect to, respectively, even or odd time reversal of the active forcing so that the rate of “mutual information production” is different in the two cases; compare Eqs. (81a) and (81b). Their difference reads

$$\langle \sigma_{\mathcal{I}}^+ \rangle - \langle \sigma_{\mathcal{I}}^- \rangle = k_B \frac{D_a}{D} \left(\frac{1}{1 + \frac{k\tau_a}{\gamma}} \right) \left(\frac{1}{\tau_a} - \frac{k}{\gamma} \right). \quad (82)$$

We conclude that for the active bath, the relevant timescale the trajectory duration τ is measured against is the

correlation time of fluctuations τ_a , whereas it is the system’s relaxation time in the harmonic potential γ/k for the self-propelled particle.

VII. CONCLUSIONS AND DISCUSSION

Our present work contributes to assessing the out-of-equilibrium character of active matter [28,29,42,50,51,61,122,123] based on its observable dynamical behavior. Having in mind that in a typical experiment the central observables are particle trajectories, we quantify irreversibility in active matter systems based on particle trajectories alone, without resolving the microscopic mechanisms and associated dissipation underlying the active fluctuations which drive particle motion. For the thermodynamiclike features emerging from the dynamical behavior, it is presumably of little relevance what kind of microscopic processes dissipate how much entropy in generating the active fluctuations in the system (e.g., via self-propulsion), as long as these processes are not altered due to the motion of the particle. We are therefore interested in how (ir)reversible a specific particle trajectory is, out of the set of all possible trajectories which can be generated by the combined influence of thermal and active fluctuations, but not in how (ir)reversible the processes which underly the active fluctuations are. In that spirit, we treat the active fluctuations as an active (nonthermal) bath the particle is in contact with in addition to the thermal bath, with statistical properties which are completely independent of the internal dynamical state of the particle.

We implement this active bath by following the common approach to include stochastic active forces into the equations of motion, which emulate the directional persistence and the nonequilibrium character of the active fluctuations [124]. A particularly successful description of active fluctuations along these lines consists in a colored noise model, more specifically, a Gaussian Ornstein-Uhlenbeck process [7,15–18,21–24,28,42,49–61]. For this model class, we calculate the exact expression for the probability density of particle trajectories by integrating over all possible realizations of the active fluctuations [see Eq. (38)]. Since the colored noise renders the particle’s dynamics non-Markovian, the standard Onsager-Machlup path integral [84–87] cannot be applied directly to obtain the path probabilities. While expressions for a single colored noise source exist in the literature [105,106], here we derive for the first time the path weight for the superposition of colored Ornstein-Uhlenbeck noise (the active fluctuations) and white thermal noise.

Building on this result, we then relate the probabilities of time-forward and time-backward processes and establish an integral fluctuation theorem [32–35] for their log ratio $\Delta\Sigma$, a functional over the forward particle trajectory [see Eqs. (44) and (42b)] which quantifies the time (ir)reversibility of the particle dynamics. From the integral fluctuation theorem, we directly obtain a corresponding

second-law-like relation for $\Delta\Sigma$ [see Eq. (45)]. With $\Delta\Sigma$, we therefore provide an explicit expression to (exactly) evaluate the fluctuation theorem for a Brownian particle in contact not only with a thermal bath but at the same time also with an active nonequilibrium bath. In particular, it applies to trajectories of arbitrary finite duration. We expect that it can be tested experimentally with state-of-the-art micro(fluidic) technology for biological or synthetic active matter systems, as used, e.g., in Refs. [5,22–24,127–131]. We point out again that our explicit results for the path weight and the fluctuation theorem are based on the assumption that the active fluctuations are Gaussian with exponential correlations in time generated by an Ornstein-Uhlenbeck process.

Like in the case of usual Brownian motion in contact with a thermal bath only, the path probability ratio (42a) is equal to the identity if there are no external forces acting on the particle. Therefore, for $\mathbf{f} = 0$ any time-backward trajectory is equally likely to occur as its time-forward twin so that the particle dynamics looks reversible and equilibriumlike. Indeed on a coarse-grained timescale (beyond τ_a) they appear very similar to free Brownian diffusion in thermal equilibrium. Accordingly, in the absence of external forces, the probability ratio of particle trajectories does not reveal the nonequilibrium nature of the system, even though the whole system is out of equilibrium due to the active fluctuations. This observation should be true for any stationary (and unbiased) nonequilibrium bath, not just the Ornstein-Uhlenbeck implementation considered here. In order to detect the irreversibility connected with the active fluctuations, we would have to resolve the corresponding d.o.f. and analyze their behavior under time reversal [41,71,72].

However, as we can see from comparing Eqs. (19) and (42b), for nonvanishing forces \mathbf{f} , the irreversibility measure $\Delta\Sigma$ is distinctively different from the entropy production of purely Brownian motion because it contains the nonlocal memory kernel $\Gamma_\tau(t, t')$. Driving the particle by an external force thus reveals the non-Markovian and nonequilibrium character of the environment; even in the limit $\tau_a \rightarrow 0$ of δ -correlated active fluctuations the kernel $\Gamma_\tau(t, t')$ yields a nontrivial contribution (see also Appendix C), which renders $\Delta\Sigma$ different from the entropy production in the actual thermal bath. We leave for future exploration how this observation that external forces may reveal the non-equilibrium character of the environment may be used to probe properties of the active bath. Preliminary results indicate that the external force \mathbf{f} has to be nonlinear or time dependent, as a simple linear \mathbf{f} leads to $\Delta\Sigma = 0$ already on the level of individual trajectories [100]; see also Ref. [28].

Our irreversibility measure $\Delta\Sigma$ from Eq. (42) quantifies the combined “dissipation” into the thermal and the active bath, which occurs along a (time-forward) particle trajectory $\bar{\mathbf{x}}$ but cannot be interpreted easily as entropy production or dissipated heat. However, if we keep track of a

specific realization $\bar{\boldsymbol{\eta}}$ of the active noise as a fluctuating force affecting the particle dynamics, we find that $\Delta\Sigma$ can be split into two parts which have a direct physical interpretation: the usual entropy production in the thermal environment and a complementary dissipative component in the active bath, which is expressed as the difference of pathwise mutual information $\Delta\mathcal{I}$ accumulated along the time-forward process $(\bar{\mathbf{x}}, \bar{\boldsymbol{\eta}})$ versus its time-backward twin process [see Eqs. (58) and (64)]. This partition of $\Delta\Sigma$ is valid for any particle trajectory $\bar{\mathbf{x}}$ and any realization of the active noise fluctuation $\bar{\boldsymbol{\eta}}$ but with process-dependent amounts of dissipation in the two baths.

All these general results and interpretations are independent of how we choose the active fluctuations to behave under time reversal, even [Secs. IV A 1 and V A] or odd [Secs. IV A 2 and V B]. The quantitative details, however, are different; compare, e.g., Eq. (61) with Eq. (67). These quantitative differences are a consequence of the different amounts of heat exchanged with the thermal environment along a displacement $d\mathbf{x}$ in the two cases; see Eqs. (25) and (29). When we interpret the active fluctuations $\boldsymbol{\eta}(t)$ as an active environment the particle is moving through, the viscous friction forces from the thermal bath which balance these active fluctuations are included in the heat exchange. In contrast, they are not counted as contributing to heat in the case of self-propulsion because self-propelled motion occurs without external forces; i.e., on the coarse-grained level of description based on particle d.o.f., self-propulsion appears to be force- and thus dissipation-free. We argue in Sec. IV A that the interpretation of the active fluctuations $\boldsymbol{\eta}(t)$ as an active environment requires $\boldsymbol{\eta}(t)$ to be even under time reversal, while their interpretation as self-propulsion corresponds to $\boldsymbol{\eta}(t)$ being odd [109].

In any case, the corresponding stochastic energetics for active systems related to displacements of the particles, i.e., the definition of heat and work based on particle trajectories, appears as a natural and consistent generalization of Sekimoto’s stochastic energetics of passive Brownian motion [31] (see Sec. IV A), with a clear-cut connection to the irreversibility measure $\Delta\Sigma$ as obtained from the path probability ratio. The consistency and unambiguity of these results rely on fully incorporating the thermal fluctuations from the equilibrium heat bath into our description [cf. Eqs. (5)]. A number of previous works [28,42,50,51,122,123] attempted to assess irreversibility in active particle systems and its connection to heat and entropy production using the model class (5)–(7), including, however, only viscous friction effects from the heat bath and neglecting thermal fluctuations [$D = 0$ in Eq. (5) in our notation]. In this way, they disregarded an essential part of the energy exchange between system and heat bath [42], leading to ambiguities in the definition of heat and entropy production. Likewise, no consensus could be reached on the thermodynamic interpretation of the path probability ratio in terms of entropy production (see Ref. [42] for a

clear and enlightening summary, as well as the comment [122] and the corresponding reply [123]). We believe that here we resolve these problems in a unique and consistent way by incorporating the thermal fluctuations in our analysis [see model (5)–(7)]. In particular, this leads to a clear interpretation of the path probability ratio in terms of entropy production in the thermal bath (connected to the actual heat dissipated in the bath) and mutual information “production” with the active bath [see Eqs. (58) and (64)].

