

Mutual information in changing environments: Nonlinear interactions, out-of-equilibrium systems, and continuously varying diffusivities

Giorgio Nicoletti¹ and Daniel Maria Busiello^{2,*}

¹Laboratory of Interdisciplinary Physics, Department of Physics and Astronomy “G. Galilei,” University of Padova, Padova 35121, Italy

²Institute of Physics, École Polytechnique Fédérale de Lausanne–EPFL, 1015 Lausanne, Switzerland



(Received 4 April 2022; revised 1 June 2022; accepted 14 July 2022; published 29 July 2022)

Biochemistry, ecology, and neuroscience are examples of prominent fields aiming at describing interacting systems that exhibit nontrivial couplings to complex, ever-changing environments. We have recently shown that linear interactions and a switching environment are encoded separately in the mutual information of the overall system. Here we first generalize these findings to a broad class of nonlinear interacting models. We find that a new term in the mutual information appears, quantifying the interplay between nonlinear interactions and environmental changes, and leading to either constructive or destructive information interference. Furthermore, we show that a higher mutual information emerges in out-of-equilibrium environments with respect to an equilibrium scenario. Finally, we generalize our framework to the case of continuously varying environments. We find that environmental changes can be mapped exactly into an effective spatially varying diffusion coefficient, shedding light on modeling of biophysical systems in inhomogeneous media.

DOI: [10.1103/PhysRevE.106.014153](https://doi.org/10.1103/PhysRevE.106.014153)

I. INTRODUCTION

An accurate description of real-world systems should capture both their internal interactions and their couplings with noisy, ever-changing environments. The main difficulty stems from the fact that often environmental changes are not directly observable, hence leading to the necessity of more simplified yet informative approaches. The more simplistic one might be to ignore environmental effects. However, it is now well understood that these are fundamental ingredients in many different fields, from biology to neuroscience [1–8]. A slightly more complete understanding of real-world systems would come from the estimation of effective couplings, in principle affected also by the presence of a changing environment [9–11]. Although this idea might lead to descriptive models, it makes it impossible to understand whether the observed behaviors originate from internal interactions or are sheer consequences of a shared environment.

Examples of interacting systems affected by an ever-changing environment pervades nature. Species in ecological networks interact among them, being constantly affected by sudden changes in their surrounding ecosystems (e.g., climatic, artificial) [2]. Similarly, neurons form an intricate network of connections and are affected by external stimuli, either externally controlled or unknown and not observable [8]. Other examples are proteins in the cytoplasm, subjects to ceaseless chemical reactions [12], tracers in active crowded baths [13,14], and particles diffusing in inhomogeneous media [15]. In all these scenarios, the environment can be considered independent from the internal degrees of freedom and acting on a different timescale.

In this intricate scenario, information theory might be the leading framework to determine the role of different coupling sources in shaping complex systems’ behaviors. In particular, a key quantity is the mutual information associated with two stationary processes $x_1(t)$ and $x_2(t)$,

$$I = \int dx_1 dx_2 p(x_1, x_2) \log \frac{p(x_1, x_2)}{p(x_1)p(x_2)}, \quad (1)$$

which is nothing but the Kullback-Leibler divergence between $p(x_1, x_2)$, the joint stationary probability distribution, and $p(x_1)p(x_2)$, the product of their marginalized stationary distributions [16]. I quantifies the overall dependency between x_1 and x_2 . Recently, we showed that the mutual information of systems with linear interactions and a switching discrete-state environment receive disentangled contributions from environmental and internal interactions [17]. This result revealed that the properties of the information content of complex systems can be particularly informative, and that tools from information theory may greatly help to capture their essential features [18].

Here we generalize our previous results to more complex systems, highlighting criticalities and potentialities of the proposed approach. First, we show that the presence of nonlinear interactions may give rise to a new interference term in the mutual information. This additional contribution may lead to both an increase and a decrease of the mutual information with respect to the sum of the contributions associated with the environment and the internal interactions. A careful analysis of the system under investigation here reveals the phenomenological origin of this constructive or destructive information interference.

Then we show that in systems placed out-of-equilibrium by the presence of a multiplicative noise (e.g., a thermal gradient), in the absence of interactions, the environmental information increases with the magnitude of the nonequilibrium

*Present address: Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany (busiello@pks.mpg.de).

term. Finally, we consider the case in which the environment is described by a continuous process [15,19]. We show that the effect of the environment at stationarity can be mapped into a heterogeneous diffusion coefficient, i.e., an effective inhomogeneous medium. Thus, the presence of changing environments cannot always be mapped into effective interactions among degrees of freedom, as one may naively believe, but sometimes it manifests into indirect spatial couplings. This result highlights a potential warning for inference methods trying to estimate interactions from measured data.

II. TIMESCALE SEPARATION APPROACH

Consider a system of N possibly interacting particles that share the same changing environment, whose effect is to modify the overall diffusion coefficient. In general, we assume that we have a finite number M of environmental states, i.e., the diffusion coefficient of the system only takes discrete values. We will eventually relax this condition.

This framework is analogous to the one introduced in Ref. [17]. It is described by the following Fokker-Planck equation,

$$\begin{aligned} \partial_t p_i(\mathbf{x}, t) = & \sum_{\mu=1}^N \partial_{\mu} [F_{\mu}(\mathbf{x}) p_i(\mathbf{x}, t)] + \sum_{\mu=1}^N \partial_{\mu}^2 [D_{\mu} p_i(\mathbf{x}, t)] \\ & + \sum_{j=1}^M [W_{j \rightarrow i} p_j(\mathbf{x}, t) - W_{i \rightarrow j} p_i(\mathbf{x}, t)], \end{aligned} \quad (2)$$

where $\mathbf{x} = (x_1, \dots, x_N)$ indicates all internal degrees of freedom, $F_{\mu}(\mathbf{x})$ is the μ th component of a generic force field, $W_{i \rightarrow j}$ is the transition rate from the i th to the j th environmental state, and D_i is the diffusion coefficient associated with such states. We are interested in the stationary solution of Eq. (2), whose finding is, in general, a particularly challenging task. Therefore, we resort to a timescale separation approach in which the environment can be either much faster or much slower than all timescales at which the internal dynamics operates.

Let us assume that τ is the fastest timescale associated with the force field \mathbf{F} , whereas the jump process between the environmental states occurs on a typical timescale τ_{env} . For instance, if $M = 2$ we would have $\tau_{\text{env}} = (W_{1 \rightarrow 2} + W_{2 \rightarrow 1})^{-1}$. Then, we first consider the limit $\tau/\tau_{\text{env}} := \delta \ll 1$ and seek a formal solution of the form

$$p_i(\mathbf{x}, t) = p_i^{(0)}(\mathbf{x}, t) + \delta p_i^{(1)}(\mathbf{x}, t) + \mathcal{O}(\delta^2). \quad (3)$$

Rescaling the time by the slowest timescale, that is $t \rightarrow t/\tau_{\text{env}}$, we end up with

$$\begin{aligned} \partial_t p_i^{(0)} = & \frac{1}{\delta} \sum_{\mu=1}^N \{ \partial_{\mu} [\tilde{F}_{\mu}(\mathbf{x}) p_i^{(0)}] + \partial_{\mu}^2 [\tilde{D}_i p_i^{(0)}] \} \\ & + \sum_{\mu=1}^N \{ \partial_{\mu} [\tilde{F}_{\mu}(\mathbf{x}) p_i^{(1)}] + \partial_{\mu}^2 [\tilde{D}_i p_i^{(1)}] \} \\ & + \sum_{j=1}^M [\tilde{W}_{j \rightarrow i} p_j^{(0)} - \tilde{W}_{i \rightarrow j} p_i^{(0)}] + \mathcal{O}(\delta), \end{aligned}$$

where $\tilde{F}_{\mu} := \tau F_{\mu}$, $\tilde{D}_i := \tau D_i$, and $\tilde{W}_{i \rightarrow j} := \tau_{\text{env}} W_{i \rightarrow j}$. The leading δ^{-1} order corresponds to the stationary solution of the Fokker-Planck equation associated with the fastest dynamics alone. Here, this is equal to the distribution $P_i^{\text{st}}(\mathbf{x})$ that solves the interacting dynamics at a fixed environmental state D_i :

$$0 = \sum_{\mu=1}^N \{ \partial_{\mu} [F_{\mu}(\mathbf{x}) P_i^{\text{st}}(\mathbf{x})] + \partial_{\mu}^2 [D_i P_i^{\text{st}}(\mathbf{x})] \}. \quad (4)$$

We can always assume that the zeroth-order solution of Eq. (2) can be written as $p_i^{(0)}(\mathbf{x}, t) = \pi_i(t) P_i^{\text{st}}(\mathbf{x})$ (see also Ref. [20]). Then, by integrating over \mathbf{x} , the $\mathcal{O}(1)$ order gives

$$\partial_t \pi_i(t) = \sum_{j=1}^M [\tilde{W}_{j \rightarrow i} \pi_j(t) - \tilde{W}_{i \rightarrow j} \pi_i(t)].$$

Hence, the zeroth order for the steady state reads

$$p_{\text{slow}}(\mathbf{x}) := \sum_{i=1}^M p_i^{(0)}(\mathbf{x})|_{\delta \ll 1} = \sum_{i=1}^M [\pi_i P_i^{\text{st}}(\mathbf{x})], \quad (5)$$

where π_i are the stationary probabilities of the jump process alone, and the subscript "slow" refers to the fact the environment is the slowest process in this limit. Equation (5) is a mixture distribution, where the mixture components are the stationary solutions obtained with a fixed environmental state i , $P_i^{\text{st}}(\mathbf{x})$.

These calculations can be easily carried out in the opposite limit, $\tau_{\text{env}}/\tau = \delta^{-1} \ll 1$, i.e., when the environment is much faster than the internal processes. In this case, the stationary joint probability distribution $p_{\text{fast}}(\mathbf{x})$ that solves Eq. (2) is given by the solution of

$$0 = \sum_{\mu=1}^N \left\{ \partial_{\mu} [F_{\mu}(\mathbf{x}) p_{\text{fast}}(\mathbf{x})] + \partial_{\mu}^2 \left[\left(\sum_i \pi_i D_i \right) p_{\text{fast}}(\mathbf{x}) \right] \right\}. \quad (6)$$

The system feels an effective diffusion coefficient $\sum_i \pi_i D_i$, which is the stationary ensemble average of all environmental states, as a consequence of the presence of the environment.

