


## Generalized Glauber Dynamics for Inference in Biology

Xiaowen Chen<sup>1</sup>,<sup>✉</sup> Maciej Winiarski,<sup>2</sup> Alicja Puścian,<sup>2</sup> Ewelina Knapska,<sup>2</sup>  
Aleksandra M. Walczak,<sup>1,\*</sup><sup>‡</sup> and Thierry Mora<sup>1,†</sup><sup>‡</sup>

<sup>1</sup>Laboratoire de Physique de l'École normale supérieure, ENS, Université PSL, CNRS,  
Sorbonne Université, Université Paris Cité, F-75005 Paris, France

<sup>2</sup>Nencki-EMBL Partnership for Neural Plasticity and Brain Disorders-BRAINCITY,  
Nencki Institute of Experimental Biology of Polish Academy of Sciences,  
Pasteur 3 Street, 02-093 Warsaw, Poland

 (Received 21 September 2022; revised 15 August 2023; accepted 17 November 2023; published 19 December 2023)

Large interacting systems in biology often exhibit emergent dynamics, such as coexistence of multiple timescales, manifested by fat tails in the distribution of waiting times. While existing tools in statistical inference, such as maximum entropy models, reproduce the empirical steady-state distributions, it remains challenging to learn dynamical models. We present a novel inference method, called generalized Glauber dynamics. Constructed through a non-Markovian fluctuation dissipation theorem, generalized Glauber dynamics tunes the dynamics of an interacting system, while keeping the steady-state distribution fixed. We motivate the need for the method on real data from Eco-HAB, an automated habitat for testing behavior in groups of mice under seminaturalistic conditions, and present it on simple Ising spin systems. We show its applicability for experimental data by inferring dynamical models of social interactions in a group of mice that reproduce both its collective behavior and the long tails observed in individual dynamics.

DOI: [10.1103/PhysRevX.13.041053](https://doi.org/10.1103/PhysRevX.13.041053)

Subject Areas: Biological Physics, Complex Systems,  
Statistical Physics

### I. INTRODUCTION

From collective information encoding in neurons [1–4] to emergence dynamics in collective animal motion [5–8] and population dynamics in ecological communities [9], collective behavior emerges from dynamical interaction among individual components. In recent years, large scale data acquisition in precisely controlled experiments [10–14] allow researchers to address these questions by constructing statistical and dynamical data-based models that reproduce the correlated activity in spiking and nonspiking neurons [15–18] and collective animal motion [19,20].

Living systems are intrinsically dynamic and out of equilibrium, as manifested by the coexistence of multiple timescales and the breaking of Markovian rule in animal behavior [21–24] and neuron activities [25,26]. However, most recent work focuses on inferring the static properties of collective behavior: analyzing the joint probability

distributions of the interacting components and relating these global states to functional behavior. In many cases, inferred pairwise interaction models successfully reproduce the correlation structure of the data [27–29], leading to identifying the empirical rules for collective neuronal encoding [2,15,30–32], the interaction structure of bird flocks [19,33], or contact maps of proteins [34]. While these approaches do not directly address the dynamics, attempts have been made to reconcile them with dynamical models of neurons [18,35] and flocks [36] by using classical rules of equilibrium dynamics.

Recent methods to learn the dynamics of interacting systems focus on extensions to second order dynamics for continuous systems [37–40]. For discrete spiking neurons, dynamical inference focuses on reproducing pairwise correlation functions between different time points [41–44], or inferring transition probabilities or causal dependencies [45]. The extension of maximum entropy to reproduce time-delayed cross-correlations (called maximum caliber [46]) is computationally expensive and requires a lot of data to train [41,42]. However, many possible dynamical models can generate the same steady-state distribution. More fundamentally, in general the dynamics cannot be automatically related to the steady-state distribution, especially if the learned models are out of equilibrium. In transition models, the transition rate of a component is given by the history of itself and other inputs, stimuli, and interacting components in the network. By incorporating autoregressively generated noise

\*Corresponding author: [aleksandra.walczak@phys.ens.fr](mailto:aleksandra.walczak@phys.ens.fr)

†Corresponding author: [thierry.mora@phys.ens.fr](mailto:thierry.mora@phys.ens.fr)

‡These authors contributed equally to this work.

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

in the transition rate, recent developments in generalized master equation (GME) models have introduced the capacity to encode latent variables [47]. The generalized linear model (GLM), a representative of this class used in neuroscience, successfully reproduces different types of neuronal dynamics and firing patterns in many brain regions [45]. However, since GLMs are constructed by definition to predict only the next time point forward, they often generate unstable trajectories and produce inconsistent steady-state distributions with respect to the training data [48–51]. To our knowledge, no existing model tractably infers the collective dynamics while also reproducing the observed steady-state distribution.