In the present work, our focus is on deriving the path probability ratio $\tilde{\mathfrak{p}}[\underline{\mathbf{x}}|\underline{\mathbf{x}}_0]/\mathfrak{p}[\underline{\mathbf{x}}|\underline{\mathbf{x}}_0]$ as a measure of irreversibility in active matter systems and on establishing its interpretation in thermodynamic terms. As a fundamental concept in stochastic thermodynamics, it contains essential information about the system and entails many interesting implications. With the fluctuation theorems and the second-law-like relations, we derive the most immediate ones here. Further potential applications of our results include the exploration of features in $\Delta\Sigma$ and $\Delta\mathcal{I}$ which are characteristic for the different phases of active matter [20], the analysis of linear response under small external perturbations and its potential for probing properties of the active bath [132], the connection between violations of the fluctuation-response relation and irreversibility [132,133], the characterization of universal statistics of infima, stopping times, and passage probabilities of $\Delta\Sigma$ and $\Delta\mathcal{I}$ based on the properties of $\tilde{\mathfrak{p}}[\underline{\mathbf{x}}|\underline{\mathbf{x}}_0]/\mathfrak{p}[\underline{\mathbf{x}}|\underline{\mathbf{x}}_0]$ [134], the connection of $\Delta\Sigma$ and $\Delta\mathcal{I}$ to the arrow of time [116,135] in active matter systems, and universal properties in the efficiency fluctuations of stochastic heat engines operating between active baths [96,136,137].

Here, we analyze the model (5), (6) in great detail from the viewpoint of a Brownian particle moving under the influence of active fluctuations, which are represented by the Ornstein-Uhlenbeck process (7). However, most of our results, in particular, the path weight (38) and the associated integral fluctuation theorem (44), as well as its formulation in information-theoretic terms in Sec. V, are a direct consequence of the mathematical structure of the model and are therefore valid in a much broader physical context. In principle, our results can be applied to any Brownian dynamics that is driven by an additional Gaussian Ornstein-Uhlenbeck process. For instance, in Ref. [138], an information-theoretic analysis similar to ours has been conducted for a colored-noise-driven Brownian model. Other examples are Brownian motion in a harmonic trap with fluctuating location [78,139–141] and thermodynamic nonequilibrium processes using external (artificial) colored noise sources [117,142].

ACKNOWLEDGMENTS

S. B. and R. E. acknowledge financial support from the Swedish Research Council (Vetenskapsrådet) under the Grants No. 621-2013-3956, No. 638-2013-9243, and No. 2016-05412. L. D. acknowledges financial support

by the Deutsche Forschungsgemeinschaft under Grant No. RE 1344/10-1 and within the Research Unit FOR 2692 under Grant No. 397303734, as well as by the Stiftung der Deutschen Wirtschaft. L. D. also thanks Nordita for the hospitality and support during an internship in 2016. We moreover thank Erik Aurell, Aykut Argun, Giovanni Volpe, and Jan Wehr for illuminating discussions.

APPENDIX A: INTEGRATING OVER THE ACTIVE FLUCTUATIONS η_t

Here we show that the path integral (36) evaluates to expression (38). Gaussian path integrals of this form are ubiquitous in field theories [101,102] with well-established results. Using the abbreviation $\mathbf{w}_t = [(\sqrt{2D_a})/(2D)](\dot{\mathbf{x}}_t - \mathbf{v}_t)$, the relevant terms in Eq. (36) involving the active noise variable η_t become

$$\int \mathcal{D}\bar{\eta} \exp \left[-\frac{1}{2} \int_0^\tau dt \int_0^\tau dt' \eta_t^\top \hat{V}_\tau(t, t') \eta_{t'} + \int_0^\tau dt \eta_t^\top \mathbf{w}_t \right]. \quad (\text{A1})$$

Completing the square, we wish to write this as

$$\int \mathcal{D}\bar{\eta} \exp \left\{ \frac{1}{2} \int_0^\tau dt \int_0^\tau dt' [\mathbf{w}_t^\top \Gamma_\tau(t, t') \mathbf{w}_{t'} - (\eta_t + \boldsymbol{\varepsilon}_t)^\top \hat{V}_\tau(t, t') (\eta_{t'} + \boldsymbol{\varepsilon}_{t'})] \right\}, \quad (\text{A2})$$

where $\boldsymbol{\varepsilon}_t$ and $\Gamma_\tau(t, t')$ are yet unknown functions. We can then shift the active noise histories and integrate over $\eta_t' = \eta_t + \boldsymbol{\varepsilon}_t$ instead of η_t . Since the path integral effectively integrates over all possible states of η_t from $-\infty$ to $+\infty$ for any point in time t , this shift of trajectories does not alter the domain of integration. Moreover, the Jacobian associated with the transformation is the identity. Performing the remaining functional integral over $\bar{\eta}'$, expression (A2) then reduces to

$$(\text{Det} \hat{V}_\tau)^{-1/2} \exp \left[\frac{1}{2} \int_0^\tau dt \int_0^\tau dt' \mathbf{w}_t^\top \Gamma_\tau(t, t') \mathbf{w}_{t'} \right], \quad (\text{A3})$$

assuming proper normalization of the integration measure $\mathcal{D}\bar{\eta}'$. Absorbing the path-independent functional determinant $(\text{Det} \hat{V}_\tau)^{-1/2}$ into the normalization, we obtain the path weight stated in Eq. (38).

However, we still have to show that $\Gamma_\tau(t, t')$ is the operator inverse of $\hat{V}_\tau(t, t')$; i.e., we have to verify Eq. (39). Requiring equality of Eqs. (A1) and (A2), we find that

$$\int_0^\tau dt \eta_t^\top \mathbf{w}_t \stackrel{!}{=} \frac{1}{2} \int_0^\tau dt \int_0^\tau dt' [\mathbf{w}_t^\top \Gamma_\tau(t, t') \mathbf{w}_{t'} - \boldsymbol{\varepsilon}_t^\top \hat{V}_\tau(t, t') \boldsymbol{\varepsilon}_{t'} - 2\eta_t^\top \hat{V}_\tau(t, t') \boldsymbol{\varepsilon}_{t'}]. \quad (\text{A4})$$

Note that we use $\hat{V}_\tau(t, t') = \hat{V}_\tau(t', t)$. We now observe that the term on the left-hand side as well as the last term on the right-hand side are of first order in $\boldsymbol{\eta}_t$, whereas the remaining terms on the right-hand side do not contain $\boldsymbol{\eta}_t$. Therefore, these two types of expressions must cancel individually, i.e.,

$$\int_0^\tau dt \int_0^\tau dt' \boldsymbol{\epsilon}_t^\top \hat{V}_\tau(t, t') \boldsymbol{\epsilon}_{t'} = \int_0^\tau dt \int_0^\tau dt' \boldsymbol{w}_t^\top \Gamma_\tau(t, t') \boldsymbol{w}_{t'}, \quad (\text{A5a})$$

$$\int_0^\tau dt \boldsymbol{\eta}_t^\top \boldsymbol{w}_t = - \int_0^\tau dt \boldsymbol{\eta}_t^\top \int_0^\tau dt' \hat{V}_\tau(t, t') \boldsymbol{\epsilon}_{t'}. \quad (\text{A5b})$$

From Eq. (A5b), we immediately infer

$$\boldsymbol{w}_t = - \int_0^\tau dt' \hat{V}_\tau(t, t') \boldsymbol{\epsilon}_{t'}. \quad (\text{A6})$$

Substituting this result into the left-hand side of Eq. (A5a), we obtain

$$- \int_0^\tau dt \boldsymbol{\epsilon}_t^\top \boldsymbol{w}_t = \int_0^\tau dt \boldsymbol{w}_t^\top \int_0^\tau dt' \Gamma_\tau(t, t') \boldsymbol{w}_{t'}, \quad (\text{A7})$$

implying

$$\boldsymbol{\epsilon}_t = - \int_0^\tau dt' \Gamma_\tau(t, t') \boldsymbol{w}_{t'}. \quad (\text{A8})$$

This intermediate result gives the shift $\boldsymbol{\epsilon}_t$ required to complete the square in Eq. (A1) [see also Eq. (A2)] as a function of $\Gamma_\tau(t, t')$ and \boldsymbol{w}_t . Moreover, substituting Eq. (A8) back into Eq. (A6), we find

$$\boldsymbol{w}_t = \int_0^\tau dt' \int_0^\tau dt'' \hat{V}_\tau(t, t') \Gamma_\tau(t', t'') \boldsymbol{w}_{t''}, \quad (\text{A9})$$

which implies Eq. (39) and thus establishes that $\Gamma_\tau(t, t')$ is indeed the Green's function of the differential operator $\hat{V}_\tau(t, t')$.

APPENDIX B: CONSTRUCTION OF THE GREEN'S FUNCTION $\Gamma_\tau(t, t')$

In this Appendix, we construct the Green's function (40a) by solving its defining equation (39) with the differential operator $\hat{V}_\tau(t, t')$ from Eq. (37). Exploiting the ‘‘diagonal’’ structure of $\hat{V}_\tau(t, t')$, we can directly evaluate the integral and obtain

$$[\hat{V}(t) + \hat{V}_0(t) + \hat{V}_\tau(t)] \Gamma_\tau(t, t') = \delta(t - t'). \quad (\text{B1})$$

This is a linear second-order ordinary differential equation with a δ inhomogeneity. We calculate its solution in two steps. First, we compute the Green's function $\bar{\Gamma}(t, t')$ of the ordinary component $\hat{V}(t)$ with vanishing boundary conditions, i.e., $\bar{\Gamma}(0, t') = \bar{\Gamma}(\tau, t') = 0$. Then we add a solution $\Gamma_{0,\tau}(t, t')$ of the corresponding homogeneous problem $\hat{V}(t) \Gamma_{0,\tau}(t, t') = 0$ that fixes the two boundary terms. Their sum $\Gamma_\tau(t, t') = \bar{\Gamma}(t, t') + \Gamma_{0,\tau}(t, t')$ satisfies Eq. (B1) and thus gives the desired solution.