III. MUTUAL INFORMATION: DEPENDENCIES AND BOUNDS

For the sake of simplicity, here we focus on the case of two particles moving in a one-dimensional (1D) space, $\mathbf{x} = (x_1, x_2)$, and two environmental states, specified by the diffusion coefficients D_- and D_+ . In what follows, the environmental states will be indexed by $i \in \{-, +\}$. Accordingly, w_{\pm} is the rate of transition into the state $i = \pm$. The multi-dimensional generalization is straightforward. In Ref. [17], we showed that, in the presence of linear interactions the mutual information associated with the stationary solution of Eq. (2) can be exactly disentangled into two independent contributions. The first one depends solely on the environmental dimensionless parameters, D_-/D_+ and w_-/w_+ . The second, instead, only depends on the internal interactions between the two particles. Hence, the dependencies between x_1 and x_2 induced by the environment and by the internal interactions are fully disentangled. Notably, in this case Eq. (5) corresponds to a Gaussian mixture, which allows for analytical calculations.

We now relax the assumption of linear interactions and explore the effects of nonlinear couplings. In particular, we already know that the mutual information in the fast-jumps limit, Eq. (6), only contains the contribution from the internal interactions, since the environment results in a constant effective diffusion. Conversely, the slow-jumps limit is much more intriguing. In this case, both environment and internal couplings will contribute to the mutual information between x_1 and x_2 , Eq. (1), but their interplay is far from being easy to predict.

As a general remark, we notice that the mutual information can only depend on dimensionless quantities since it is itself dimensionless. These, in turn, may depend on environmental features, internal parameters, or combinations of both. In the slow-jumps limit, by inspecting Eq. (5), the parameter w_+/w_- can only enter through $\pi_{\pm} = w_{\pm}/(w_{\pm} + w_{\mp})$. Then, the stationary solution of the dynamics at a fixed environment, P_{\pm}^{st} , determines all the other dimensionless parameters in play.

Moreover, the slow-jumps limit allows us to consider some simple bounds [16] on the entropy of a mixture distribution, and thus on the mutual information. Let H_{12} be the joint entropy associated to the probability distribution $p_{\text{slow}}(x_1, x_2)$ defined by Eq. (5), and let H_{12}^i be the entropy of its i th component $P_i^{\text{st}}(x_1, x_2)$. This joint entropy is bounded by (see Appendix A)

$$\sum_i \pi_i H_{12}^i \leq H_{12} \leq \sum_i \pi_i [H_{12}^i - \log \pi_i]. \quad (7)$$

Analogous bounds can be cast for H_1 and H_2 , i.e., the entropies of the marginal distributions $p_{\text{slow}}(x_1)$ and $p_{\text{slow}}(x_2)$, respectively. Then, a lower (upper) bound on the mutual information, $I = H_1 + H_2 - H_{12}$, can be found by taking these lower (upper) bounds on H_1 and H_2 and the upper (lower) one on H_{12} , Eq. (7). Therefore, the mutual information of the mixture distribution, Eq. (5), is bounded by

$$\sum_i \pi_i I^i - H_{\text{jumps}} \leq I \leq \sum_i \pi_i I^i + 2H_{\text{jumps}}, \quad (8)$$

where $H_{\text{jumps}} = -\sum_i \pi_i \log \pi_i$ is the entropy associated with the environmental jumps, and $I_i = H_1^i + H_2^i - H_{12}^i$ is the mutual information associated with the i th component $P_i^{\text{st}}(x_1, x_2)$ of the joint distribution.

These bounds can be greatly improved [17,21], provided our ability to compute some suitable information distances both between the components of the mixture distribution in Eq. (5) and the components of the two corresponding marginalizations. However, besides the Gaussian case, this is often challenging. Generally speaking, Eq. (8) shows that the mutual information cannot be larger than the sum of the weighted average of the mutual information in the different environmental states and twice the entropy of the jumps. Albeit loose, this upper bound shows that, in principle, the system may contain more information than the sum of the contributions stemming from the environment and the internal interactions. Therefore, on the one hand, we expect and later show that the presence of nonlinear interactions might undermine the exact disentangling holding for the linear case. On the other hand, in what follows we also report situations in which the presence of nonlinearities boost, or even suppress, the overall mutual information due to internal and

environmental couplings. These observations effectively hinder our ability to pinpoint the presence of interactions in complex systems, but reveal surprising properties of the associated information quantities.

IV. ENVIRONMENTAL CONTRIBUTION WITH NONLINEAR RELAXATION

Let us start with a nonlinear, yet noninteracting, case. Each particle diffuses in the 1D potential

$$U(x_{\mu}) = \frac{x_{\mu}^4}{4\tau}, \quad (9)$$

depicted in Fig. 1(a). Hence, the drift term in Eq. (2) is simply given by the potential gradient $F_{\mu}(x_{\mu}) = -\partial_{\mu}U(x_{\mu}) = -x_{\mu}^3/\tau$, resulting in a nonlinear relaxation. Notice that, in this case, the only dimensionless parameters are given by w_-/w_+ and D_-/D_+ .

In the slow-jumps limit, this term leads to the following mixture components:

$$P_i^{\text{st}}(x_1, x_2) = \sqrt{\frac{D_i \tau \pi^2}{2}} \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})} \exp\left(-\frac{x_1^4 + x_2^4}{4D_i \tau}\right), \quad (10)$$

which are the solutions of Eq. (4). The corresponding marginal distributions are

$$P_i^{\text{st}}(x_{\mu}) = \frac{\Gamma(\frac{3}{4})}{\pi(D_i \tau)^{1/4}} e^{-\frac{x_{\mu}^4}{4D_i \tau}} \quad (11)$$

and, clearly, $P_i^{\text{st}}(x_1, x_2) = P_i^{\text{st}}(x_1)P_i^{\text{st}}(x_2)$ since the particles are not interacting. The joint mixture distribution, defined in Eq. (5), and its factorization,

$$\prod_{\mu=1}^2 p_{\text{slow}}(x_{\mu}) = \prod_{\mu=1}^2 \left\{ \sum_{i=\{+,-\}} [\pi_i P_i^{\text{st}}(x_{\mu})] \right\},$$

are plotted in Figs. 1(b) and 1(c). Notably, the effects of the environment on the joint distribution, with respect to the factorized probability distribution, reflect into a suppression of the tails along the axes. We are interested in the mutual information I of the joint probability distribution $p_{\text{slow}}(x_1, x_2)$. Since there are no interaction between x_1 and x_2 , the only contribution to I can come from the shared environment. Hence, for simplicity, we name this mutual information I_{env} .

As for the linear case [17], we are able to compute the pairwise distance bounds [21] on I_{env} analytically, starting from Eqs. (10) and (11) (see Appendix B). In particular, we find that the lower bound obeys

$$I_{\text{env}}^{\text{low}}\left(\frac{D_-}{D_+}, \frac{w_-}{w_+}\right) = -\pi_+ \log \frac{[\pi_+ + h_C^1(\frac{D_-}{D_+})\pi_-]^2}{\pi_+ + h_D^2(\frac{D_-}{D_+})\pi_-} - \pi_- \log \frac{[h_C^1(\frac{D_{\pm}}{D_{\pm}})\pi_+ + \pi_-]^2}{h_D^2(\frac{D_{\pm}}{D_{\pm}})\pi_+ + \pi_-} \quad (12)$$

where $h_C^1(x) = (4x)^{1/8}/(1+x)^{1/4}$ and $h_D^1(x) = (x-1-\log x)/4$ are, respectively, the 1/2-Chernoff and Kullback-Leibler divergence between the $-$ and $+$ components of the