We propose an inference method, called generalized Glauber dynamics (GGD), that combines the power of steady-state inference with dynamical inference. Constructed through a non-Markovian fluctuation dissipation theorem, the generalized Glauber dynamics tunes the dynamics of an interacting system, while keeping the steady-state equilibrium distribution fixed. In practice, this method allows for the inference to be separated into two parts: first, inference of the steady-state distribution using maximum entropy models, and then, tuning the dynamics to match the data. The basic idea behind the GGD is similar

to generalized Langevin dynamics: coupled degrees of freedom are integrated out to generate an effective memory kernel, such that the dynamics of the system depends on its history. Interestingly, the functional form of the GGD is similar to the GLM but differs dramatically in its link to the steady-state distribution. We demonstrate the power of GGD to predict the colocalization pattern of groups of socially interacting mice.

## II. RESULTS

### A. Collective behavior of social mice

We studied the interaction structure of groups of animals in a controlled environment. We analyzed data generated from the Eco-HAB experiment [11] and presented in Ref. [52], where a group of  $N = 15$  freely moving male mice live in an artificial ecological environment resembling natural burrows [Fig. 1(a)]. The Eco-HAB consists of four chambers, two of them with food, connected by tunnels. The experiment lasts for 10 days, with alternating light conditions of darkness and brightness, each lasting 12 hours, to simulate the day-night cycle. Mice are able to behave and interact freely, without any experimental constraints or manipulation. Each mouse is equipped with a unique radio

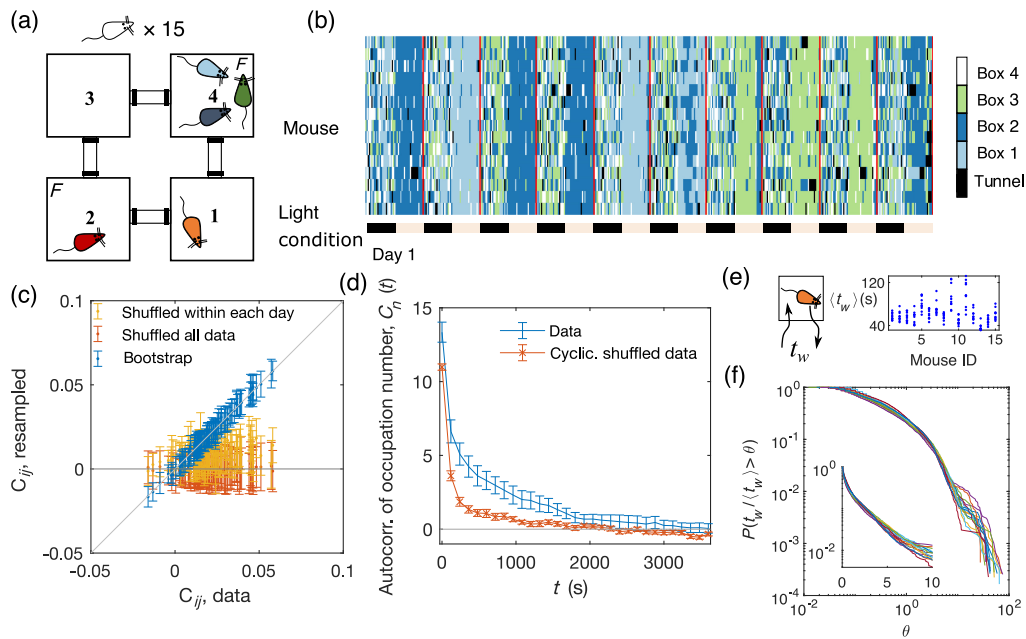


FIG. 1. Collective dynamics among social mice. (a) The Eco-HAB experimental setup (top view) with C57BL6/J male mice ( $N = 15$ ) placed in four interconnected chambers. The two chambers with food are labeled by letter  $F$ . The location of each mouse is recorded using mouse-embedded RFID and antennas at the edge of the tubes (indicated by black bars). (b) Example trace for the colocalization patterns of a group of 15 mice over 10 full days, consisting of alternating dark and light cycles (12 h). The red vertical lines indicate the beginning of each dark cycle. (c) Pairwise connected correlation function of mice colocalization  $C_{ij}$  (with error bar computed by bootstrap). Cyclically shuffled data show no correlation, while data shuffled within the same day show a strongly reduced correlation. (d) Autocorrelation function for the number of mice in a given box  $C_n(t)$  as a function of time difference, computed from the mice position between 13:00 and 19:00 each day, a period of the intensified activity chosen for the presented analysis. Error bars are the standard error from the mean across 10 days. (e) Mean waiting time for each mouse, defined to be the time a mouse spent staying in a given box before exits. Each dot indicates a different day. (f) Distribution of waiting times normalized by their mean for each mice. Distributions collapse across all mice, and decay slower than exponentially, indicating the existence of long timescales.

frequency identification transmitting chip (RFID) that allow for the detection of their position every time they pass by the 8 recording antennas [marked in black in Fig. 1(a)]. The data consist of time points with 1-ms resolution at which each mouse passed a given recording device, which allows us to identify the location of each mouse as a function of time,  $\sigma_i(t) = 1, 2, 3, \text{ or } 4$  [Fig. 1(b)].