We can construct both these parts $\bar{\Gamma}(t, t')$ and $\Gamma_{0,\tau}(t, t')$ from the solution of the homogeneous problem associated with $\hat{V}_\tau(t, t)$, which reads [see Eq. (37)]

$$[-\tau_a^2 \partial_t^2 + (1 + D_a/D)] \Gamma(t) = 0. \quad (\text{B2})$$

We make an exponential ansatz $\Gamma(t) \sim e^{\lambda t}$ and easily obtain

$$\Gamma(t) = \alpha^+ e^{\lambda t} + \alpha^- e^{-\lambda t}, \quad \lambda = \frac{1}{\tau_a} \sqrt{1 + \frac{D_a}{D}}, \quad (\text{B3})$$

with constants α^\pm to be determined by the boundary conditions or the δ inhomogeneity.

There exists a standard recipe [103,104] for the construction of Green's functions of boundary value problems for ordinary differential equations, which we follow here to calculate $\bar{\Gamma}(t, t')$. Splitting the interval $[0, \tau]$ at $t = t'$, we write

$$\bar{\Gamma}(t, t') = \Theta(t' - t) \bar{\Gamma}_<(t, t') + \Theta(t - t') \bar{\Gamma}_>(t, t'), \quad (\text{B4})$$

where Θ is the Heaviside step function and both $\bar{\Gamma}_<(t, t')$ and $\bar{\Gamma}_>(t, t')$ satisfy the homogenous problem; i.e., they are of the form (B3) with constants α_\pm^\pm and α_\pm^\pm , respectively, to be determined. The constants are fixed by the boundary conditions,

$$\bar{\Gamma}_<(0, t') = 0 \quad \text{and} \quad \bar{\Gamma}_>(\tau, t') = 0, \quad (\text{B5a})$$

and the requirements that $\bar{\Gamma}$ is continuous at $t = t'$,

$$\bar{\Gamma}_<(t', t') = \bar{\Gamma}_>(t', t'), \quad (\text{B5b})$$

and has a jump of size $-1/\tau_a^2$ in its first derivative at $t = t'$,

$$[\partial_t \bar{\Gamma}_>(t, t') - \partial_t \bar{\Gamma}_<(t, t')]_{t=t'} = -\frac{1}{\tau_a^2}. \quad (\text{B5c})$$

The last two conditions ensure the desired δ discontinuity at $t = t'$ in the second derivative. Solving the resulting system of linear equations for α_\pm^+ , α_\pm^- , α_\pm^+ , and α_\pm^- , we obtain

$$\bar{\Gamma}_>(t, t') = \left(\frac{1}{2\tau_a^2\lambda} \right) \frac{e^{\lambda\tau}[e^{-\lambda(t-t')} - e^{-\lambda(t+t')}] + e^{-\lambda\tau}[e^{\lambda(t-t')} - e^{\lambda(t+t')}]}{e^{\lambda\tau} - e^{-\lambda\tau}} \quad (\text{B6})$$

and $\bar{\Gamma}_<(t, t') = \bar{\Gamma}_>(t', t)$. With Eq. (B4), we can thus write the Green's function of the ordinary component on the entire interval $[0, \tau]$ in the form

$$\bar{\Gamma}(t, t') = \left(\frac{1}{2\tau_a^2\lambda} \right) \frac{e^{-\lambda|t-t'|} - e^{-\lambda(t+t')} + e^{-\lambda(2\tau-|t-t'|)} - e^{-\lambda(2\tau-t-t')}}{1 - e^{-2\lambda\tau}}. \quad (\text{B7})$$

We can immediately check that indeed $\hat{V}(t)\bar{\Gamma}(t, t') = \delta(t-t')$ as well as $\bar{\Gamma}(0, t') = 0$ and $\bar{\Gamma}(\tau, t') = 0$. Moreover, we note that $\bar{\Gamma}(t, t')$ is symmetric in its arguments, $\bar{\Gamma}(t, t') = \bar{\Gamma}(t', t)$.

As announced earlier, we now add a homogeneous solution $\Gamma_{0,\tau}(t, t') = \alpha^+ e^{\lambda t} + \alpha^- e^{-\lambda t}$ to the ordinary inverse $\bar{\Gamma}(t, t')$. Since $\hat{V}(t)\Gamma_{0,\tau}(t, t') = 0$ and $\bar{\Gamma}(t, t')$ vanishes at both boundaries of the interval $[0, \tau]$, the differential equation (B1) for the sum $\Gamma_\tau(t, t') = \bar{\Gamma}(t, t') + \Gamma_{0,\tau}(t, t')$ becomes

$$\begin{aligned} \delta(t-\tau)[\tau_a^2 \partial_t \bar{\Gamma}_>(t, t')|_{t=\tau} + \alpha^+ \tau_a \kappa_+ e^{\lambda\tau} + \alpha^- \tau_a \kappa_- e^{-\lambda\tau}] \\ + \delta(t)[-\tau_a^2 \partial_t \bar{\Gamma}_<(t, t')|_{t=0} + \alpha^+ \tau_a \kappa_- + \alpha^- \tau_a \kappa_+] = 0, \end{aligned} \quad (\text{B8})$$

with $\kappa_\pm = 1 \pm \lambda\tau_a = 1 \pm \sqrt{1 + (D/D_a)}$. We require the two terms in square brackets to vanish individually and thus obtain two linear equations for the two unknowns α^+ and α^- . Solving them and combining the results for $\bar{\Gamma}(t, t')$ and $\Gamma_{0,\tau}(t, t')$ into $\Gamma_\tau(t, t') = \bar{\Gamma}(t, t') + \Gamma_{0,\tau}(t, t')$, we finally find the Green's function (40a).

APPENDIX C: LIMITING CASES FOR THE PATH WEIGHT

Here we analyze three relevant limiting cases of the path weight (38), namely, $D_a \rightarrow 0$ (usual Brownian particle without active fluctuations), $\tau_a \rightarrow 0$ (memoryless active fluctuations), and $D \rightarrow 0$ (no thermal fluctuations). All three limits reduce the Langevin equation (5) to simpler setups, for which the path probabilities are already known in the literature. We recover all these results when performing the respective limits in our general expression (38).

1. Usual Brownian motion $D_a \rightarrow 0$

Removing the effects of the active fluctuations from the system amounts to setting $\boldsymbol{\eta}(t) = 0$ or, equivalently, $D_a = 0$ in the equation of motion (5). Performing this limit for the path weight is straightforward: In this case, $\lambda = 1/\tau_a$ is well behaved, so that the term containing the Green's function simply drops out in Eq. (38). The remaining path

probability is just the standard Onsager-Machlup expression (16) for a Brownian particle in a thermal bath.

2. Vanishing correlation time $\tau_a \rightarrow 0$

In the limit $\tau_a \rightarrow 0$, the correlator (8) of the active noise approaches a δ distribution, i.e., $\langle \eta_i(t)\eta_j(t') \rangle \rightarrow \delta_{ij}\delta(t-t')$. This means that $\boldsymbol{\eta}(t)$ becomes just another Gaussian white noise with diffusion constant D_a . The equation of motion (5) thus contains two independent Gaussian white noises with zero mean and variance $2D$ and $2D_a$, respectively. Their sum is itself a Gaussian white noise with zero mean but variance $2(D + D_a)$. Therefore, we expect Eq. (38) to reduce to

$$\mathfrak{p}[\mathbf{x}|\mathbf{x}_0]_{\tau_a \rightarrow 0} \propto \exp \int_0^\tau dt \left[-\frac{(\dot{\mathbf{x}}_t - \mathbf{v}_t)^2}{4(D + D_a)} - \frac{\mathbf{V} \cdot \mathbf{v}_t}{2} \right] \quad (\text{C1})$$

as $\tau_a \rightarrow 0$.