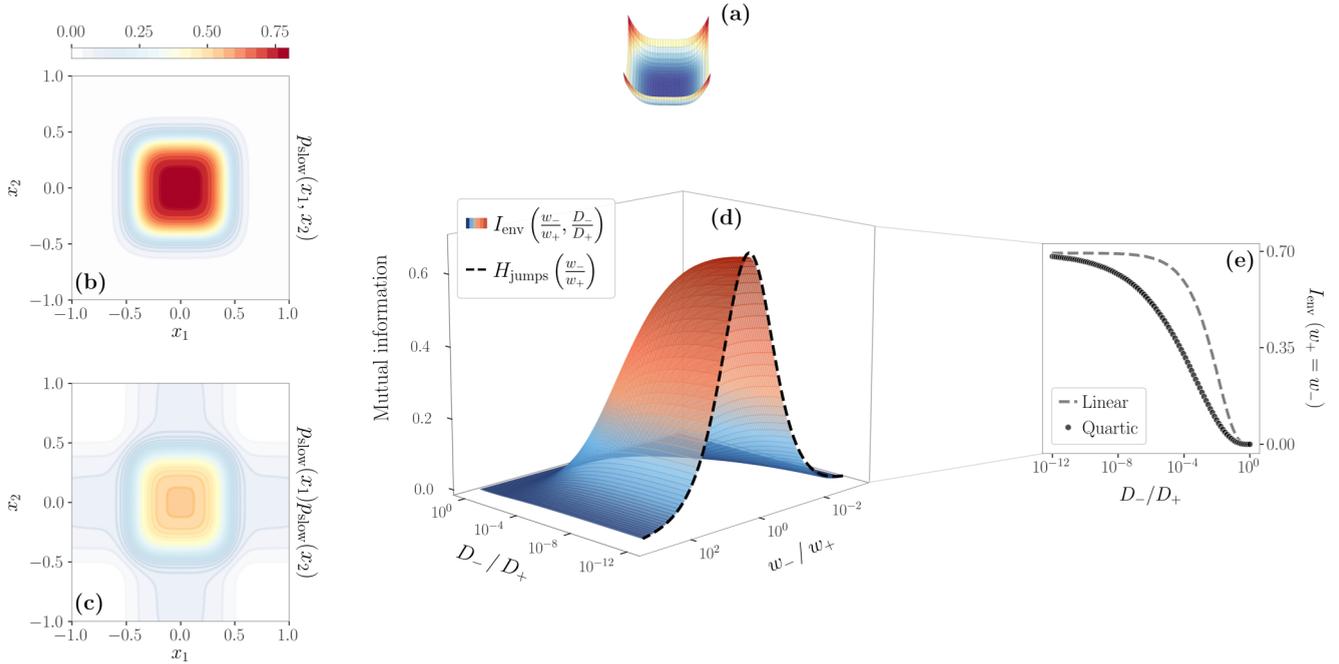


FIG. 1. The environmental contribution to the mutual information as a function of D_-/D_+ and w_-/w_+ in a quartic potential and in the slow-jumps limit. For all plots, $\tau = 1$. (a) The quartic potential considered here. [(b) and (c)] Contour plots of the joint probability distribution and its factorization, respectively, for $D_+ = 10$, $D_- = 10^{-2}$, $w_- = w_+$. Notice that the marginalized probability has much longer tails along the axis $x = 0$ and $y = 0$. (d) The colored surface is the result of a Monte Carlo integration with importance sampling of the mutual information. In the $D_-/D_+ \rightarrow 0$ limit, I_{env} approaches H_{jumps} , the black dashed line, which is also its maximum value. (e) Compared with the linear case, the nonlinear relaxation reflects into a considerably slower convergence toward H_{jumps} of the mutual information.

marginal distribution. The divergences between the analogous components of the joint distribution are, respectively, $h_C^2(x) = 2h_C^1(x)$ and $h_D^2(x) = 2h_D^1(x)$. The upper bound $I_{\text{env}}^{\text{up}}$ is identical to Eq. (12), with the exchange $h_C \leftrightarrow h_D$.

Crucially, these bounds converge to the same limits of the linear case, namely,

$$I_{\text{env}}\left(\frac{D_-}{D_+}, \frac{w_-}{w_+}\right) = \begin{cases} H_{\text{jumps}} & \text{if } D_-/D_+ \ll 1, \\ 0 & \text{if } D_-/D_+ \approx 1, \end{cases} \quad (13)$$

but their convergence rate is slower than the one obtained in the linear regime [see Fig. 1(e)]. This is perhaps unsurprising, since the nonlinear relaxation increases the typical autocorrelation timescale and thus reduces the impact of environmental changes. In Fig. 1(d), we show the mutual information in this slow-jumps limit, computed via importance sampling [22]. In particular, we sample the components of the joint distribution starting from the potential in Eq. (9) via Hamiltonian Monte-carlo [23,24]. Then, each component is weighted according to the stationary distribution of the environment, for any given w_-/w_+ , to obtain samples of Eq. (5). In Fig. 1(e) we see that, at a given value of D_-/D_+ , the mutual information due to the environment is typically smaller than the case of linear relaxation.

It is possible to show that, for any potential of the form $U(x_\mu) \propto x_\mu^{2n}$, with n a positive integer, the bounds in Eq. (12) always saturate to H_{jump} when $D_-/D_+ \rightarrow 0$ and vanish when $D_- \rightarrow D_+$ (see Appendix B). This result, which is consistent with the one presented for linear interactions in Ref. [17], remarks that, when the variability of the environment is maximal, any two noninteracting degrees of freedom share the

information contained into the Shannon entropy associated with the external jump process, H_{jumps} . Importantly, in all these noninteracting cases, the only dimensionless parameters we can build are once more w_-/w_+ , which determines the persistence of the two environmental states, and D_-/D_+ , describing how similar the environmental states are. Although the probability distributions in Eq. (11) and Eq. (10) do not depend only on such combinations, the mutual information does (see also Ref. [17]).

V. MUTUAL INFORMATION IN NONLINEAR POTENTIALS

We now consider nonlinear interactions between the two particles. To keep things analytically tractable, we assume that the drift term in Eq. (2) can be written as the gradient of a potential of the form

$$V(x_1, x_2) = \sum_{\mu} U(x_{\mu}) + V_{\text{int}}(x_1, x_2), \quad (14)$$

where $U(x_{\mu}) = x_{\mu}^4/4\tau$, so that the solution of Eq. (4) is given by $P_i^{\text{st}}(x_1, x_2) \propto \exp[-V(x_1, x_2)/D_i]$. Hence, we are focusing on equilibrium systems with nonlinear relaxation and nonlinear interactions. In particular, V_{int} depends on parameters that may combine with τ and D_i to form a set of dimensionless parameters $\{\psi\}$. In what follows, it will be useful to distinguish between those combinations that include D_i , which we call $\{\psi_i\}$, and those that only depend on the parameters of the potential V , termed $\{\psi_{\text{int}}\}$.

For convenience, let us identify all the different contributions to the mutual information of the overall system. First, we call I_{env} the mutual information stemming from the shared environment alone, which we obtained in the previous section. It can only depend on the environmental dimensionless parameters, D_-/D_+ and w_-/w_+ . Second, we call $I^{\text{int}}(\{\psi_{\text{int}}\}, \{\psi_i\})$ the mutual information stemming from the components $P_i^{\text{st}}(x_1, x_2)$ of the joint distribution, which can only depend on $\{\psi_{\text{int}}\}$ and on $\{\psi_i\}$. Given that such i th dependence may be present, we write in full generality

$$I_{\text{pot}}(\{\psi\}) = \sum_i \pi_i I^{\text{int}}(\{\psi_{\text{int}}\}, \{\psi_i\}). \quad (15)$$

This term stems from the presence of internal interactions, since in the absence of V_{int} such components are factorizable and their mutual information is zero. Notice that this form is consistent with the first term of the upper bound in Eq. (8) and, if no dimensionless parameter depends on i , it reduces to $I_{\text{pot}} = I^{\text{int}}(\{\psi_{\text{int}}\})$ which is the correct expression for linear interactions [17].

Therefore, we choose to decompose the mutual information of the overall system as

$$I\left(\frac{w_-}{w_+}, \frac{D_-}{D_+}, \{\psi\}\right) = I_{\text{env}}\left(\frac{w_-}{w_+}, \frac{D_-}{D_+}\right) + I_{\text{pot}}(\{\psi\}) + \Xi\left(\frac{w_-}{w_+}, \frac{D_-}{D_+}, \{\psi\}\right), \quad (16)$$

where the term Ξ quantifies the contributions due to the presence of both the environment and the interactions at once. In general, this is not a mutual information, i.e., it needs not to be positive, and may depend on all dimensionless parameters. For these reasons, we name this term as *information interference*. In Ref. [17], we showed that this term is always zero in the case of linear interactions. However, we often cannot find analytical expressions for all the terms in Eq. (16), and we have to resort to numerical integration.

Hence, unless otherwise specified, the mutual information in Eq. (16) is obtained as outlined in the previous section. We employ Hamiltonian Montecarlo to sample the joint distribution associated with the potential in Eq. (14), and we weight these samples according to the corresponding mixture distribution, Eq. (5). Then, the mutual information integral is evaluated by importance sampling. Crucially, importance sampling requires the knowledge of the analytical expressions of both the joint and the marginal mixture components [22], which we need to compute for every choice of the potential. Hence, we need to solve Eq. (4) to find the expression for $P_i^{\text{st}}(x_1, x_2)$, as well as its marginalization.

A. Destructive information interference

We first study the case

$$V(x_1, x_2) = \frac{x_1^4 + x_2^4}{4\tau} - g \frac{x_1^2 x_2^2}{2} := V_{\text{sw}}(x_1, x_2), \quad (17)$$

where for stability $g > 0$. This single-well potential, depicted in Fig. 2(a), has one stable minima at $(x_1, x_2) = (0, 0)$. In

the slow-jumps limit, the mixture components of the joint distributions follows a Boltzmann-like distribution

$$P_i^{\text{st}}(x_1, x_2) = \frac{1}{\mathcal{N}_{\text{sw}}} e^{-V_{\text{sw}}(x_1, x_2)/D_i}, \quad (18)$$

where the normalization \mathcal{N}_{sw} can be computed analytically. The corresponding marginal components are

$$P_i^{\text{st}}(x_\mu) = \sqrt{\frac{g x_\mu^2 \tau}{2 \mathcal{N}_{\text{sw}}^2}} K_{\frac{1}{4}}\left(\frac{g^2 x_\mu^4 \tau}{8 D_i}\right) e^{(-2+g^2 \tau^2) \frac{x_\mu^4}{8 D_i \tau}}, \quad (19)$$

where $K_n(x)$ is the modified Bessel function of the second kind. We plot the corresponding mixture distributions in see Figs. 2(b) and 2(c).

With this choice of the potential, the only dimensionless parameters appearing in the mutual information are D_-/D_+ , w_-/w_+ , and $g\tau$. The first two belong to the environment, while the last one is the sole quantity characterizing the interactions. That is, we expect that $I_{\text{pot}} = I_{\text{sw}}^{\text{int}}(g\tau)$, although it is not possible to find an exact expression for this term. Notice that we only have to inspect the dynamics for a fixed environment to determine the dimensionless relevant quantities, and add w_-/w_+ that modulates the mixture in the slow-jumps limit. In other words, and as for the linear case, the mutual information of the joint distribution, I_{sw} , cannot depend separately on D_- and D_+ thus being independent of the environmental state.