Mice are nocturnal animals with increased activity during dark periods. For the purpose of method development we focus on dark periods and restrict our analysis to a 6 h period of stable activity, which consists of the first and bigger of the two nocturnal activity peaks characterizing the used strain of mice, C57BL6/J [see Supplemental Material (SM) [53], Fig. S1(a)]. The individual dynamics are characterized by a basal activity rate of moving between boxes, and by the tendency to explore the next box rather than come back to the previous one, which we term “roaming” [see SM, Fig. S1(b)]. Dominant mice tend to chase others more frequently [52], resulting in chaser-chased dynamics that are significant within a short time-scale of a few seconds [see SM, Fig. S1(c)].

To measure collective behavior at the group scale, we examine the excess frequency of finding two mice in the same box. The mean frequency of mouse  $i$  in box  $\alpha$  reads  $\langle \delta_{\sigma_i, \alpha} \rangle$ , where  $\delta_{a,b} = 1$  if  $a = b$ , and 0 otherwise. The excess pairwise frequency is given by the correlated pairwise correlation  $C_{ij} = \langle \delta_{\sigma_i, \sigma_j} \rangle - \sum_{\alpha} \langle \delta_{\sigma_i, \alpha} \rangle \langle \delta_{\sigma_j, \alpha} \rangle$ , where the first average is over all box combinations [Fig. 1(c)]. Nonzero correlations between mice strongly decrease when the data are shuffled across time within the same day, and completely go away when the data are shuffled across days. This suggests that interactions between mice drive the correlation, rather than the environment (i.e., the day).

To study the dynamics of that collective behavior, one can look at the temporal autocorrelation of the total number of mice in each box  $n_{\alpha}(t)$ , defined as  $C_n(t) = \sum_{\alpha} \langle n_{\alpha}(0)n_{\alpha}(t) \rangle - \langle n_{\alpha} \rangle^2$ . The excess of this autocorrelation over its counterpart in the shuffled dataset [Fig. 1(d)] is a signature of group behavior, which slows down the overall dynamics of occupancy by keeping individuals in heavily occupied boxes longer. This is confirmed by the long tails in the distribution of the waiting time, i.e., the duration between transitioning events [Figs. 1(e) and 1(f)].

## B. Modeling the steady-state distribution

These experimental observations suggest collective effects driven by direct interactions between mice that lead to effective phenomena scanning a broad range of time-scales. Our goal is to find a set of effective equations that describe the evolution of the system and are consistent with the properties of the observables in Fig. 1. Predicting the full dynamical collective behavior requires defining both the static distribution of box occupancies and the type of dynamics that governs the transitioning of the mice between boxes. We separate the inference problem into two

steps: First, we infer the steady-state distribution  $P_s(\sigma)$  for the macrostates  $\sigma = (\sigma_1, \dots, \sigma_N)$  ( $N = 15$ ), using a maximum entropy approach, and then we infer the dynamics while keeping the steady-state distribution fixed.

The maximum entropy approach has been applied in a wide range of biological contexts [2,15,16,19,34,54]. It generates approximations to the steady-state distribution that match the expectation values of a chosen set of observables while keeping the model otherwise as random as possible. Here we constrain the colocalization probabilities of all pairs of mice,  $C_{ij}$ , as well as the single-mice box occupancy functions  $\langle \delta_{\sigma_i, \alpha} \rangle$ . The maximum entropy distribution then takes the form of a Boltzmann’s law:

$$P_s(\sigma) = \frac{e^{-E(\sigma)}}{Z_s}, \quad E(\sigma) = -\sum_{i,\alpha} \left( h_{i,\alpha} - \sum_{j \neq i} J_{ij} \delta_{\sigma_j, \alpha} \right) \delta_{\sigma_i, \alpha}, \quad (1)$$

where  $h_{i,\alpha}$  and  $J_{ij}$  are Lagrange multipliers that must be tuned to satisfy the constraints,  $E(\sigma)$  is interpreted as an energy by analogy with statistical mechanics, and  $Z_s$  enforces normalization. We fit this model to the Eco-HAB and show that it correctly predicts collective statistics of occupancy that were not fit in the model, such as triad correlations and the probability of pairs of mice to be in a specific box (see SM, Fig. S2 [53]).