We first note that λ diverges in this limit so that we have to take more care when analyzing the Green's function (40a). Since the limit is expressed more compactly as $\lambda \rightarrow \infty$, we rewrite all occurrences of τ_a in Eq. (40a) in terms of λ , D , and D_a . By definition [see Eqs. (40b) and (B3)], $\tau_a^2 = [1 + (D_a/D)]/\lambda^2$. Thus, the leading-order behavior of $\Gamma_\tau(t, t')$ is

$$\Gamma_\tau(t, t') \sim \left(\frac{\lambda}{2} \right) \frac{e^{-\lambda|t-t'|} - \frac{\kappa_-}{\kappa_+} [e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}]}{1 + \frac{D_a}{D}}. \quad (\text{C2})$$

We observe that

$$\lim_{\lambda \rightarrow \infty} \frac{\lambda}{2} e^{-\lambda|t-t'|} = \delta(t-t'), \quad (\text{C3a})$$

$$\lim_{\lambda \rightarrow \infty} \frac{\lambda}{2} e^{-\lambda t} \Theta(t) = \frac{1}{2} \delta(t), \quad (\text{C3b})$$

such that the exponentials become δ distributions as $\lambda \rightarrow \infty$:

$$\Gamma_\tau(t, t') \sim \frac{\delta(t-t') - \frac{\kappa_-}{2\kappa_+} [\delta(t+t') + \delta(2\tau-t-t')]}{1 + \frac{D_a}{D}}.$$

However, the last two δ distributions map either of the two times t and t' outside of the interval of integration when integrating over the other. Therefore, they do not contribute when integrating over both t and t' . We can see this also from partially integrating the corresponding term in Eq. (C2) with two test functions f and g ,

$$\begin{aligned} & \int_0^\tau dt \int_0^\tau dt' f(t)g(t') \frac{\lambda}{2} [e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}] \\ &= \int_0^\tau dt f(t) \left\{ g(t') \frac{1}{2} [-e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}] \right\} \Big|_{t'=0}^\tau \\ & \quad - \int_0^\tau dt' g(t') \frac{1}{2} [-e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}] \\ & \rightarrow 0 \quad \text{as } \lambda \rightarrow \infty. \end{aligned}$$

Therefore, $\Gamma_\tau(t, t') \sim \delta(t - t')/[1 + (D_a/D)]$ as $\tau_a \rightarrow 0$. Substituting this finding into the path weight (38) leads precisely to the expected limit (C1).

3. No thermal fluctuations $D \rightarrow 0$

The limit of vanishing thermal white noise is a little more involved. If we let $D \rightarrow 0$ in the equation of motion (5), we are left with a system driven by a single colored noise source. The resulting path probability density is known in the literature [105,106] and given by Eq. (41). However, it is not immediately obvious how we can obtain this result from Eq. (38) because both the prefactor $1/(4D)$ and the exponent λ diverge as $D \rightarrow 0$.

We first rewrite Eq. (41) so that the action can be expressed in the form of Eq. (38). To this end, we introduce the abbreviation $\mathbf{w}_t = \dot{\mathbf{x}}_t - \mathbf{v}_t$ in Eq. (41) and remember that we choose p_s to be the stationary distribution (35) of the colored noise. After partial integration of the $\dot{\mathbf{w}}_t^2$ term, we find

$$\begin{aligned} \mathfrak{p}[\mathbf{x}|\mathbf{x}_0]_{D \rightarrow 0} &\propto \exp \left\{ -\frac{1}{4D_a} \left[\int_0^\tau dt (-\tau_a^2 \mathbf{w}_t^\top \dot{\mathbf{w}}_t + \mathbf{w}_t^2) \right. \right. \\ & \quad \left. \left. + \tau_a^2 \dot{\mathbf{w}}_t^\top \mathbf{w}_t \Big|_0^\tau + \tau_a \mathbf{w}_t^2 \Big|_0^\tau + 2\tau_a^2 \mathbf{w}_0^2 \right] \right. \\ & \quad \left. - \int_0^\tau dt \frac{\nabla \cdot \mathbf{v}_t}{2} \right\}. \end{aligned} \quad (\text{C4})$$

We now show that the action

$$\mathcal{A}[\bar{\mathbf{x}}] = \frac{1}{4D} \int_0^\tau dt \int_0^\tau dt' \mathbf{w}_t^\top \left[\delta(t - t') - \frac{D_a}{D} \Gamma_\tau(t, t') \right] \mathbf{w}_{t'} \quad (\text{C5})$$

of the full path weight $\mathfrak{p}[\mathbf{x}|\mathbf{x}_0] \propto e^{-\mathcal{A}[\bar{\mathbf{x}}] - \int_0^\tau dt \nabla \cdot \mathbf{v}_t/2}$ as given in Eq. (38) reduces to the form of the action in Eq. (C4) in the limit $D \rightarrow 0$.

We rewrite the limit $D \rightarrow 0$ again as $\lambda \rightarrow \infty$; i.e., we express all occurrences of D in terms of λ , D_a , and τ_a according to $(D_a/D) = \lambda^2 \tau_a^2 - 1$ [see Eqs. (40b) and (B3)]. For the leading-order behavior of the Green's function $\Gamma_\tau(t, t')$, we then find

$$\Gamma_\tau(t, t') \sim \frac{e^{-\lambda|t-t'|} + [1 - \frac{2}{\tau_a \lambda} + \mathcal{O}(\lambda^{-2})][e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}]}{2\tau_a^2 \lambda}, \quad (\text{C6})$$

which implies

$$\begin{aligned} \mathcal{A}[\bar{\mathbf{x}}]_{D \rightarrow 0} &\sim \frac{1}{4D_a} \int_0^\tau dt \int_0^\tau dt' \mathbf{w}_t^\top \mathbf{w}_{t'} \left\{ (\lambda^2 \tau_a^2 - 1) \delta(t - t') \right. \\ & \quad \left. - \left(\frac{\lambda^3 \tau_a^2}{2} - \lambda \right) [e^{-\lambda|t-t'|} + e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}] \right. \\ & \quad \left. + [\lambda^2 \tau_a + \mathcal{O}(\lambda)] [e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}] \right\}. \end{aligned}$$

We divide our further analysis into three parts by writing $\mathcal{A}[\bar{\mathbf{x}}] \sim [1/(4D_a)](\mathcal{B}_1[\bar{\mathbf{x}}] + \mathcal{B}_2[\bar{\mathbf{x}}] + \mathcal{B}_3[\bar{\mathbf{x}}])$ with

$$\begin{aligned} \mathcal{B}_1[\bar{\mathbf{x}}] &= \int_0^\tau dt \int_0^\tau dt' \mathbf{w}_t^\top \mathbf{w}_{t'} \left[(\lambda^2 \tau_a^2 - 1) \delta(t - t') \right. \\ & \quad \left. - \left(\frac{\lambda^3 \tau_a^2}{2} - \lambda \right) e^{-\lambda|t-t'|} \right], \end{aligned} \quad (\text{C7a})$$

$$\begin{aligned} \mathcal{B}_2[\bar{\mathbf{x}}] &= - \int_0^\tau dt \int_0^\tau dt' \mathbf{w}_t^\top \mathbf{w}_{t'} \left(\frac{\lambda^3 \tau_a^2}{2} - \lambda \right) \\ & \quad \times [e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}], \end{aligned} \quad (\text{C7b})$$

$$\begin{aligned} \mathcal{B}_3[\bar{\mathbf{x}}] &= \int_0^\tau dt \int_0^\tau dt' \mathbf{w}_t^\top \mathbf{w}_{t'} [\lambda^2 \tau_a + \mathcal{O}(\lambda)] \\ & \quad \times [e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}], \end{aligned} \quad (\text{C7c})$$

Upon repeated partial integration using $\lambda e^{-\lambda t} = -\partial_t e^{-\lambda t}$ to remove powers of λ , and upon performing the $\lambda \rightarrow \infty$ limiting procedure for well-behaved terms, we find for the first part

$$\mathcal{B}_1[\bar{\mathbf{x}}] \sim \int_0^\tau dt [-\tau_a^2 \mathbf{w}_t^\top \dot{\mathbf{w}}_t + \mathbf{w}_t^2] + (\mathbf{w}_0^2 + \mathbf{w}_\tau^2) \frac{\lambda \tau_a^2}{2}.$$

Similarly, the second part reduces to

$$\mathcal{B}_2[\bar{\mathbf{x}}] \sim -(\mathbf{w}_0^2 + \mathbf{w}_\tau^2) \frac{\lambda \tau_a^2}{2} + \tau_a^2 (\mathbf{w}_\tau^\top \dot{\mathbf{w}}_\tau - \mathbf{w}_0^\top \dot{\mathbf{w}}_0).$$

For the third part, we first remark that the terms of order λ vanish here again under double time integrals because they become δ distributions that map one of the times outside the interval of integration, similar to the case we have for the limit of vanishing correlation time. Hence,

$$\begin{aligned}
\mathcal{B}_3[\bar{\mathbf{x}}] &\sim \int_0^\tau dt \int_0^\tau dt' \mathbf{w}_t^\top \mathbf{w}_{t'} \lambda^2 \tau_a [e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}] \\
&= \int_0^\tau dt \int_0^\tau dt' \mathbf{w}_t^\top \mathbf{w}_{t'} \lambda \tau_a \partial_{t'} [-e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}] \\
&= \int_0^\tau dt \mathbf{w}_t^\top \mathbf{w}_{t'} \lambda \tau_a [-e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}] \Big|_{t'=0}^\tau \\
&\quad - \int_0^\tau dt \int_0^\tau dt' \dot{\mathbf{w}}_t^\top \dot{\mathbf{w}}_{t'} \lambda \tau_a [-e^{-\lambda(t+t')} + e^{-\lambda(2\tau-t-t')}] \\
&\sim \int_0^\tau dt \mathbf{w}_t^\top \{ \mathbf{w}_\tau \tau_a \lambda [-e^{-\lambda(\tau+t)} + e^{-\lambda(\tau-t)}] \} \\
&\quad - \int_0^\tau dt \mathbf{w}_t^\top \{ \mathbf{w}_0 \tau_a \lambda [-e^{-\lambda t} + e^{-\lambda(2\tau-t)}] \} \\
&\sim \tau_a (\mathbf{w}_\tau^2 - \mathbf{w}_0^2) + 2\tau_a \mathbf{w}_0^2.
\end{aligned}$$

Combining these three results, we finally find

$$\begin{aligned}
\mathcal{A}[\bar{\mathbf{x}}]_{D \rightarrow 0} &= \frac{1}{4D_a} \left[\int_0^\tau dt (-\tau_a^2 \mathbf{w}_t^\top \dot{\mathbf{w}}_t + \mathbf{w}_t^2) \right. \\
&\quad \left. + \tau_a^2 (\mathbf{w}_\tau^\top \dot{\mathbf{w}}_\tau - \mathbf{w}_0^\top \dot{\mathbf{w}}_0) + \tau_a (\mathbf{w}_\tau^2 - \mathbf{w}_0^2) + 2\tau_a \mathbf{w}_0^2 \right], \quad (\text{C8})
\end{aligned}$$

which is identical to the path weight (C4).