Consequently, we write I_{sw} as

$$I_{\text{sw}}\left(\frac{w_-}{w_+}, \frac{D_-}{D_+}, \tau g\right) = I_{\text{env}}\left(\frac{w_-}{w_+}, \frac{D_-}{D_+}\right) + I_{\text{sw}}^{\text{int}}(\tau g) + \Xi_{\text{sw}}\left(\frac{w_-}{w_+}, \frac{D_-}{D_+}, \tau g\right). \quad (20)$$

Let us investigate separately the impact of interactions and environmental changes in this example. First, the effect of the interactions in the joint probability distribution reflects into the appearance of tails along the axes $x_1 = 0$ and $x_2 = 0$ [see Figs. 2(b) and 2(c)]. The higher is g , the longer the tails. Conversely, the environment affects the joint distribution by suppressing such tails as the difference between D_- and D_+ becomes more pronounced. Since these two terms operate in opposite ways, the mutual information takes contrasting contributions. As a consequence of this interplay between changing environment and nonlinear interactions, in Figs. 2(d)–2(f), we see the mutual information of the overall system, I_{sw} , is always smaller than the sum of I_{env} and $I_{\text{sw}}^{\text{int}}$, and can also be smaller than I_{env} for some values of $(w_-/w_+, D_-/D_+)$. This means that $\Xi_{\text{sw}} \leq 0$ in the entire space, and $\Xi_{\text{sw}} < -I_{\text{sw}}^{\text{int}}$ in some regions of the parameter space [see Fig. 2(f)]. Naively speaking, nonlinear interactions can mask environmental information, by counteracting the dependency induced by a switching environment and effectively reducing the information that x_1 and x_2 share. We name this phenomenon *destructive information interference*.

However, the limiting behaviors of I_{sw} can be understood as discussed in Ref. [17], and the disentangling is possible provided some knowledge of the environmental states in these regimes. Indeed, when $D_-/D_+ \rightarrow 1$, the only contribution to the mutual information comes from the interactions alone, $I_{\text{sw}}^{\text{int}}(\tau g)$. Similarly, in the opposite limit $D_-/D_+ \rightarrow 0$, the

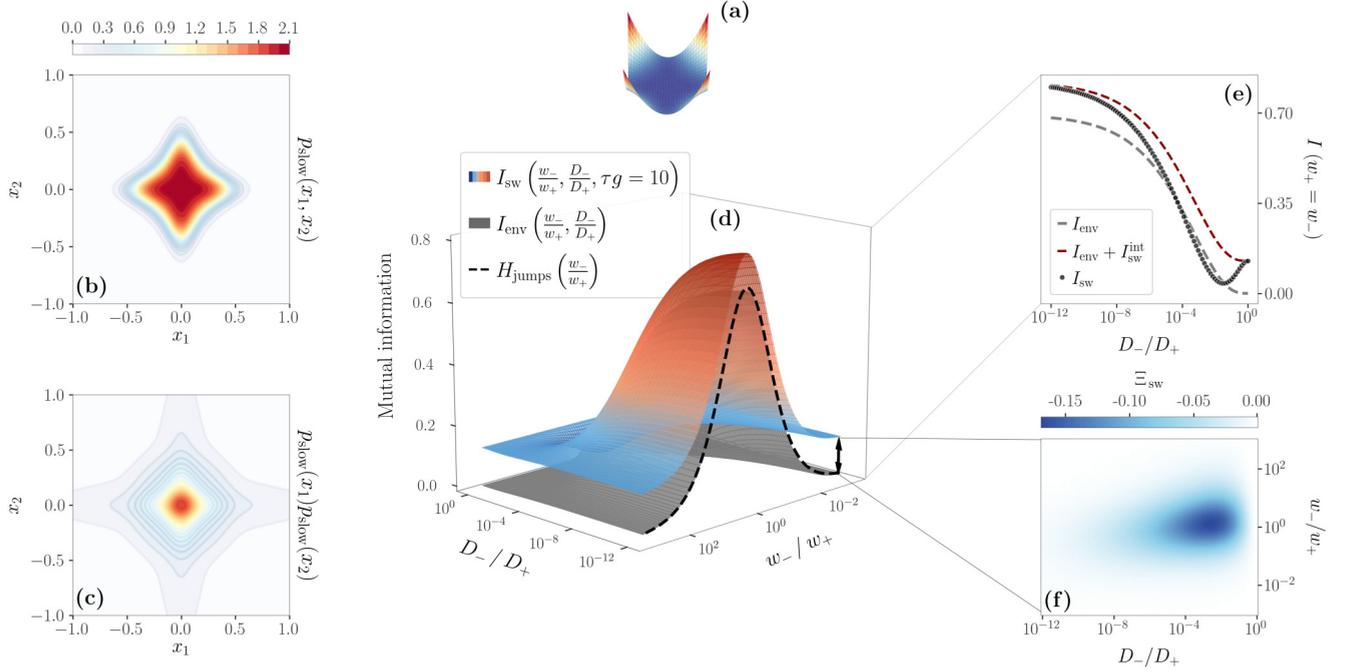


FIG. 2. The total mutual information I_{sw} in the single-well case, as a function of D_-/D_+ and w_-/w_+ , in the slow-jumps limit. For all plots, $\tau g = 10$. (a) The single-well potential considered here. [(b) and (c)] Contour plots of the joint probability distribution and its factorization, respectively, for $D_+ = 10$, $D_- = 10^{-2}$, $w_- = w_+$. The marginalized probability has much longer tails along the axis $x = 0$ and $y = 0$, which tend to be suppressed by the environment. However, these tails are still present in the joint probability as a consequence of the interactions. (d) The colored surface is the result of a Monte Carlo integration with importance sampling of I_{sw} , whereas the gray surface represent the environmental contribution alone I_{env} . (e) I_{sw} (black dots) can be smaller than I_{env} (gray dashed line) and, in general, it is lower than the sum of $I_{\text{sw}}^{\text{int}}$ and I_{env} (red dashed line). (f) In fact, the term Ξ_{sw} is always negative, showing that the effects of the environment and of the interactions are reciprocally masked at low-enough values of D_-/D_+ . For $D_-/D_+ \rightarrow 0, 1$ we find $\Xi_{\text{sw}} = 0$, hence the environmental and the interactions contributions are disentangled.

numerical integration shows that the two contributions to the mutual information are exactly disentangled, i.e.,

$$I_{\text{sw}}\left(\frac{w_-}{w_+}, \frac{D_-}{D_+}, \tau g\right) = \begin{cases} H_{\text{jumps}} + I_{\text{sw}}^{\text{int}}(\tau g) & \text{if } \frac{D_-}{D_+} \ll 1 \\ I_{\text{sw}}^{\text{int}}(\tau g) & \text{if } \frac{D_-}{D_+} \approx 1 \end{cases}, \quad (21)$$

that means that in both limits $\Xi_{\text{sw}} \rightarrow 0$. In Fig. 2(e), we compare this fully disentangled form (in red) with the mutual information at fixed w_-/w_+ and for different values of D_-/D_+ (black dots). We see that indeed this disentangling is only achieved in the limits in Eq. (21), whereas at intermediate values of D_-/D_+ destructive interference reduces I_{sw} .

B. Constructive information interference

In the previous section, we argued that the destructive information interference stems from the fact that interactions and environment operate on the same axes in opposite ways. Indeed, we now show that a rotation of the interaction term in Eq. (17) of an angle $\pi/4$ generates instead a cooperation of the two terms that can boost the overall mutual information. In analogy with the previous case, this feature is named *constructive information interference*.

Thus, the potential governing the system, shown in Fig. 3(a), is

$$V_{\text{swr}}(x_1, x_2) = \frac{x_1^4 + x_2^4}{4\tau} - g \frac{\|R_{\pi/4}(x, y)\|^2}{2}, \quad (22)$$

where R_θ is the rotation matrix of angle θ , $\|\cdot\|^2$ is the L_2 norm and $g > 0$.

In this scenario, the joint and marginal mixture components in the slow-jumps limit can be again found analytically, and are given by

$$P_i^{\text{st}}(x_1, x_2) = \frac{1}{\mathcal{N}_{\text{swr}}} e^{-V_{\text{swr}}(x_1, x_2)/D_i} \quad (23)$$

and

$$P_i^{\text{st}}(x_\mu) = \sqrt{\frac{g\pi^2}{2\alpha}} \frac{|x_\mu| [I_{-\frac{1}{4}}(\beta_i x_\mu^4) + I_{\frac{1}{4}}(\beta_i x_\mu^4)]}{2\mathcal{N}_{\text{swr}}} e^{\frac{x_\mu^4 (g^2 - 128\alpha)}{128\alpha D_i}}, \quad (24)$$

where $I_n(x)$ is the modified Bessel function of the first kind, $\alpha = \tau^{-1}/4 + g/8$ and $\beta_i = g^2/(128\alpha D_i)$.

As in the previous case, the dimensionless parameters appearing in the mutual information are w_-/w_+ , D_-/D_+ , and $g\tau$. This implies once more that in Eq. (16) we have $I_{\text{pot}} = I_{\text{swr}}^{\text{int}}(g\tau)$. Hence, the interference term in Eq. (16), Ξ_{swr} , will again depend on all of them. However, in this case, the role of the interactions is to introduce tails along the bisectors of the (x_1, x_2) plane, whereas the environment keeps acting on