## C. Glauber dynamics fails to capture the longtime behavior

The same steady-state distribution, Eq. (1), can be generated by many different dynamical models. The simplest assumption inspired by statistical physics is to assume that the transition rate of a mouse from one box  $\alpha$  to an adjacent one  $\beta$  (assuming that transition is instantaneous so that two mice never transition at the same time) is a function of the difference in energies between the ending and starting states,  $\Delta E_{i,\alpha \rightarrow \beta}(\sigma)$ . Writing the transition rate between adjacent boxes as  $W_{i,\alpha \rightarrow \beta} = \mu_i f(\Delta E_{i,\alpha \rightarrow \beta})$ , where  $\mu_i$  is the overall activity of mouse  $i$  and  $f(\Delta E)$  is a function, a sufficient and necessary condition on  $f$  for these rates to admit Eq. (1) as steady state is given by detailed balance:  $f(\Delta E)/f(-\Delta E) = e^{-\Delta E}$ .

We tested some classical forms for  $f(\Delta E)$  on the data [Fig. 2(a)]. We found that the empirical normalized transition rates  $W_{i,\alpha \rightarrow \beta}/\mu_i$  are well reproduced by the form of the Glauber dynamics,  $f(\Delta E) = 1/(1 + e^{\Delta E})$ , but not by the Metropolis-Hasting prescription,  $f(\Delta E) = \min(1, e^{-\Delta E})$ . However, even the Glauber dynamics do not reproduce the long tails in the waiting time distribution, nor does a nonparametric form of transition dynamics  $f(\Delta E)$  directly estimated from the data [Fig. 2(b); see also SM, Fig. S3 [53]]. It did not have to be the case: While these dynamics are Markovian and memoryless at the group level, long

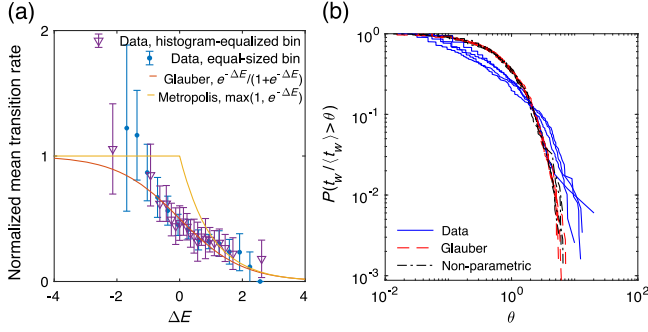


FIG. 2. Glauber dynamics based on the inferred steady-state distribution, Eq. (1), fails to predict the longtime dynamics of mice. (a) The normalized mean transition rate is well reproduced by Glauber dynamics with the parameters of the steady-state model, but not by Metropolis-Hasting dynamics. (b) However, Glauber dynamics with the parameters of the steady-state model do not reproduce the long tails of the waiting time distribution. The dynamics are also not reproduced by a nonparametric estimation of the transition rates (see SM, Fig. S3 [53]).

timescales may nonetheless emerge from interactions, as, for example, during critical slowing-down. The failure of the model suggests that, by themselves, these concurrent and colocalized pairwise interactions are not strong enough for such long timescales to emerge. Additionally, the transition probability conditioned on the elapsed time after the last transition exhibits tails (SM Fig. S4). Together, these results suggest that the dynamics may have a Glauber form, but that additional memory effects must be incorporated. Here, the term “memory” describes how the dynamics depends on the past location time series.

#### D. Generalized Glauber dynamics

Our goal is to add long-term memory effects to the equilibrium dynamics described above, while keeping maximum entropy distribution valid. For concreteness, we start from Glauber dynamics, and call the method the generalized Glauber dynamics, although the approach can be generalized to other equilibrium dynamics.

Our approach is general and applies to any group of  $N$  correlated, categorical variables taking  $q$  possible values,  $\sigma_i = 1, \dots, q$  ( $q = 4$  in the Eco-HAB). For simplicity of exposition, here we outline the derivation for a single binary (Ising) spin ( $N = 1$ ,  $q = 2$ ),  $\sigma = \pm 1$  in spin convention. Relating to the Eco-HAB, this is equivalent to a single mouse placed in an experiment apparatus with two connected chambers (denoted  $-1$  and  $+1$ ). The general derivation for arbitrary  $N$  and  $q$  is given in the SM [53].