APPENDIX D: DETAILS FOR THE HARMONICALLY TRAPPED PARTICLE

In this Appendix, we summarize some details and a few key steps behind the calculations for the Brownian particle in a harmonic potential from Sec. VI.

1. Dynamics

The equations of motion for the joint Markovian system of particle and active fluctuations combined from Eqs. (74) and (7) (for $d = 1$) read

$$\begin{pmatrix} \dot{x}_t \\ \dot{\eta}_t \end{pmatrix} = -\mathbf{A} \begin{bmatrix} x_t \\ \eta_t \end{bmatrix} - \begin{pmatrix} u \\ 0 \end{pmatrix} t + \mathbf{B} \begin{pmatrix} \xi_t \\ \zeta_t \end{pmatrix} \quad (\text{D1a})$$

with

$$\mathbf{A} = \begin{pmatrix} k/\gamma & -\sqrt{2D_a} \\ 0 & 1/\tau_a \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} \sqrt{2D} & 0 \\ 0 & 1/\tau_a \end{pmatrix}. \quad (\text{D1b})$$

Because of its Markovian character, all statistical properties of this joint system are encoded in the propagator $p(\mathbf{q}, t | \mathbf{q}_0, t_0)$, where $\mathbf{q} = (x, \eta)$, $\mathbf{q}_0 = (x_0, \eta_0)$ are ‘‘generalized coordinates’’ summarizing particle position and state

of the active fluctuations. This propagator represents the probability to be at ‘‘position’’ \mathbf{q} at time t when having been at \mathbf{q}_0 at an earlier time $t_0 < t$. We obtain its explicit form by solving the Fokker-Planck equation associated with Eq. (D1),

$$p(\mathbf{q}, t | \mathbf{q}_0, t_0) = \frac{e^{-\frac{1}{2}[\mathbf{q} - \boldsymbol{\mu}(t | \mathbf{q}_0, t_0)]^\top \mathbf{C}(t | t_0)^{-1} [\mathbf{q} - \boldsymbol{\mu}(t | \mathbf{q}_0, t_0)]}}{\sqrt{(2\pi)^2 \det \mathbf{C}(t | t_0)}}. \quad (\text{D2a})$$

Here, the expectation vector and covariance matrix conditioned on the initial time point are given by

$$\boldsymbol{\mu}(t | \mathbf{q}_0, t_0) = (t - \mathbf{A}^{-1})\mathbf{u} + e^{-(t-t_0)\mathbf{A}}[\mathbf{q}_0 - (t_0 - \mathbf{A}^{-1})\mathbf{u}], \quad (\text{D2b})$$

$$\mathbf{C}(t | t_0) = \mathbf{C}(\infty) - e^{-(t-t_0)\mathbf{A}}\mathbf{C}(\infty)e^{-(t-t_0)\mathbf{A}^\top} \quad (\text{D2c})$$

with $\mathbf{u} = (u, 0)$ and the stationary covariance matrix

$$\mathbf{C}(\infty) = \begin{pmatrix} \frac{\gamma}{k} \left[D + \frac{D_a}{1 + \frac{k\tau_a}{\gamma}} \right] & \sqrt{\frac{D_a}{2}} \frac{1}{1 + \frac{k\tau_a}{\gamma}} \\ \sqrt{\frac{D_a}{2}} \frac{1}{1 + \frac{k\tau_a}{\gamma}} & \frac{1}{2\tau_a} \end{pmatrix}. \quad (\text{D2d})$$

As we describe in the main text, we consider the situation in which initially the distribution of particle positions is independent of the active fluctuations (which are in their stationary state) so that the initial distribution for the joint system factorizes as $p_0(x_0, \eta_0) = p_0(x_0)p_s(\eta_0)$. We further assume that $p_0(x_0)$ is Gaussian with zero mean and variance c_0^2 , whereas $p_s(\eta_0)$ is given in Eq. (35). The initial probability density $p_0(x_0, \eta_0)$ is thus also Gaussian with mean $\boldsymbol{\mu}(0) = 0$ and covariance matrix

$$\mathbf{C}(0) = \begin{pmatrix} c_0^2 & 0 \\ 0 & 1/2\tau_a \end{pmatrix}. \quad (\text{D3})$$

Because of the linearity of the system [see Eq. (D1)], the distribution remains Gaussian for all later times $t > t_0$, with the expectation values and covariances evolving according to

$$\boldsymbol{\mu}(t) = [t - (1 - e^{-tA})\mathbf{A}^{-1}]\mathbf{u}, \quad (\text{D4a})$$

$$\mathbf{C}(t) = \mathbf{C}(\infty) + e^{-tA}[\mathbf{C}(0) - \mathbf{C}(\infty)]e^{-tA^T}. \quad (\text{D4b})$$

In order to calculate averages of, e.g., $\Delta\Sigma$ or $\Delta\mathcal{I}_\pm$, we have to evaluate correlators of particle positions and/or active fluctuations at two different time points (see next section). Using the propagator (D2) and the time-dependent probability density (D4), we find

$$\begin{aligned} & \langle [\mathbf{q}(t) - \boldsymbol{\mu}(t)]_i [\mathbf{q}(t') - \boldsymbol{\mu}(t')]_j \rangle \\ &= \begin{cases} [e^{-(t-t')A}\mathbf{C}(t')]_{ij} & \text{for } t \geq t', \\ [e^{-(t'-t)A}\mathbf{C}(t)]_{ji} & \text{for } t < t'. \end{cases} \end{aligned} \quad (\text{D5})$$

2. Evaluation of averages

We are interested in the averages $\langle \Delta S_{\text{sys}} \rangle$, $\langle \Delta\Sigma \rangle$, and $\langle \Delta\mathcal{I}_\pm \rangle$, where the general, trajectorywise expressions of all these quantities are given in Eqs. (22), (42b), and (61), (67), respectively. While calculating $\langle \Delta S_{\text{sys}} \rangle$ is relatively straightforward (and basically amounts to computing the variances of the Gaussians at initial and final time), the evaluation of $\langle \Delta\Sigma \rangle$ and $\langle \Delta\mathcal{I}_\pm \rangle$ is more complicated. It involves the nonlocal memory kernel $\Gamma_\tau(t, t')$ and the averages $\langle \dot{x}(t)k[x(t') - ut'] \rangle$ [for $\langle \Delta\Sigma \rangle$ and $\langle \Delta\mathcal{I}_\pm \rangle$, see Eqs. (42b) and (61), (67)], and $\langle \dot{x}(t)\eta(t') \rangle$ or $\langle \eta(t)k[x(t') - ut'] \rangle$ [for $\langle \Delta\mathcal{I}_\pm \rangle$, see Eqs. (61) and (67)]. Here, we use $f_{t'} = f(x(t'), t') = k[x(t') - ut']$; see Eq. (74). It turns out to be convenient to move the time derivative from $\dot{x}(t)$ over to $\Gamma_\tau(t, t')$ by partial integration. Then all the correlations reduce to $\langle x(t)x(t') \rangle$ and $\langle x(t)\eta(t') \rangle$, which we already calculated in Eq. (D5).

We exemplify the procedure in more detail for $\langle \Delta\Sigma \rangle$, $\langle \Delta\mathcal{I}_\pm \rangle$ can be evaluated in an analogous way. For the explicit calculation, it is useful to split Eq. (42b) into a white-noise and a colored-noise contribution,

$$\Delta\Sigma_w = \frac{1}{T} \int_0^\tau dx_t f_t, \quad (\text{D6a})$$

$$\Delta\Sigma_c = -\frac{1}{T} \left(\frac{D_a}{D} \right) \int_0^\tau dt \int_0^\tau dt' \dot{x}_t f_{t'} \Gamma_\tau(t, t'), \quad (\text{D6b})$$

such that $\Delta\Sigma = \Delta\Sigma_w + \Delta\Sigma_c$. The white-noise part $\Delta\Sigma_w$ is an ordinary stochastic integral, and its average can be evaluated using standard techniques; we recall that we use

the Stratonovich convention throughout this work. For evaluating $\Delta\Sigma_c$, we observe that $\Gamma_\tau(t, t')$ is continuously differentiable except at $t = t'$ [see Eq. (B5c)]. Splitting the t integral into the intervals $[0, t']$ and $[t', \tau]$, we can transfer the time derivative from the trajectory $\dot{x}(t)$ to the memory kernel by partial integration, so that

$$\begin{aligned} \Delta\Sigma_c &= \frac{1}{T} \left(\frac{D_a}{D} \right) \int_0^\tau dt' \left[x_0 f_{t'} \Gamma_{<}(0, t') - x_\tau f_{t'} \Gamma_{>}(\tau, t') \right. \\ &\quad \left. + \int_0^{t'} dt x_t f_{t'} \partial_t \Gamma_{<}(t, t') + \int_{t'}^\tau dt x_t f_{t'} \partial_t \Gamma_{>}(t, t') \right], \end{aligned} \quad (\text{D7})$$

where $\Gamma_{<}(t, t')$ and $\Gamma_{>}(t, t')$ are defined in analogy to Eq. (B4). Now, the average $\langle \Delta\Sigma_c \rangle$ involves correlations $\langle x_t f_{t'} \rangle = k \langle x_t [x_{t'} - ut'] \rangle$ between different time points. Using the autocorrelation functions (D5) and the result (40) for $\Gamma_\tau(t, t')$, it can thus be evaluated as an ordinary integral.