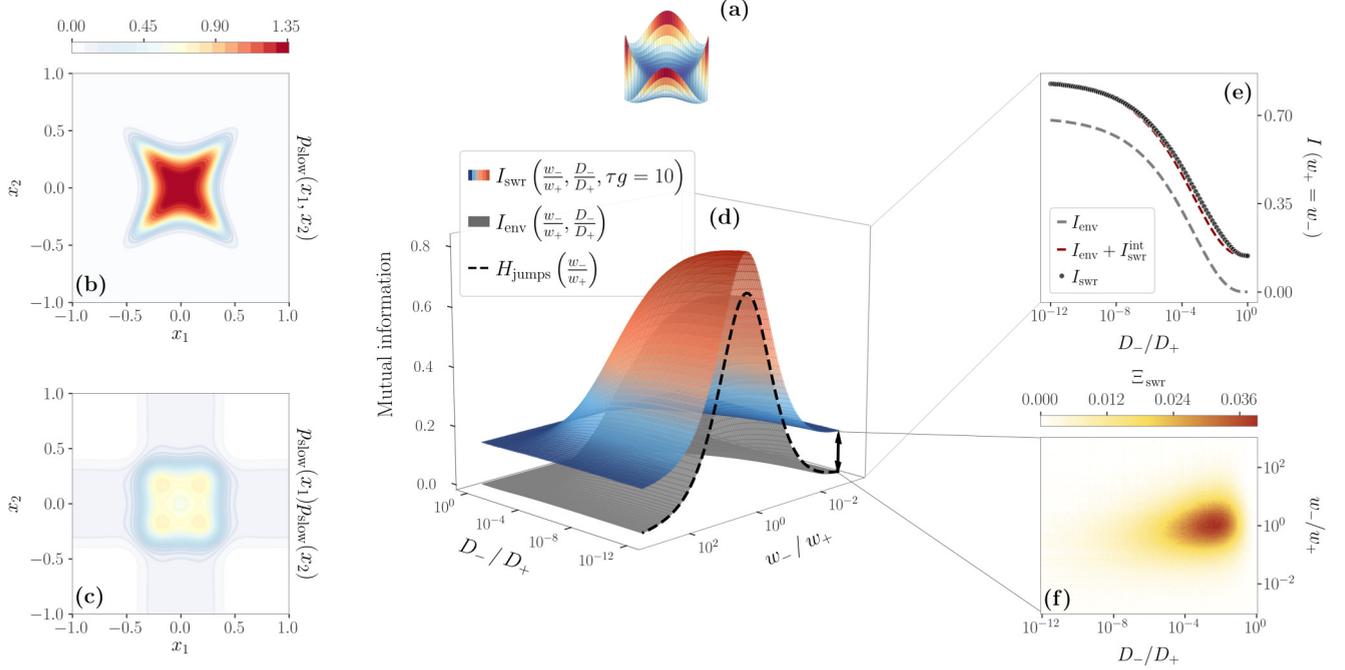


FIG. 3. The total mutual information I_{swr} in the rotated single-well case, as a function of D_-/D_+ and w_-/w_+ , in the slow-jumps limit. For all plots, $\tau g = 10$. (a) The rotated single-well potential considered here. [(b) and (c)] Contour plots of the joint probability distribution and its factorization, respectively, for $D_+ = 10$, $D_- = 10^{-2}$, $w_- = w_+$. The marginalized probability has much longer tails along the axis $x = 0$ and $y = 0$, which are suppressed in the joint probability as a consequence of the environment. Contrary to the single-well case, the interactions here trigger the presence of tails along the bisectors of the (x, y) plane. (d) The colored surface is the result of a Monte Carlo integration with importance sampling of I_{swr} , whereas the gray surface represent the environmental contribution alone I_{env} . (e) I_{swr} (black dots) is always greater than I_{env} (gray dashed line) and, in general, it is greater than the sum of $I_{\text{swr}}^{\text{int}}$ and I_{env} (red dashed line). (f) The term Ξ_{swr} is always positive, and in particular it is different from zero at high enough values of D_-/D_+ . For $D_-/D_+ \rightarrow 0$, we end up with $\Xi_{\text{swr}} = 0$ as expected.

the $x_1 = 0$ and $x_2 = 0$ axes. Hence, nonlinear interactions do not counteract the dependency induced by the environment. As a consequence, as shown in Figs. 3(d) and 3(e), the mutual information of the overall system is very close to the sum of the environmental and the interaction terms, i.e., $\Xi_{\text{swr}} \approx 0$. Moreover, there is a region in the parameter space in which $\Xi_{\text{swr}} > 0$, meaning that x_1 and x_2 share more information than the one coming from the changing environment and their sheer couplings.

We remark that the limiting behaviors of the mutual information exhibit an exact disentangling, as before. Thus, when $D_-/D_+ \rightarrow 0$, $\Xi_{\text{swr}} \rightarrow 0$, and $I_{\text{env}} \rightarrow H_{\text{jumps}}$, while for $D_- \rightarrow D_+$ only $I_{\text{swr}}^{\text{int}}$ survives.

C. Information peaks in bistable systems

As a last example, here we consider the slightly more complex case of a bistable system. In particular, we have the following potential:

$$V_{\text{dw}}(x_1, x_2) = \frac{x_1^4 + x_2^4}{4\tau} - gxy, \quad (25)$$

where $g > 0$. This potential, depicted in Fig. 4(a), has two stable minima at $(x_1, x_2) = (\pm\sqrt{g\tau}, \pm\sqrt{g\tau})$. The joint and the

marginal mixture components in the slow-jumps limit are

$$P_i^{\text{st}}(x_1, x_2) = \frac{1}{\mathcal{N}_{\text{dw}}} e^{-V_{\text{dw}}(x_1, x_2)/D_i} \quad (26)$$

and

$$P_i^{\text{st}}(x_\mu) = \left[(D_i^9 \tau)^{\frac{1}{4}} \Gamma\left(\frac{1}{4}\right) {}_0F_2\left(\frac{1}{2}, \frac{3}{4}; \alpha_i x_\mu^4\right) + g^2 x_\mu^2 (D_i \tau)^{\frac{3}{4}} \right. \\ \left. \times \Gamma\left(\frac{3}{4}\right) {}_0F_2\left(\frac{5}{4}, \frac{3}{2}; \alpha_i x_\mu^4\right) \right] \frac{e^{-\frac{x_\mu^4}{4D_i \tau}}}{\sqrt{2} D_i^2 \mathcal{N}_{\text{dw}}}, \quad (27)$$

where ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x)$ is the generalized hypergeometric function, $\alpha_i = g^4 \tau / (64 D_i^3)$, and \mathcal{N}_{dw} can be found analytically. As we can see in Figs. 4(b) and 4(c), the joint probability distribution has two peaks corresponding to the two minima of the potential, whereas the factorized distribution presents four peaks with connections among them that reflect the influence of a switching environment.

The first crucial difference between this case and the previous ones is that the dimensionless parameters appearing in the mutual information mix environmental and interaction features. Indeed, they are w_-/w_+ , $D_-/\tau g^2$, and $D_+/\tau g^2$, with D_-/D_+ resulting from a combination of the last two parameters. Hence, differently from the previous cases, we can write

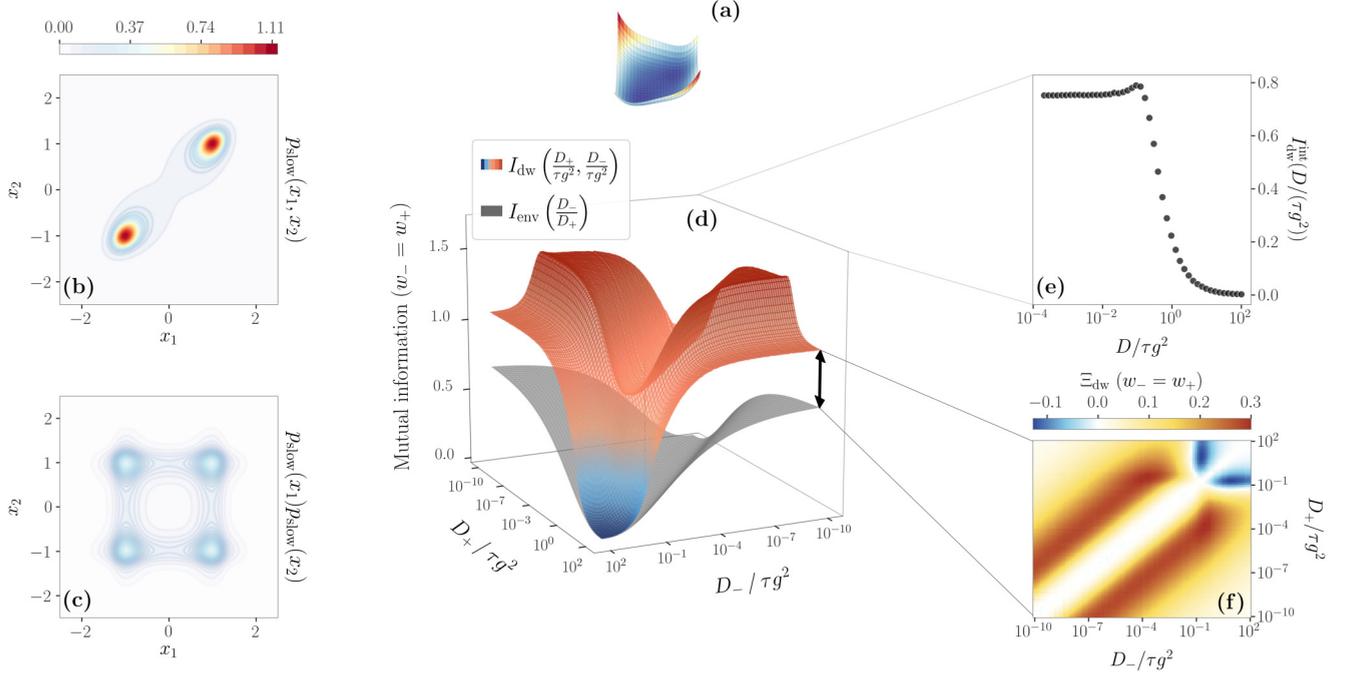


FIG. 4. The total mutual information I_{dw} in the double-well case, as a function of the adimensional parameters $D_{\pm}/(\tau g^2)$, in the slow-jumps limit. For all plots, $w_-/w_+ = 1$, $\tau = g = 1$. (a) The double-well potential considered here. [(b) and (c)] Contour plots of the joint probability distribution and its factorization, respectively, for $D_+ = 10$, $D_- = 10^{-2}$. The joint probability has two peaks, corresponding to the two minima of the potential. On the contrary, the marginalized probability is markedly different, with four peaks. (d) The colored surface is the result of a Monte Carlo integration with importance sampling of I_{dw} , whereas the gray surface represents the environmental contribution alone I_{env} , in the plane $(D_-/\tau g^2, D_+/\tau g^2)$. (e) The mutual information $I_{\text{dw}}^{\text{int}}$ of the interactions only. At large D , we expect the two minima to be less relevant, and indeed the mutual information vanishes. At small D , instead, $I_{\text{dw}}^{\text{int}}$ is markedly different from zero since the particles are typically trapped in one of the two minima. At intermediate values the mutual information peaks due to an interplay between trapping and diffusion. (e) The term Ξ_{dw} can be either positive and negative, meaning that, at different values of $D_{\pm}/(\tau g^2)$, we find both constructive and destructive interference. Clearly, when $D_- \approx D_+$, we have $\Xi_{\text{dw}} \approx 0$.