The maximum entropy distribution is given by the energy,  $E(\sigma) = -h\sigma$ , and the Glauber dynamics is defined by the transition rates:  $W(\sigma \rightarrow -\sigma) = \mu e^{-h\sigma} / (e^h + e^{-h})$ . To include memory, we take inspiration from multidimensional Markov systems with equilibrium dynamics, such as hidden Markov models and generalized Langevin

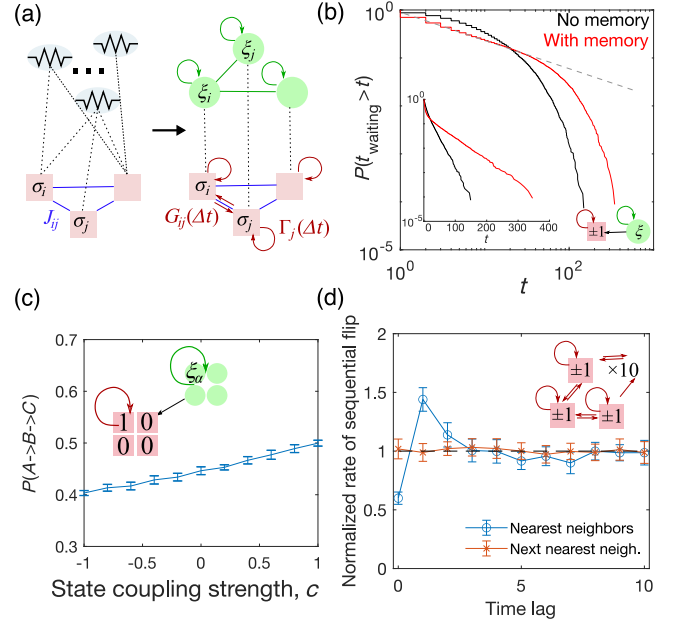


FIG. 3. Toy models with generalized Glauber dynamics. (a) Schematics of the GGD. In addition to classical couplings  $J_{ij}$ , by considering coupling to an oscillator bath, the observed degrees of freedom are coupled to their own memory [through kernel  $\Gamma_i(t)$ ], to that of their neighbors [through kernels  $G_{ij}(t)$ ], as well as to noise sources  $\xi_i(t)$  correlated across time and variables. Panel (b) shows for a single Ising spin that adding an exponentially decaying memory kernel can create long tails in the waiting distribution. (c) A single four-state variable with memory can create a bias in the tendency to continue transitioning in the same direction as in the previous transition, versus going back to the previous state. (d) In a multiple spin system, the dynamics can depend on the history of other spins, illustrated by 10 Ising spins arranged in a loop.

equations. The idea is to consider a larger equilibrium system coupling both the observed spins and some hidden degrees of freedom. While this augmented system is Markovian, the subsystem formed by the observed spins may exhibit memory.

In practice, we couple the spins to a heat bath of harmonic oscillators [see Fig. 3(a) for schematics, Ref. [55] for the standard derivation for continuous variables, and SM for a detailed derivation for categorical variables [53]]. After integrating out the hidden degrees of freedom, the transition rates of the GGD take a Glauber-like form,  $W(\sigma \rightarrow -\sigma) = \mu e^{-h_{\text{eff}}\sigma} / (e^{h_{\text{eff}}} + e^{-h_{\text{eff}}})$ , but with an effective, time-dependent field,

$$h_{\text{eff}}(t) = h + \Gamma(0)\sigma(t) - \int_0^t dt' \Gamma(t-t') \frac{d\sigma(t')}{dt'} + \xi(t), \quad (2)$$

where the noise correlation satisfies the generalized fluctuation-dissipation relation:

$$\langle \xi(t)\xi(t') \rangle = \Gamma(t-t'). \quad (3)$$

$\Gamma(t)$  is an arbitrary function that specifies how the spectrum of oscillators couples to the spin (see SM, Sec. 1B). The second and the third terms depend on the memory kernel, and is defined to be  $h_{\text{mem}}(t)$ .

The first term of Eq. (2) is the local field learned from the maximum entropy model, already present in classical Glauber dynamics. This term generalizes readily to the case of multiple interacting spins as the local field  $h_i(\sigma)$  acting on spin  $i$  (defined as half the energy difference between the configurations with  $\sigma_i = -1$  and  $\sigma_i = +1$ , the other spins being fixed). The second and the third terms depend on the history of the spin  $\sigma(t)$ , and add memory to the dynamics. The last term is the colored noise that results from the coupling with the memory kernel. When extending to  $q$  states, the fields  $h$  and  $h_{\text{eff}}$  become vectors of length  $q$ , and  $\Gamma(t)$  takes the form of a  $q \times q$  matrix coupling the different states together. Memory kernels  $G_{ij}(t)$  may also be added to couple different degree of freedom  $i$  to the memory of  $j$  (see SM, Sec. 1C [53]). By construction, the process is reversible, as a subsystem of a larger equilibrium system including both spins and the oscillator bath, and its steady state is still given by Boltzmann's law, Eq. (1). The choice for the memory kernel  $\Gamma(t)$  is general and can be chosen from a large family of functions.