In its full general form, the resulting expression for $\langle \Delta\Sigma \rangle = \langle \Delta\Sigma_w \rangle + \langle \Delta\Sigma_c \rangle$ is rather lengthy so that we omit it here. A few relevant limiting cases are given in Sec. VI.

-
- [1] S. Ramaswamy, *The Mechanics and Statistics of Active Matter*, *Annu. Rev. Condens. Matter Phys.* **1**, 323 (2010).
 - [2] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, *Active Brownian Particles*, *Eur. Phys. J. Spec. Top.* **202**, 1 (2012).
 - [3] M. E. Cates, *Diffusive Transport without Detailed Balance in Motile Bacteria: Does Microbiology Need Statistical Physics?*, *Rep. Prog. Phys.* **75**, 042601 (2012).
 - [4] M. C. Marchetti, J.-F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha, *Hydrodynamics of Soft Active Matter*, *Rev. Mod. Phys.* **85**, 1143 (2013).
 - [5] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichardt, G. Volpe, and G. Volpe, *Active Particles in Complex and Crowded Environments*, *Rev. Mod. Phys.* **88**, 045006 (2016).
 - [6] S. Ramaswamy, *Active Matter*, *J. Stat. Mech.* (2017) 054002.
 - [7] É. Fodor and M. C. Marchetti, *The Statistical Physics of Active Matter: From Self-Catalytic Colloids to Living Cells*, *Physica (Amsterdam)* **504A**, 106 (2018).
 - [8] J. K. Parrish and W. M. Hamner, *Animal Groups in Three Dimensions: How Species Aggregate* (Cambridge University Press, Cambridge, England, 1997).
 - [9] A. Cavagna and I. Giardina, *Bird Flocks as Condensed Matter*, *Annu. Rev. Condens. Matter Phys.* **5**, 183 (2014).
 - [10] A. Cavagna, I. Giardina, and T. S. Grigera, *The Physics of Flocking: Correlation as a Compass from Experiments to Theory*, *Phys. Rep.* **728**, 1 (2018).
 - [11] J. Elgeti, R. G. Winkler, and G. Gompper, *Physics of Microswimmers—Single Particle Motion and Collective Behavior: A Review*, *Rep. Prog. Phys.* **78**, 056601 (2015).

- [12] A. E. Patteson, A. Gopinath, and P. E. Arratia, *Active Colloids in Complex Fluids*, *Curr. Opin. Colloid Interface Sci.* **21**, 86 (2016).
- [13] E. Ben-Jacob, I. Cohen, and H. Levine, *Cooperative Self-Organization of Microorganisms*, *Adv. Phys.* **49**, 395 (2000).
- [14] F. Peruani, A. Deutsch, and M. Bär, *Nonequilibrium Clustering of Self-Propelled Rods*, *Phys. Rev. E* **74**, 030904(R) (2006).
- [15] Y. Fily and M. C. Marchetti, *Athermal Phase Separation of Self-Propelled Particles with No Alignment*, *Phys. Rev. Lett.* **108**, 235702 (2012).
- [16] C. Maggi, U. M. B. Marconi, N. Gnan, and R. Di Leonardo, *Multidimensional Stationary Probability Distribution for Interacting Active Particles*, *Sci. Rep.* **5**, 10742 (2015).
- [17] U. M. B. Marconi and C. Maggi, *Towards a Statistical Mechanical Theory of Active Fluids*, *Soft Matter* **11**, 8768 (2015).
- [18] M. Paoluzzi, C. Maggi, U. M. B. Marconi, and N. Gnan, *Critical Phenomena in Active Matter*, *Phys. Rev. E* **94**, 052602 (2016).
- [19] C. Dombrowski, L. Cisneros, S. Chatkaew, R. E. Goldstein, and J. O. Kessler, *Self-Concentration and Large-Scale Coherence in Bacterial Dynamics*, *Phys. Rev. Lett.* **93**, 098103 (2004).
- [20] M. E. Cates and J. Tailleur, *Motility-Induced Phase Separation*, *Annu. Rev. Condens. Matter Phys.* **6**, 219 (2015).
- [21] T. F. F. Farage, P. Krinninger, and J. M. Brader, *Effective Interactions in Active Brownian Suspensions*, *Phys. Rev. E* **91**, 042310 (2015).
- [22] C. Maggi, M. Paoluzzi, N. Pellicciotta, A. Lepore, L. Angelani, and R. Di Leonardo, *Generalized Energy Equipartition in Harmonic Oscillators Driven by Active Baths*, *Phys. Rev. Lett.* **113**, 238303 (2014).
- [23] A. Argun, A.-R. Moradi, E. Pinçe, G. B. Bagci, A. Imparato, and G. Volpe, *Non-Boltzmann Stationary Distributions and Nonequilibrium Relations in Active Baths*, *Phys. Rev. E* **94**, 062150 (2016).
- [24] C. Maggi, M. Paoluzzi, L. Angelani, and R. Di Leonardo, *Memory-Less Response and Violation of the Fluctuation-Dissipation Theorem in Colloids Suspended in an Active Bath*, *Sci. Rep.* **7**, 17588 (2017).
- [25] J. Tailleur and M. E. Cates, *Statistical Mechanics of Interacting Run-and-Tumble Bacteria*, *Phys. Rev. Lett.* **100**, 218103 (2008).
- [26] T. Speck, J. Bialké, A. M. Menzel, and H. Löwen, *Effective Cahn-Hilliard Equation for the Phase Separation of Active Brownian Particles*, *Phys. Rev. Lett.* **112**, 218304 (2014).
- [27] S. C. Takatori, W. Yan, and J. F. Brady, *Swim Pressure: Stress Generation in Active Matter*, *Phys. Rev. Lett.* **113**, 028103 (2014).
- [28] É. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, *How Far from Equilibrium is Active Matter?*, *Phys. Rev. Lett.* **117**, 038103 (2016).
- [29] C. Nardini, É. Fodor, E. Tjhung, F. van Wijland, J. Tailleur, and M. E. Cates, *Entropy Production in Field Theories without Time-Reversal Symmetry: Quantifying the Non-Equilibrium Character of Active Matter*, *Phys. Rev. X* **7**, 021007 (2017).
- [30] K. Sekimoto, *Langevin Equation and Thermodynamics*, *Prog. Theor. Phys. Suppl.* **130**, 17 (1998).
- [31] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics Vol. 799 (Springer, New York, 2010).
- [32] U. Seifert, *Stochastic Thermodynamics: Principles and Perspectives*, *Eur. Phys. J. B* **64**, 423 (2008).
- [33] C. Jarzynski, *Equalities and Inequalities: Irreversibility and the Second Law of Thermodynamics at the Nanoscale*, *Annu. Rev. Condens. Matter Phys.* **2**, 329 (2011).
- [34] U. Seifert, *Stochastic Thermodynamics, Fluctuation Theorems and Molecular Machines*, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [35] C. Van den Broeck and M. Esposito, *Ensemble and Trajectory Thermodynamics: A Brief Introduction*, *Physica (Amsterdam)* **418A**, 6 (2015).
- [36] U. Seifert, *Stochastic Thermodynamics: From Principles to the Cost of Precision*, *Physica (Amsterdam)* **504A**, 176 (2018).
- [37] G. E. Crooks, *Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences*, *Phys. Rev. E* **60**, 2721 (1999).
- [38] U. Seifert, *Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem*, *Phys. Rev. Lett.* **95**, 040602 (2005).
- [39] R. Chetrite and K. Gawędzki, *Fluctuation Relations for Diffusion Processes*, *Commun. Math. Phys.* **282**, 469 (2008).
- [40] C. Jarzynski, *Nonequilibrium Equality for Free Energy Differences*, *Phys. Rev. Lett.* **78**, 2690 (1997).
- [41] P. Pietzonka and U. Seifert, *Entropy Production of Active Particles and for Particles in Active Baths*, *J. Phys. A* **51**, 01LT01 (2018).
- [42] A. Puglisi and U. M. B. Marconi, *Clausius Relation for Active Particles: What Can We Learn from Fluctuations*, *Entropy* **19**, 356 (2017).
- [43] We could therefore say that this active bath is an “athermal bath.” However, the term athermal bath has been established for noisy environments with non-Gaussian properties [44–46], such that we do not use it here.
- [44] K. Kanazawa, T. Sagawa, and H. Hayakawa, *Stochastic Energetics for Non-Gaussian Processes*, *Phys. Rev. Lett.* **108**, 210601 (2012).
- [45] K. Kanazawa, T. G. Sano, T. Sagawa, and H. Hayakawa, *Minimal Model of Stochastic Athermal Systems: Origin of Non-Gaussian Noise*, *Phys. Rev. Lett.* **114**, 090601 (2015).
- [46] K. Kanazawa, *Statistical Mechanics for Athermal Fluctuation: Non-Gaussian Noise in Physics* (Springer, New York, 2017).
- [47] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, 2nd ed. (Springer, New York, 1985).
- [48] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, New York, 1992), Vol. 1.
- [49] S. Chaki and R. Chakrabarti, *Entropy Production and Work Fluctuation Relations for a Single Particle in Active Bath*, *Physica (Amsterdam)* **511A**, 302 (2018).
- [50] U. M. B. Marconi, A. Puglisi, and C. Maggi, *Heat, Temperature and Clausius Inequality in a Model for Active Brownian Particles*, *Sci. Rep.* **7**, 46496 (2017).