Eq. (16) as

$$I_{\text{dw}}\left(\frac{w_-}{w_+}, \frac{D_-}{\tau g^2}, \frac{D_+}{\tau g^2}\right) = I_{\text{env}}\left(\frac{w_-}{w_+}, \frac{D_-}{D_+}\right) + \sum_i \pi_i I_{\text{dw}}^{\text{int}}\left(\frac{D_i}{\tau g^2}\right) + \Xi_{\text{dw}}\left(\frac{w_-}{w_+}, \frac{D_-}{\tau g^2}, \frac{D_+}{\tau g^2}\right) \quad (28)$$

and we plot it in Fig. 4(d).

The dependence on the diffusion coefficient of the interaction term, $I_{\text{dw}}^{\text{int}}$, shown in Fig. 4(e), can be explained on an intuitive basis. Indeed, the distance between $P_i^{\text{st}}(x_1, x_2)$ and its factorization receives the most contributions from the fact that the latter has four peaks, due to the implicit assumption of independence between x_1 and x_2 . However, when D is large, the system can easily escape the potential minima, and thus they will not contribute to $I_{\text{dw}}^{\text{int}}$, which vanishes as D grows.

Conversely, small values of D weight more the potential minima, since the system is substantially trapped in them. In this limit, $I_{\text{dw}}^{\text{int}}$ converges to a nonzero value due to the fact that only two of the peaks of the factorized distribution are present in the joint distribution. Finally, we observe an emerging peak of $I_{\text{dw}}^{\text{int}}$ at a finite value of D . This optimal diffusion naively allows the system to explore both minima from time to time, still being trapped for a consistent amount of time during each stochastic realization.

In Fig. 4(f), we report Ξ_{dw} for the specific case $w_- = w_+$. All other choices do not qualitatively change the picture. In this scenario, Ξ_{dw} can be either positive and negative, exhibiting a nontrivial pattern of constructive and destructive information interference. This pattern, although hard to understand analytically, is intuitively a consequence of the system switching from a state in which it is trapped in one single minimum, to a state in which it can freely explore larger regions of the (x_1, x_2) plane.

We also remark that in this case it is difficult to define the usual limiting behaviors of the mutual information of the overall system in which the disentangling is recovered. Indeed, D_-/D_+ is not the only relevant parameter of the system and the limit $D_-/D_+ \rightarrow 0$ is not particularly informative anymore.

VI. MUTUAL INFORMATION IN NONEQUILIBRIUM ENVIRONMENTS

So far, we investigated systems coupled with an ever-changing environment that eventually relax to equilibrium. However, nature usually operates out-of-equilibrium, and most of the environments of biochemical, neural, and ecological systems are in nonequilibrium conditions. To study a minimal model encompassing this additional feature, we

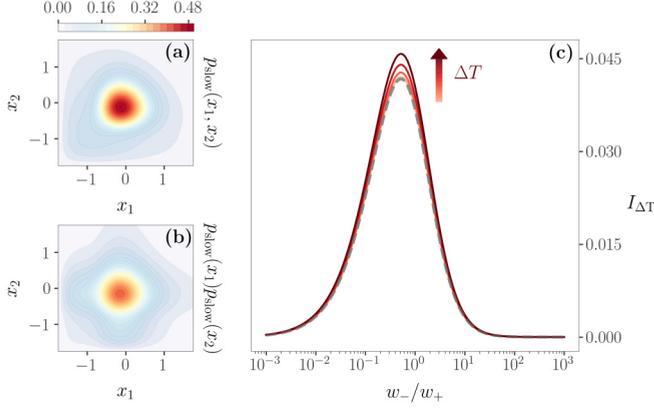


FIG. 5. Mutual information in the presence of multiplicative noise. [(a) and (b)] The joint and the marginal probability distributions in the slow-jumps limit. (c) As we increase the temperature gradient ΔT , the mutual information increases. Here the different curves are for $\Delta T = \{0.15, 0.29, 0.43, 0.57\}$, $T_0 = \tau = 1$, $\gamma_+ = 1.5$, and $\gamma_- = 0.5$.

consider the following dynamics:

$$\dot{x}_\mu = -\frac{1}{\tau}x_\mu + \gamma_{i(t)}\sqrt{2T(x_\mu)}\xi_\mu, \quad (29)$$

with $i(t)$ is a realization of the stochastic process governing the environment, $\mu = 1, 2$, and $T(x_\mu) = T_0 + x_\mu \Delta T$ is a linear temperature gradient, for the sake of simplicity. This picture can capture the key features of a diffusing molecule that can live in two conformational states [25–28]. Alternatively, it can be a simple way to describe proteins in an environment with patches of different density (e.g., liquid condensates [12,29]) subject to an external gradient. Since diffusion and temperature are connected by the Einstein relation, $D(x) \propto \gamma T(x)$, here the environment may act as a modification of the viscosity, in the case of patches of different density, or the motility, when the switching describes two different conformational states. These diffusive properties are encoded into $\gamma_{i(t)}$ that can take two values, γ_- and γ_+ , replacing the role of D_- and D_+ of the previous models.

The peculiarity of this model is the presence of a multiplicative noise proportional to x_μ . Since there are no interactions, the joint component is just the product of the two mixture components. Here $x_\mu \in [-T_0/\Delta T, +\infty]$, and $\Delta T < \sqrt{T_0/\gamma\tau}$ to ensure flat derivatives at the boundaries, so that no particles can escape the system. Hence, the mixture components read:

$$P_i(x_\mu) = \mathcal{N}e^{-\frac{1}{\tau}\frac{x_\mu}{\gamma\Delta T}} \left(1 + \frac{\Delta T}{T_0}x_\mu\right)^{\frac{1}{\tau}\frac{T_0/\Delta T}{\gamma\Delta T}-1}. \quad (30)$$

In Figs. 5(a) and 5(b), we respectively show the joint and factorized distributions of this system, that are both symmetric with respect to the bisector $x_1 = x_2$. In this case, we compute the mutual information of $p_{\text{slow}}(x_1, x_2)$ as a standard numerical integral, which we plot in Fig. 5(c), as a function of

w_-/w_+ for increasing ΔT . We can effectively conclude that the presence of a multiplicative noise increase the shared information between two (noninteracting) degrees of freedom.

Finally, as in the presence of a non multiplicative noise, the mutual information vanishes when $\gamma_- \rightarrow \gamma_+$, whereas it converges to H_{jumps} for $\gamma_-/\gamma_+ \rightarrow 0$, hence preserving the limiting behaviors that are crucial to perform an exact disentangling [17].

VII. MUTUAL INFORMATION IN CONTINUOUS ENVIRONMENTS

Finally, we consider the case in which two particles are not interacting, but share the same continuous environment [15,19]. To fix the ideas, let us consider the paradigmatic example of two Ornstein-Uhlenbeck processes,

$$\begin{cases} \dot{x}_\mu = -x_\mu/\tau_X + \sqrt{2D}\xi_\mu \\ \dot{D} = -D/\tau_D + \sqrt{2\theta}\xi_D \end{cases}, \quad (31)$$

where the only adimensional parameter of the system is now τ_X/τ_D , which governs the timescale separation of the two dynamics. Hence, contrary to the case of a discrete-state environment, we cannot define the separation between environmental states (previously quantified by D_-/D_+) nor their relative persistence, which was given by w_-/w_+ .

The corresponding stationary Fokker-Planck equation is given by

$$\begin{aligned} 0 = \sum_{\mu=1}^2 \left[\partial_\mu \left(\frac{x_\mu}{\tau_X} p(\mathbf{x}, D) \right) + D^2 \partial_\mu^2 p(\mathbf{x}, D) \right] \\ + \partial_D \left[\frac{D}{\tau_D} p(\mathbf{x}, D) \right] + \theta \partial_D^2 p(\mathbf{x}, D) \end{aligned} \quad (32)$$

and, as before, we are interested in the marginalization $p(\mathbf{x}, t) = \int dD p(\mathbf{x}, D, t)$. Notably, if we explicitly marginalize Eq. (32), then at stationarity we obtain

$$0 = \sum_{\mu=1}^2 \left[\partial_\mu \left(\frac{x_\mu}{\tau_X} p(\mathbf{x}) \right) + \partial_\mu^2 (\hat{D}^2(\mathbf{x}) p(\mathbf{x})) \right], \quad (33)$$

where $\hat{D}^2(\mathbf{x}) = \int dD D^2 p(D|\mathbf{x})$ is an effective spatial diffusion coefficient. Therefore, we can interpret the effective dependencies induced by the environment as arising from an inhomogeneous medium, rather than associated with effective couplings between x_1 and x_2 . It is also worth noting that, in principle, space-dependent diffusion coefficients, interpreted in the Ito sense, might always emerge from the variations of an external stochastic environment, which is also the sole responsible for a nonzero mutual information. This result might shed some light on the controversial topic of the Ito-Stratonovich dilemma in diffusing chemical systems. A similar perspective,

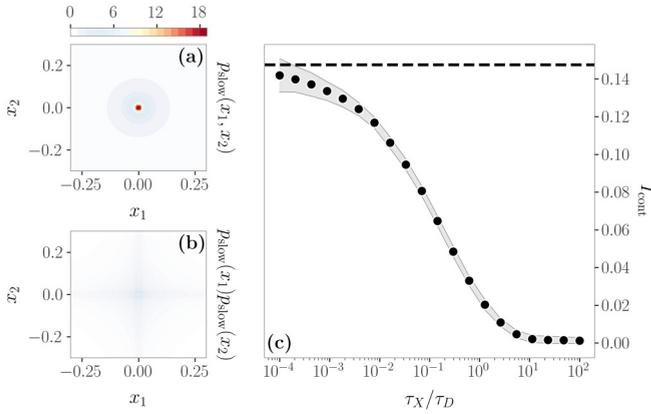


FIG. 6. The mutual information associated to the Langevin equations in Eq. (31). [(a) and (b)] Plots of the joint and the factorized distribution in the limit of a much slower environment, $\tau_X/\tau_D \ll 1$. (c) Mutual information at different values of τ_X/τ_D estimated through a k -nearest-neighbors estimator. As expected, in the limit $\tau_X/\tau_D \ll 1$ the mutual information converges to Eq. (36), whereas it vanishes in the opposite limit.

where the internal states play an analogous role of a changing environment, is presented in Ref. [28].