### E. Range of possible dynamics of GGD

We now illustrate the range of possible dynamics generated by GGD, by simulating simple toy models of Ising or Potts spins (see Sec. IV D for details on the simulations). We start by asking whether our illustrative example of a single spin variable with memory can generate non-Markovian tails of the waiting time distribution, as observed in the mice experiments [Fig. 2(b)]. We define a GGD for a single spin with an exponentially decaying memory kernel,  $\Gamma(t) = A \exp(-t/\tau)$ , where  $\tau$  is the timescale of the self-memory, and compare to the classical Glauber dynamics ( $A = 0$ ). By construction, both dynamics predict the same steady-state distribution, characterized by  $\langle \sigma \rangle = \tanh(h)$  (see SM, Fig. S5 [53]). However, the GGD predicts a long tail in the waiting time distribution, whereas the naive Glauber dynamics yields an exponential distribution of waiting times [Fig. 3(b)]. Thus, even a simple form of the memory kernel can create long memory effects similar to those observed in data.

Second, we illustrate the model's ability to account for non-Markovian flow between states. In the Eco-HAB data, different mice have different levels of roaming—the ratio of probabilities of moving forward versus moving backward in two consecutive transitions. This effect is non-Markovian and arises from memory effects. In the GGD, this memory can be encoded in nondiagonal elements of the  $4 \times 4$  matrix  $\Gamma(t)$ . We define a GGD on a four-state variable through a kernel  $\Gamma(t)$  with cyclic symmetry and exponentially decaying terms (see detailed parametrization in SM, Sec. 1D [53]). The model can generate a bias

between transitioning to the next state ( $A \rightarrow B \rightarrow C$ ) rather than coming back to the previous one ( $A \rightarrow B \rightarrow A$ ), and this tendency is tuned by the strength  $c$  of the off-diagonal elements of the memory kernel with a maximum value of  $P(A \rightarrow B \rightarrow C) = 0.5$  [Fig. 3(c); see also SM, Sec. 1D].

Mice tend to chase each other in the Eco-HAB experiment [SM, Fig. S1(c)], suggesting that transitions also depend on the history of other mice's behavior. This can be encoded in the GGD through cross-individual memory kernels  $G_{ij}(t)$ . This memory coupling enforces the flipping rate of a degree of freedom (in more general terms, called the follower) to depend on the recent transition of another degree of freedom (called the leader), such that the transition rate of the leader-follower pairs has a distinguished characteristic timescale that is not visible in other pairs [Fig. 3(d); see also SM, Sec. 1C, for the memory kernel]. The symmetry of the memory kernel enforces the memory dependence between a given pair to be symmetric.

### F. Inference

How do we fit the GGD to data? Assuming that a maximum entropy distribution has already been learned, we need to solve the inverse problem of finding the memory kernel  $\Gamma$  that reproduces the experimentally observed dynamics. We assume an exponential form for the kernel,  $\Gamma_i(t) = A e^{-t/\tau}$ , which allows for rewriting  $\xi_i(t)$  as an Ornstein-Uhlenbeck process. We maximize the likelihood of the discretized data series  $\sigma(t)$  (with some small time bin) over the three parameters  $\theta = (\mu, A, \tau)$ , using the expectation-maximization (EM) algorithm [56] to deal with the hidden variables  $\xi_i(t)$  (see SM [53]). Specifically, we adopt the EM algorithm used to infer neural firing dynamics with hidden noise [47,57–59]. The key difference is that for our inference problem of the GGD, the non-Markovian fluctuation-dissipation relation [Eq. (3)] acts as an extra constraint between the parameters generating the noise  $\xi(t)$  and the memory kernel  $\Gamma_i(t)$ , while studies applying EM algorithm to infer neural firing dynamics do not assume the fluctuation-dissipation relation. In addition, the transition dynamics are of the Glauber form (i.e., logistic function, instead of exponential as in models of neural spiking dynamics), which does not lead to simple mathematical expressions and requires Monte Carlo sampling in the computation (see SM).