- [51] D. Mandal, K. Klymko, and M. R. DeWeese, *Entropy Production and Fluctuation Theorems for Active Matter*, *Phys. Rev. Lett.* **119**, 258001 (2017).
- [52] N. Koumakis, C. Maggi, and R. Di Leonardo, *Directed Transport of Active Particles over Asymmetric Energy Barriers*, *Soft Matter* **10**, 5695 (2014).
- [53] G. Szamel, *Self-Propelled Particle in an External Potential: Existence of an Effective Temperature*, *Phys. Rev. E* **90**, 012111 (2014).
- [54] G. Szamel, E. Flenner, and L. Berthier, *Glassy Dynamics of Athermal Self-Propelled Particles: Computer Simulations and a Nonequilibrium Microscopic Theory*, *Phys. Rev. E* **91**, 062304 (2015).
- [55] E. Flenner, G. Szamel, and L. Berthier, *The Nonequilibrium Glassy Dynamics of Self-Propelled Particles*, *Soft Matter* **12**, 7136 (2016).
- [56] U. M. B. Marconi, N. Gnan, M. Paoluzzi, C. Maggi, and R. Di Leonardo, *Velocity Distribution in Active Particles Systems*, *Sci. Rep.* **6**, 23297 (2016).
- [57] G. Szamel, *Evaluating Linear Response in Active Systems with No Perturbing Field*, *Europhys. Lett.* **117**, 50010 (2017).
- [58] C. Sandford, A. Y. Grosberg, and J.-F. Joanny, *Pressure and Flow of Exponentially Self-Correlated Active Particles*, *Phys. Rev. E* **96**, 052605 (2017).
- [59] L. Caprini, U. M. B. Marconi, and A. Vulpiani, *Linear Response and Correlation of a Self-Propelled Particle in the Presence of External Fields*, *J. Stat. Mech.* (2018) 033203.
- [60] L. Berthier and J. Kurchan, *Non-Equilibrium Glass Transitions in Driven and Active Matter*, *Nat. Phys.* **9**, 310 (2013).
- [61] S. Shankar and M. C. Marchetti, *Hidden Entropy Production and Work Fluctuations in an Ideal Active Gas*, *Phys. Rev. E* **98**, 020604(R) (2018).
- [62] R. Kubo, *The Fluctuation-Dissipation Theorem*, *Rep. Prog. Phys.* **29**, 255 (1966).
- [63] U. M. B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani, *Fluctuation-Dissipation: Response Theory in Statistical Physics*, *Phys. Rep.* **461**, 111 (2008).
- [64] E. M. Purcell, *Life at Low Reynolds Number*, *Am. J. Phys.* **45**, 3 (1977).
- [65] I. Snook, *The Langevin and Generalised Langevin Approach to the Dynamics of Atomic, Polymeric and Colloidal Systems* (Elsevier, New York, 2006).
- [66] R. M. Mazo, *Brownian Motion: Fluctuations, Dynamics, and Applications* (Oxford University Press, New York, 2002), Vol. 112.
- [67] A. Einstein, *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, *Ann. Phys. (Berlin)* **322**, 549 (1905).
- [68] H. Nyquist, *Thermal Agitation of Electric Charge in Conductors*, *Phys. Rev.* **32**, 110 (1928).
- [69] H. B. Callen and T. A. Welton, *Irreversibility and Generalized Noise*, *Phys. Rev.* **83**, 34 (1951).
- [70] D. T. N. Chen, A. W. C. Lau, L. A. Hough, M. F. Islam, M. Goulian, T. C. Lubensky, and A. G. Yodh, *Fluctuations and Rheology in Active Bacterial Suspensions*, *Phys. Rev. Lett.* **99**, 148302 (2007).
- [71] P. Gaspard and R. Kapral, *Communication: Mechanochemical Fluctuation Theorem and Thermodynamics of Self-Phoretic Motors*, *J. Chem. Phys.* **147**, 211101 (2017).
- [72] T. Speck, *Active Brownian Particles Driven by Constant Affinity*, *Europhys. Lett.* **123**, 20007 (2018).
- [73] F. Zamponi, F. Bonetto, L. F. Cugliandolo, and J. Kurchan, *A Fluctuation Theorem for Non-Equilibrium Relaxational Systems Driven by External Forces*, *J. Stat. Mech.* (2005) P09013.
- [74] T. Ohkuma and T. Ohta, *Fluctuation Theorems for Non-Linear Generalized Langevin Systems*, *J. Stat. Mech.* (2007) P10010.
- [75] U. M. B. Marconi, N. Gnan, M. Paoluzzi, C. Maggi, and R. Di Leonardo, *Velocity Distribution in Active Particles Systems*, *Sci. Rep.* **6**, 23297 (2016).
- [76] P. Hänggi and P. Jung, *Colored Noise in Dynamical Systems*, *Advances in Chemical Physics Vol. 89*, edited by I. Prigogine and S. A. Rice (Wiley & Sons, Inc., Hoboken, NJ, 1994), Chap. 4, pp. 239–326.
- [77] Note that we do not introduce a special symbol (like \circ , for instance) to denote Stratonovich products, but rather we use the usual dot for scalar products because we adopt the convention throughout this manuscript that all products involving stochastic processes (or functions with similar properties) are to be interpreted in the Stratonovich sense.
- [78] J. R. Gomez-Solano, L. Bellon, A. Petrosyan, and S. Ciliberto, *Steady-State Fluctuation Relations for Systems Driven by an External Random Force*, *Europhys. Lett.* **89**, 60003 (2010).
- [79] T. Schmiedl and U. Seifert, *Optimal Finite-Time Processes in Stochastic Thermodynamics*, *Phys. Rev. Lett.* **98**, 108301 (2007).
- [80] S. Ciliberto, *Experiments in Stochastic Thermodynamics: Short History and Perspectives*, *Phys. Rev. X* **7**, 021051 (2017).
- [81] D. McGloin, C. McDonald, and Y. Belotti, *Colloidal Interactions with Optical Fields: Optical Tweezers, in Fluids, Colloids and Soft Materials: An Introduction to Soft Matter Physics*, edited by A. Fernandez Nieves and A. M. Puertas (John Wiley & Sons, Inc., New York, 2016), pp. 111–130.
- [82] T. Speck, J. Mehl, and U. Seifert, *Role of External Flow and Frame Invariance in Stochastic Thermodynamics*, *Phys. Rev. Lett.* **100**, 178302 (2008).
- [83] D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud, and A. Petrosyan, *Entropy Production and Time Asymmetry in Nonequilibrium Fluctuations*, *Phys. Rev. Lett.* **98**, 150601 (2007).
- [84] L. Onsager and S. Machlup, *Fluctuations and Irreversible Processes*, *Phys. Rev.* **91**, 1505 (1953).
- [85] S. Machlup and L. Onsager, *Fluctuations and Irreversible Process. II. Systems with Kinetic Energy*, *Phys. Rev.* **91**, 1512 (1953).
- [86] V. Y. Chernyak, M. Chertkov, and C. Jarzynski, *Path-Integral Analysis of Fluctuation Theorems for General Langevin Processes*, *J. Stat. Mech.* (2006) P08001.
- [87] L. F. Cugliandolo and V. Lecomte, *Rules of Calculus in the Path Integral Representation of White Noise Langevin Equations: The Onsager-Machlup Approach*, *J. Phys. A* **50**, 345001 (2017).