In the limit in which the environment is either much faster or much slower than the internal relaxation, i.e., respectively $\tau_X/\tau_D \gg 1$ and $\tau_X/\tau_D \ll 1$, we can repeat the calculations of Sec. I. In the presence of a slower environment, we find the following stationary joint probability distribution

$$p_{\text{slow}}(x_1, x_2) = \int_{-\infty}^{+\infty} dD p^{\text{st}}(D) p^{\text{st}}(x_1, x_2 | D) \\ = \frac{1}{2\pi \sqrt{(x_1^2 + x_2^2) \theta \tau_D \tau_X}} e^{-\sqrt{\frac{x_1^2 + x_2^2}{\theta \tau_D \tau_X}}}, \quad (34)$$

where $p^{\text{st}}(D) \sim \mathcal{N}(0, \tau_D \theta)$ is the stationary distribution of the diffusion coefficient and $p^{\text{st}}(x_1, x_2 | D) \sim \mathcal{N}(0, \tau_X D^2)$ is the stationary distribution of (x_1, x_2) at fixed D . Eq. (34) can be marginalized exactly over one of the two degrees of freedom, in order to evaluate the mutual information. The marginalization leads to

$$p_{\text{slow}}(x_\mu) = \frac{1}{\pi \sqrt{\theta \tau_D \tau_X}} K_0\left(\frac{|x_\mu|}{\sqrt{\theta \tau_D \tau_X}}\right). \quad (35)$$

These probability distributions are plotted in Figs. 6(a) and 6(b). The joint probability in Eq. (34) is not factorizable, and thus we expect that in this limit the mutual information of $p_{\text{slow}}(x_1, x_2)$, $I_{\text{cont}}^{\text{slow}}$, will be different from zero due to the shared environment. Since in this case there are no dimensionless parameters characterizing environmental dynamics, we also expect no parametric dependence. Indeed, we can immediately rewrite the mutual information integral after the change of variable $(x_1, x_2) = \sqrt{\theta \tau_D \tau_X} (s \cos \phi, s \sin \phi)$ as

$$I_{\text{cont}}^{\text{slow}} = \gamma_E + \log \frac{\pi}{2} - 1 \\ - \frac{1}{\pi} \int_0^\infty ds e^{-s} \int_0^{2\pi} d\phi \log K_0(s |\cos \phi|), \quad (36)$$

where γ_E is the Euler's constant. The numerical value is $I_{\text{cont}}^{\text{slow}} \approx 0.148$.

In the opposite limit, $\tau_X/\tau_D \gg 1$, since $\langle D \rangle_{p^{\text{st}}(D)} = 0$ we trivially find that $p_{\text{fast}}(x_1, x_2) = \prod_\mu \delta(x_\mu)$. Therefore, the mutual information vanishes in this limit. At intermediate values of τ_X/τ_D , we cannot solve Eq. (32) exactly. Therefore, to obtain samples from the stationary joint distribution, we simulate the Langevin equations Eq. (31). Then, from these samples, we estimate the mutual information through the k -nearest-neighbors estimator proposed in [30,31]. The results are plotted in Fig. 6(c). As expected, the mutual information changes smoothly with τ_X/τ_D and, in the limit $\tau_X/\tau_D \rightarrow 0$, approaches Eq. (36).

VIII. CONCLUSIONS

In this work, we showed that tackling the information properties of complex systems in changing environment is a feasible task, even in the presence of nonlinear interactions, nonequilibrium conditions, and continuously varying environments.

In particular, in the presence of nonlinear couplings, the resulting information structure can be interpreted as an interplay between the effects of internal interactions and environmental changes. This interplay can be generically quantified by an information interference term, which surprisingly cancels exactly in the case of linear interactions [17].

Moreover, we showed that continuously varying environments can be mapped into an effective spatial diffusion coefficient. This result might be a crucial step to understand under which conditions a shared changing environment generates effective couplings, and in which ones it does not. Additionally, the emergence of an effective space-dependent diffusion from external couplings might shed some light on the Ito-Stratonovich dilemma when describing biological and biochemical systems in inhomogeneous media.

Our results have important implications in settings where we expect nonlinear or out-of-equilibrium effects to be crucial, such as neural activity originated by external stimulation [32–34] or population growth [35–37]. Notably, it was shown that, in models with latent variables, phenomenological renormalization group approaches can give seemingly nontrivial results [38,39]. Such models are formally similar to the framework of a changing environment analyzed here, and future works should be devoted to understand the relation between our results and the underlying information properties of these models.

Further, concepts such as mutual information and disentangled representations of the data are particularly relevant in the context of machine learning [40–42]. Indeed, it will be paramount to unravel how these approaches might benefit from the results presented in this work.

Ultimately, we believe that this work highlights criticalities and potentialities of an information-theoretic approach to study more general and complex real-world systems. In particular, the unforeseen findings presented here might reveal, in the future, surprising properties of the information structure of complex systems with far-reaching consequences in different interdisciplinary fields.

APPENDIX A: DERIVATION OF THE BOUNDS ON THE JOINT ENTROPY

Here we briefly derive the bounds on the joint entropy [16] in the slow-jumps limit,

$$H_{\mathbf{X}} = - \int d\mathbf{x} p_{\text{slow}}(\mathbf{x}) \log p_{\text{slow}}(\mathbf{x}), \quad (\text{A1})$$

where p_{slow} is the mixture distribution defined by Eq. (5). This entropy lacks in general an analytical expression, but it is often the case that we can compute the entropy of the i th component of p_{slow} ,

$$H_{\mathbf{X}}^i = - \int d\mathbf{x} P_i^{\text{stat}}(\mathbf{x}) \log P_i^{\text{stat}}(\mathbf{x}). \quad (\text{A2})$$

Let us now call $H_{\mathbf{X},E}$ the joint entropy between the system and the environment, where the environment is represented by the set of probabilities $\{\pi\}$ appearing in Eq. (5). Then, the properties of the conditional entropy allow us to write

$$H_{\mathbf{X}} \geq H_{\mathbf{X}|E}, \quad (\text{A3})$$

where

$$H_{\mathbf{X}|E} = \sum_i \pi_i H_{\mathbf{X}}^i. \quad (\text{A4})$$

Similarly, we can write

$$H_{\mathbf{X}} \leq H_{\mathbf{X},E} = H_{\mathbf{X}|E} + H_E, \quad (\text{A5})$$

where $H_E = - \sum_i \pi_i \log \pi_i = H_{\text{jumps}}$. In the case of two variables presented in the main text, $H_{\mathbf{X}} = H_{12}$.

APPENDIX B: DERIVATION OF THE BOUNDS ON THE ENVIRONMENTAL MUTUAL INFORMATION WITH NONLINEAR RELAXATION

Let us derive the pairwise distance bounds [17,21] presented in Sec. IV. We consider the case of a generic nonlinear relaxation in the slow-jumps limit,

$$U(x_\mu) = \frac{x_\mu^{2n}}{2n\tau}, \quad (\text{B1})$$

where n is a positive integer. The components of the joint distribution, Eq. (5), can be written as

$$P_i^{\text{st}}(x_1, x_2) = \frac{(2D_i n \tau)^{-1/n}}{4\Gamma^2(1 + \frac{1}{2n})} e^{-\frac{x_1^{2n} + x_2^{2n}}{2D_i n \tau}}. \quad (\text{B2})$$

Similarly, their marginalization is

$$P_i^{\text{st}}(x_\mu) = \frac{n}{\Gamma(\frac{1}{2n})(2D_i n \tau)^{1/2n}} e^{-\frac{x_\mu^{2n}}{2D_i n \tau}}. \quad (\text{B3})$$

We are interested in the mutual information

$$I_{\text{env}} = \int dx_1, dx_2 p_{\text{slow}}(x_1, x_2) \log \frac{p_{\text{slow}}(x_1, x_2)}{p_{\text{slow}}(x_1)p_{\text{slow}}(x_2)},$$

where $p_{\text{slow}}(x_1, x_2) = \sum_i \pi_i P_i^{\text{st}}(x_1, x_2)$ and $p_{\text{slow}}(x_\mu) = \sum_i \pi_i P_i^{\text{st}}(x_\mu)$, for $i \in \{-, +\}$, are mixture distributions.

In order to derive the bounds proposed in Ref. [21], we compute the 1/2-Chernoff divergence between the components of the mixture distribution,

$$C_{1/2}(P_-^{\text{st}}, P_+^{\text{st}}) = - \log \int d\mathbf{x} \sqrt{P_-^{\text{st}}(\mathbf{x})P_+^{\text{st}}(\mathbf{x})}, \quad (\text{B4})$$

and the Kullback-Leibler divergence,

$$D_{\text{KL}}(P_-^{\text{st}} || P_+^{\text{st}}) = \int d\mathbf{x} P_-^{\text{st}}(\mathbf{x}) \log \frac{P_-^{\text{st}}(\mathbf{x})}{P_+^{\text{st}}(\mathbf{x})}. \quad (\text{B5})$$

These divergences can be easily calculated, and they only depend on the ratio $\epsilon = D_-/D_+$. In particular, the 1/2-Chernoff divergences between the components of the joint and of the marginal distribution, Eqs. (B2) and (B3), are given by

$$C_{1/2}^{\text{joint}}(\epsilon) = -\frac{1}{n} \log \left[\frac{2\sqrt{\epsilon}}{1+\epsilon} \right], \quad (\text{B6})$$

$$C_{1/2}^{\text{marg}}(\epsilon) = \frac{1}{2} C_{1/2}^{\text{joint}}(P_-^{\text{st}}, P_+^{\text{st}}). \quad (\text{B7})$$

Similarly,

$$D_{\text{KL}}^{\text{joint}}(\epsilon) = \frac{\epsilon - 1 - \log \epsilon}{n}, \quad (\text{B8})$$

$$D_{\text{KL}}^{\text{marg}}(\epsilon) = \frac{1}{2} D_{\text{KL}}^{\text{joint}}(P_-^{\text{st}} || P_+^{\text{st}}), \quad (\text{B9})$$

are the Kullback-Leibler divergences between the $-$ components and the $+$ components.