Figure 4(a) shows an example of a single Ising spin undergoing GGD with a single exponentially decaying memory kernel. To mimic the situation of a spin within a large interacting system subject to a changing external field  $h_i(t)$ , we consider a time-dependent field  $h(t)$  with sinusoidal form. With our EM algorithm, we are able to recover the parameters with high accuracy (see SM, Fig. S6 and Table S1 [53]), as well as estimate the hidden noise [see Fig. 4(b)]. Trajectories simulated with the inferred set of parameters reproduce the properties of the waiting time distribution [Fig. 4(c)] and the autocorrelation function

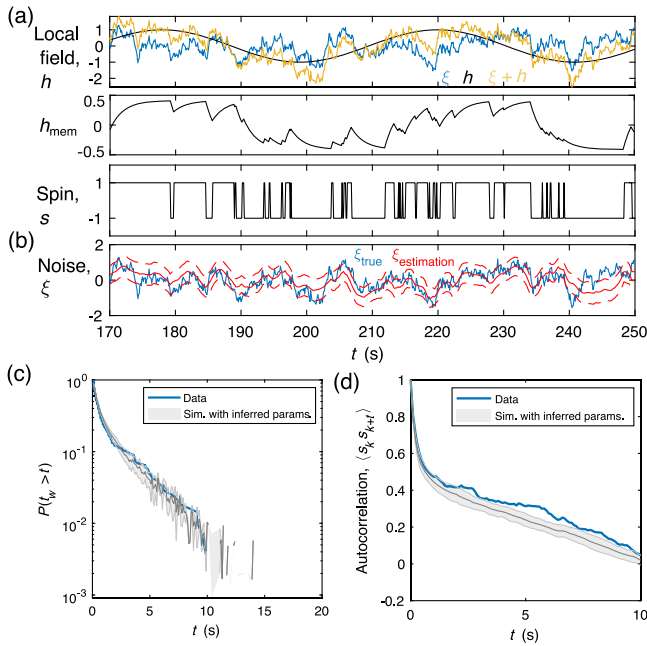


FIG. 4. Example of GGD inference with an expectation-maximization (EM) algorithm on a single Ising spin under an oscillating field  $h(t) = \sin(0.15t)$ . The memory kernel is  $\Gamma(t) = A \exp(-t/\tau)$ , with true parameters  $A = 0.8$ ,  $\tau = 19.5$  s, and the baseline transition rate  $\mu = 4$  s $^{-1}$ . The simulation was conducted for a duration of  $T = 300$  s, using a time step of  $\Delta t = 0.1$  s. (a) The combination of  $h(t)$ , the noise  $\xi$ , and the effective local field due to history of spin flips,  $h_{\text{mem}}$  [the second and third terms of Eq. (2)] generates the sample spin trajectory  $s_t$ . (b) Given the spin trajectory  $\sigma(t)$  and  $h(t)$ , the EM algorithm recovers an ensemble of possible realizations of the hidden noise  $\xi(t)$ . With parameters inferred using the EM algorithm, the simulated trajectories recover the waiting time distribution (c) and the autocorrelation decay (d) as the data. The envelope of the curve is the standard deviation.

[Fig. 4(d)] observed in the data. The error of the EM algorithm, measured as the percentage difference between the EM inferred parameters and the ground truth parameters, scales with data length  $T$  as  $\sim T^{-1/2}$ . This scaling is expected from the Cramer-Rao bound [SM, Fig. S8 and Fig. S10(a), top right], and therefore the same as one expects from maximum caliber methods and generalized linear models.

We can speed up the inference by heuristically minimizing the distance between the empirical and simulated distributions of dynamical variables, which is defined as the sum of the area between the empirical and model-simulated waiting time cumulative distributions in double logarithmic scale. Although the error in the inferred parameters is larger than when using EM (SM, Table S1), the waiting time distribution and the autocorrelation function are recovered correctly (SM, Fig. S7).

One can extend the parametrization of memory kernel to sums of exponential decays,  $\Gamma(t) = \sum_l A_l e^{-t/\tau_l}$ , to

approximate more general forms of memory kernels which decays at infinite time (Prony’s series; also see Ref. [47]). The extension of the EM algorithm is straightforward, as the noise can be written as a linear sum of Ornstein-Uhlenbeck processes. For the Eco-HAB mice data, we only used the heuristic method, because the EM algorithm becomes unreliable and hard to converge for categorical data with more possible states (see SM, Fig. S10 and Sec. 3D [53]), due to a distortion of the optimization landscape which leads to problems in convergence, which is consistent with the literature (see Chap. 3.4 in Ref. [60]).