- [88] Note that \mathbf{v} might contain forces whose sign should be inverted under time reversal for physical reasons [39]. We do not consider this case here, but the derivation can be adapted straightforwardly.
- [89] F. C. MacKintosh and C. F. Schmidt, *Active Cellular Materials*, *Curr. Opin. Cell Biol.* **22**, 29 (2010).
- [90] 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, *Probing the Stochastic, Motor-Driven Properties of the Cytoplasm Using Force Spectrum Microscopy*, *Cell* **158**, 822 (2014).
- [91] W. W. Ahmed, É. Fodor, and T. Betz, *Active Cell Mechanics: Measurement and Theory*, *Biochim. Biophys. Acta* **1853**, 3083 (2015).
- [92] D. Needleman and Z. Dogic, *Active Matter at the Interface between Materials Science and Cell Biology*, *Nat. Rev. Mater.* **2**, 17048 (2017).
- [93] A. Bernheim-Groswasser, N. S. Gov, S. A. Safran, and S. Tzilil, *Living Matter: Mesoscopic Active Materials*, *Adv. Mater.* **30**, 1707028 (2018).
- [94] E. Ben-Isaac, Y. K. Park, G. Popescu, F. L. H. Brown, N. S. Gov, and Y. Shokef, *Effective Temperature of Red-Blood-Cell Membrane Fluctuations*, *Phys. Rev. Lett.* **106**, 238103 (2011).
- [95] É. Fodor, M. Guo, N. S. Gov, P. Visco, D. A. Weitz, and F. van Wijland, *Activity-Driven Fluctuations in Living Cells*, *Europhys. Lett.* **110**, 48005 (2015).
- [96] S. Krishnamurthy, S. Ghosh, D. Chatterji, R. Ganapathy, and A. K. Sood, *A Micrometre-Sized Heat Engine Operating between Bacterial Reservoirs*, *Nat. Phys.* **12**, 1134 (2016).
- [97] T. Speck, *Stochastic Thermodynamics for Active Matter*, *Europhys. Lett.* **114**, 30006 (2016).
- [98] D. Martin, C. Nardini, M. E. Cates, and É. Fodor, *Extracting Maximum Power from Active Colloidal Heat Engines*, *Europhys. Lett.* **121**, 60005 (2018).
- [99] H. Risken, *The Fokker-Planck Equation* (Springer, New York, 1984).
- [100] L. Dabelow, S. Bo, and R. Eichhorn (to be published).
- [101] F. W. Wiegand, *Introduction to Path-Integral Methods in Physics and Polymer Science* (World Scientific, Singapore, 1986).
- [102] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, 3rd ed. (Oxford University Press, New York, 1996).
- [103] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, 1st ed. (Springer, New York, 1999).
- [104] I. Stakgold and M. Hols, *Green's Functions and Boundary Value Problems*, 3rd ed. (Wiley, New York, 2011).
- [105] A. J. McKane, H. C. Luckock, and A. J. Bray, *Path integrals and Non-Markov Processes. I. General Formalism*, *Phys. Rev. A* **41**, 644 (1990).
- [106] P. Hänggi, *Path Integral Solutions for Non-Markovian Processes*, *Z. Phys. B* **75**, 275 (1989).
- [107] A. Puglisi and D. Villamaina, *Irreversible Effects of Memory*, *Europhys. Lett.* **88**, 30004 (2009).
- [108] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, *Thermodynamics of Information*, *Nat. Phys.* **11**, 131 (2015).
- [109] Our choice of active self-propulsion being odd under time reversal is different from the approach in Ref. [61], where self-propulsion is considered even under time reversal. As we argue in the main text (mainly at the beginning of Sec. V B), we think that the time-odd choice corresponding to treating $\boldsymbol{\eta}$ as a velocitylike quantity is more appropriate for self-propulsion. In any case, if one were to treat self-propulsion as being even under time reversal, this would coincide with the model we designated active bath in the manuscript and labeled by “+” sub- and superscripts.
- [110] T. Sagawa and M. Ueda, *Fluctuation Theorem with Information Exchange: Role of Correlations in Stochastic Thermodynamics*, *Phys. Rev. Lett.* **109**, 180602 (2012).
- [111] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, *Ann. Math. Stat.* **22**, 79 (1951).
- [112] M. Lexa, Useful Facts about the Kullback-Leibler Discrimination Distance, 2004.
- [113] S. Bo and A. Celani, *Multiple-Scale Stochastic Processes: Decimation, Averaging and Beyond*, *Phys. Rep.* **670**, 1 (2017).
- [114] É. Roldán and J. M. R. Parrondo, *Estimating Dissipation from Single Stationary Trajectories*, *Phys. Rev. Lett.* **105**, 150607 (2010).
- [115] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley, New York, 2006).
- [116] É. Roldán, I. Neri, M. Dörpinghaus, H. Meyer, and F. Jülicher, *Decision Making in the Arrow of Time*, *Phys. Rev. Lett.* **115**, 250602 (2015).
- [117] P. Mestres, I. A. Martinez, A. Ortiz-Ambriz, R. A. Rica, and E. Roldan, *Realization of Nonequilibrium Thermodynamic Processes Using External Colored Noise*, *Phys. Rev. E* **90**, 032116 (2014).
- [118] R. van Zon and E. G. D. Cohen, *Stationary and Transient Work-Fluctuation Theorems for a Dragged Brownian Particle*, *Phys. Rev. E* **67**, 046102 (2003).
- [119] R. van Zon and E. G. D. Cohen, *Extended Heat-Fluctuation Theorems for a System with Deterministic and Stochastic Forces*, *Phys. Rev. E* **69**, 056121 (2004).
- [120] A. Engel, *Asymptotics of Work Distributions in Non-equilibrium Systems*, *Phys. Rev. E* **80**, 021120 (2009).
- [121] N. G. van Kampen, *The Expansion of the Master Equation*, *Adv. Chem. Phys.* **34**, 245 (1976).
- [122] L. Caprini, U. M. B. Marconi, A. Puglisi, and A. Vulpiani, *Comment on Entropy Production and Fluctuation Theorems for Active Matter*, *Phys. Rev. Lett.* **121**, 139801 (2018).
- [123] D. Mandal, K. Klymko, and M. R. DeWeese, *Mandal, Klymko, and DeWeese Reply*, *Phys. Rev. Lett.* **121**, 139802 (2018).
- [124] Alternatively to including stochastic forces representing the active fluctuations into the equations of motion, a quite different approach consists of adding velocity-dependent nonlinear active forces to model self-propulsion. These forces are typically designed in a way such that energy is fed into the system by “negative damping” (linear in the velocity) and is dissipated at higher-order contributions (such as cubic velocity terms). An analysis of the stochastic thermodynamics of this kind of model is given in, e.g., Refs. [125,126].

- [125] C. Ganguly and D. Chaudhuri, *Stochastic Thermodynamics of Active Brownian Particles*, *Phys. Rev. E* **88**, 032102 (2013).
- [126] D. Chaudhuri, *Active Brownian Particles: Entropy Production and Fluctuation Response*, *Phys. Rev. E* **90**, 022131 (2014).
- [127] J. R. Howse, R. A. L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian, *Self-Motile Colloidal Particles: From Directed Propulsion to Random Walk*, *Phys. Rev. Lett.* **99**, 048102 (2007).
- [128] K. C. Leptos, J. S. Guasto, J. P. Gollub, A. I. Pesci, and R. E. Goldstein, *Dynamics of Enhanced Tracer Diffusion in Suspensions of Swimming Eukaryotic Microorganisms*, *Phys. Rev. Lett.* **103**, 198103 (2009).
- [129] P. Bohec, F. Gallet, C. Maes, S. Safaverdi, P. Visco, and F. van Wijland, *Probing Active Forces via a Fluctuation-Dissipation Relation: Application to Living Cells*, *Europhys. Lett.* **102**, 50005 (2013).
- [130] A. Jepson, V. A. Martinez, J. Schwarz-Linek, A. Morozov, and W. C. K. Poon, *Enhanced Diffusion of Nonswimmers in a Three-Dimensional Bath of Motile Bacteria*, *Phys. Rev. E* **88**, 041002(R) (2013).
- [131] I. Buttinoni, J. Bialké, F. Kümmel, H. Löwen, C. Bechinger, and T. Speck, *Dynamical Clustering and Phase Separation in Suspensions of Self-Propelled Colloidal Particles*, *Phys. Rev. Lett.* **110**, 238301 (2013).
- [132] F. S. Gnesotto, F. Mura, J. Gladrow, and C. P. Broedersz, *Broken Detailed Balance and Non-Equilibrium Dynamics in Living Systems: A Review*, *Rep. Prog. Phys.* **81**, 066601 (2018).
- [133] T. Harada and S.-i. Sasa, *Equality Connecting Energy Dissipation with a Violation of the Fluctuation-Response Relation*, *Phys. Rev. Lett.* **95**, 130602 (2005).
- [134] I. Neri, É. Roldán, and F. Jülicher, *Statistics of Infima and Stopping Times of Entropy Production and Applications to Active Molecular Processes*, *Phys. Rev. X* **7**, 011019 (2017).
- [135] É. Roldán, J. Barral, P. Martin, J. M. R. Parrondo, and F. Jülicher, *Arrow of Time in Active Fluctuations*, *arXiv*: 1803.04743.
- [136] G. Verley, M. Esposito, T. Willaert, and C. Van den Broeck, *The Unlikely Carnot Efficiency*, *Nat. Commun.* **5**, 4721 (2014).
- [137] G. Verley, T. Willaert, C. Van den Broeck, and M. Esposito, *Universal Theory of Efficiency Fluctuations*, *Phys. Rev. E* **90**, 052145 (2014).
- [138] A. Crisanti, A. Puglisi, and D. Villamaina, *Nonequilibrium and Information: The Role of Cross Correlations*, *Phys. Rev. E* **85**, 061127 (2012).
- [139] G. Verley, C. Van den Broeck, and M. Esposito, *Work Statistics in Stochastically Driven Systems*, *New J. Phys.* **16**, 095001 (2014).
- [140] S. K. Manikandan and S. Krishnamurthy, *Asymptotics of Work Distributions in a Stochastically Driven System*, *Eur. Phys. J. B* **90**, 258 (2017).
- [141] S. K. Manikandan and S. Krishnamurthy, *Exact Results for the Finite Time Thermodynamic Uncertainty Relation*, *J. Phys. A* **51**, 11LT01 (2018).
- [142] A. Pal and S. Sabhapandit, *Work Fluctuations for a Brownian Particle in a Harmonic Trap with Fluctuating Locations*, *Phys. Rev. E* **87**, 022138 (2013).