Then, the lower bound reads

$$I_{\text{env}}^{\text{low}} = -\pi_+ \log \frac{[\pi_+ + e^{-C_{1/2}^{\text{marg}}(\epsilon)} \pi_-]^2}{\pi_+ + e^{-D_{\text{KL}}^{\text{joint}}(\epsilon)} \pi_-} - \pi_- \log \frac{[e^{-C_{1/2}^{\text{marg}}(1/\epsilon)} \pi_+ + \pi_-]^2}{e^{-D_{\text{KL}}^{\text{joint}}(1/\epsilon)} \pi_+ + \pi_-} \quad (\text{B10})$$

and the upper bound is

$$I_{\text{env}}^{\text{up}} = -\pi_+ \log \frac{[\pi_+ + e^{-D_{\text{KL}}^{\text{marg}}(\epsilon)} \pi_-]^2}{\pi_+ + e^{-C_{1/2}^{\text{joint}}(\epsilon)} \pi_-} - \pi_- \log \frac{[e^{-D_{\text{KL}}^{\text{marg}}(1/\epsilon)} \pi_+ + \pi_-]^2}{e^{-C_{1/2}^{\text{joint}}(1/\epsilon)} \pi_+ + \pi_-}. \quad (\text{B11})$$

Inserting $n = 2$ into the above expression gives the results presented in the main text.

In particular, notice that both $C_{1/2}^{\text{joint}}(\epsilon)$ and $D_{\text{KL}}^{\text{joint}}(\epsilon)$ diverge as $-\log \epsilon$ as $\epsilon \rightarrow 0$. Therefore, the upper and lower bounds on I_{env} converge to H_{jumps} in this limit, implying that $I_{\text{env}} \rightarrow H_{\text{jumps}}$.

[1] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression, *Proc. Natl. Acad. Sci. USA* **99**, 12795 (2002).

[2] C. Zhu and G. Yin, On competitive lotkavolterra model in random environments, *J. Math. Anal. Appl.* **357**, 154 (2009).

- [3] A. Hilfinger and J. Paulsson, Separating intrinsic from extrinsic fluctuations in dynamic biological systems, *Proc. Natl. Acad. Sci. USA* **108**, 12167 (2011).
- [4] C. G. Bowsher and P. S. Swain, Identifying sources of variation and the flow of information in biochemical networks, *Proc. Natl. Acad. Sci. USA* **109**, E1320 (2012).
- [5] L. S. Tsimring, Noise in biology, *Rep. Prog. Phys.* **77**, 026601 (2014).
- [6] P. Thomas, N. Popović, and R. Grima, Phenotypic switching in gene regulatory networks, *Proc. Natl. Acad. Sci. USA* **111**, 6994 (2014).
- [7] A. V. Dass, T. Georgelin, F. Westall, F. Foucher, P. De Los Rios, D. M. Busiello, S. Liang, and F. Piazza, Equilibrium and non-equilibrium furanose selection in the ribose isomerisation network, *Nat. Commun.* **12**, 1 (2021).
- [8] B. Mariani, G. Nicoletti, M. Bisio, M. Maschietto, S. Vassanelli, and S. Suweis, Disentangling the critical signatures of neural activity, *Sci. Rep.* **12**, 10770 (2022).
- [9] E. Schneidman, M. J. Berry II, R. Segev, and W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, *Nature* **440**, 1007 (2006).
- [10] T. Mora, A. M. Walczak, W. Bialek, and C. G. Callan Jr., Maximum entropy models for antibody diversity, *Proc. Natl. Acad. Sci. USA* **107**, 5405 (2010).
- [11] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, Statistical mechanics for natural flocks of birds, *Proc. Natl. Acad. Sci. USA* **109**, 4786 (2012).
- [12] A. A. Hyman, C. A. Weber, and F. Jülicher, Liquid-liquid phase separation in biology, *Annu. Rev. Cell Dev. Biol.* **30**, 39 (2014).
- [13] N. Dorsaz, C. De Michele, F. Piazza, P. De Los Rios, and G. Foffi, Diffusion-Limited Reactions in Crowded Environments, *Phys. Rev. Lett.* **105**, 120601 (2010).
- [14] M. Galanti, D. Fanelli, A. Maritan, and F. Piazza, Diffusion of tagged particles in a crowded medium, *Europhys. Lett.* **107**, 20006 (2014).
- [15] A. V. Chechkin, F. Seno, R. Metzler, and I. M. Sokolov, Brownian Yet Non-Gaussian Diffusion: From Superstatistics to Subordination of Diffusing Diffusivities, *Phys. Rev. X* **7**, 021002 (2017).
- [16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications and Signal Processing (Wiley-Interscience, New York, 2006).
- [17] G. Nicoletti and D. M. Busiello, Mutual Information Disentangles Interactions from Changing Environments, *Phys. Rev. Lett.* **127**, 228301 (2021).
- [18] G. Nicoletti, A. Maritan, and D. M. Busiello, Information-driven transitions in projections of underdamped dynamics, *Phys. Rev. E* **106**, 014118 (2022).
- [19] W. Wang, F. Seno, I. M. Sokolov, A. V. Chechkin, and R. Metzler, Unexpected crossovers in correlated random-diffusivity processes, *New J. Phys.* **22**, 083041 (2020).
- [20] D. M. Busiello, D. Gupta, and A. Maritan, Coarse-grained entropy production with multiple reservoirs: Unraveling the role of time scales and detailed balance in biology-inspired systems, *Phys. Rev. Res.* **2**, 043257 (2020).
- [21] A. Kolchinsky and B. D. Tracey, Estimating mixture entropy with pairwise distances, *Entropy* **19**, 361 (2017).
- [22] D. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, UK, 2021).
- [23] R. M. Neal, MCMC using Hamiltonian dynamics, *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. Jones, X. L. Meng (Chapman & Hall CRC Press, London, 2011).
- [24] M. Betancourt, A conceptual introduction to hamiltonian monte carlo, [arXiv:1701.02434](https://arxiv.org/abs/1701.02434).
- [25] A. Barducci and P. De Los Rios, Non-equilibrium conformational dynamics in the function of molecular chaperones, *Curr. Opin. Struct. Biol.* **30**, 161 (2015).
- [26] R. Gaspari, A. Gliozzi, and R. Ferrando, Aggregation phenomena in a system of molecules with two internal states, *Phys. Rev. E* **76**, 041604 (2007).
- [27] D. M. Busiello, S. Liang, F. Piazza, and P. De Los Rios, Dissipation-driven selection of states in non-equilibrium chemical networks, *Commun. Chem.* **4**, 16 (2021).
- [28] S. Liang, D. M. Busiello, and P. De Los Rios, The intrinsic non-equilibrium nature of thermophoresis, [arXiv:2102.03197](https://arxiv.org/abs/2102.03197).
- [29] C. A. Weber, D. Zwicker, F. Jülicher, and C. F. Lee, Physics of active emulsions, *Rep. Prog. Phys.* **82**, 064601 (2019).
- [30] A. Kraskov, H. Stögbauer, and P. Grassberger, Estimating mutual information, *Phys. Rev. E* **69**, 066138 (2004).
- [31] C. M. Holmes and I. Nemenman, Estimation of mutual information for real-valued data with error bars and controlled bias, *Phys. Rev. E* **100**, 022404 (2019).
- [32] M. D. McDonnell and L. M. Ward, The benefits of noise in neural systems: Bridging theory and experiment, *Nat. Rev. Neurosci.* **12**, 415 (2011).
- [33] S. Temereanca, E. N. Brown, and D. J. Simons, Rapid changes in thalamic firing synchrony during repetitive whisker stimulation, *J. Neurosci.* **28**, 11153 (2008).
- [34] B. Mariani, G. Nicoletti, M. Bisio, M. Maschietto, R. Oboe, A. Leparulo, S. Suweis, and S. Vassanelli, Neuronal avalanches across the rat somatosensory barrel cortex and the effect of single whisker stimulation, *Front. Syst. Neurosci.* **15**, 89 (2021).
- [35] K. Wienand, E. Frey, and M. Mobilia, Evolution of a Fluctuating Population in a Randomly Switching Environment, *Phys. Rev. Lett.* **119**, 158301 (2017).
- [36] E. Kussell and S. Leibler, Phenotypic diversity, population growth, and information in fluctuating environments, *Science* **309**, 2075 (2005).
- [37] P. Visco, R. J. Allen, S. N. Majumdar, and M. R. Evans, Switching and growth for microbial populations in catastrophic responsive environments, *Biophys. J.* **98**, 1099 (2010).
- [38] G. Nicoletti, S. Suweis, and A. Maritan, Scaling and criticality in a phenomenological renormalization group, *Phys. Rev. Res.* **2**, 023144 (2020).
- [39] M. C. Morrell, A. J. Sederberg, and I. Nemenman, Latent dynamical Variables Produce Signatures of Spatiotemporal Criticality in Large Biological Systems, *Phys. Rev. Lett.* **126**, 118302 (2021).
- [40] H. Kim and A. Mnih, Disentangling by factorising, in Proceedings of the 35th International Conference on Machine Learning 80, 2649 (2018).

- [41] R. T. Q. Chen, X. Li, R. B. Grosse, and D. K. Duvenaud, Isolating sources of disentanglement in variational autoencoders, *Advances in Neural Information Processing Systems* **31** (2018), <https://papers.nips.cc/paper/2018/hash/1ee3dfcd8a0645a25a35977997223d22-Abstract.html>.
- [42] F. Locatello, S. Bauer, M. Lucic, G. Raetsch, S. Gelly, B. Schölkopf, and O. Bachem, Challenging common assumptions in the unsupervised learning of disentangled representations, in *Proceedings of the 36th International Conference on Machine Learning* 97, 4114 (2019).