## G. GGD of social mice

We now go back to our original problem of 15 mice living in an Eco-HAB and ask if we can distinguish properties of individual animals from emergent behavior resulting from interactions. Atop the static maximum entropy model we learned previously, we learn the GGD model to fit the waiting time distribution. Since the three non-Markovian features (self-memory, individual-specific inertia, and chaser-chased dynamics) occur on different timescales, in principle we should construct a memory kernel whose diagonal and off-diagonal terms have very different timescales. To simplify the task, in addition to the activity prefactors  $\mu_i$ , we only learn the self-memory kernels  $\Gamma_i(t) = A_i e^{-t/\tau_i}$ , as this memory occurs on the longest timescale and contributes the most to the observed fat tail in the waiting time distribution. Recall from Fig. 1(f) that for all mice the waiting time distribution collapses after we divide by its mean, so we can further reduce the number of parameters by assuming  $A_i = A_0$ , and  $\mu_i \tau_i = \text{const}$ . We learn this reduced set of dynamical parameters by minimizing the total distance between the observed and predicted waiting time distribution for all mice, computed independently for each mouse while fixing the trajectories of other mice to their experimental values. The optimized parameters are found to be  $A_i = 2.75$ ,  $\mu_i = 0.25 \pm 0.08$  s $^{-1}$ ,  $\tau_i = 22 \pm 9$  s.

We then simulate the dynamics for all 15 mice, using the static parameters learned by pairwise maximum entropy model ( $h_{i,\alpha}, J_{ij}$ ) and the dynamical parameters learned by GGD ( $\mu_i, A_i, \tau_i$ ). Simulations correctly capture the tails of the waiting time distribution [Figs. 5(a) and 5(b)]. By construction, the GGD model reproduces the static observables [SM, Fig. S11(a) [53]], which the GLM model fails to reproduce [SM, Fig. S11(b)]. Since the memory kernel consists of single exponential decays, it suggests a biologically plausible mechanism for its encoding by mice using an iterative leaky integrator of their internal state, without the need to remember their entire past behavior.

As discussed previously, while the GGD for Potts spins can tune the probability of consecutive forward transitions within a certain range, the “roaming” effect exhibited by the Eco-HAB mice is more pronounced than the GGD allows for in its current form. Specifically, while GGD for a single

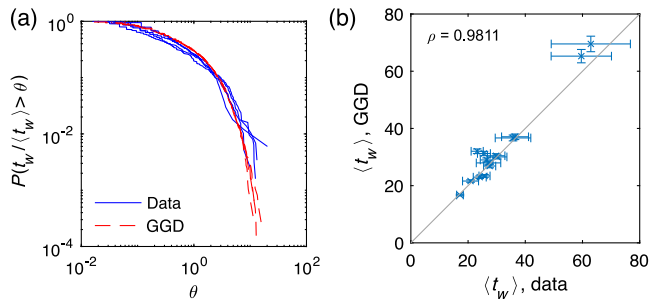


FIG. 5. The generalized Glauber dynamics on top of the static maximum entropy model on Eco-HAB data reproduces the long tail of both the shape (a) and the mean (b) of the waiting time distributions. The Pearson correlation coefficient  $\rho$  is given in the plot.

Potts spin allows a maximum value of  $P(A \rightarrow B \rightarrow C) = 0.5$ , in the Eco-HAB mice data,  $P(A \rightarrow B \rightarrow C)_{\text{mice}}$  is almost always greater than 0.5.

### III. DISCUSSION

Existing approaches for inferring collective behavior focus on reproducing either the steady-state distribution or the dynamics of the system. This study is motivated by the need for a single model that describes both types of observables while still being tractable and interpretable. Inspired by data from the Eco-HAB experiment monitoring the spontaneous collective behavior of mice, we developed the generalized Glauber dynamics (GGD) approach, which infers the dynamics in a way that the steady-state distribution is guaranteed to be reproduced. GGD simplifies dynamical inference by separating the learning of the steady-state distribution and the learning of the dynamics. Steady-state inference is performed with the well-developed method of maximum entropy models. Then, a family of reversible dynamics is constructed by adding both a memory kernel and a colored noise, which are related through a non-Markovian fluctuation-dissipation theorem. In addition to providing detailed derivations of the GGD, we presented, as a proof of principle, examples of dynamics using toy models of Ising and Potts spins, and described two protocols of inference.

Compared to existing approaches, GGD makes an explicit and analytically tractable link between the dynamical and steady-state properties of interacting systems. It combines the notion of a probability landscape and dynamics, allowing us to describe non-Markovian dynamical transitions on a well-defined energy landscape. Practically, GGD has several advantages compared to existing methods of dynamical inference—maximum caliber [46] and generalized linear models [45]. In contrast to maximum caliber methods, GGD is defined by an explicit form of the transition matrix, making it easier to simulate and interpret. Maximum caliber methods also require us to learn many more parameters. For a system of size  $N$ , for each time

delay of  $\Delta t$ , maximum caliber needs to simultaneously fit to  $N(N-1)/2$  equal time correlations, and the  $N(N-1)$  cross-time correlations. A total time lag of  $L$  will lead to a total of  $N(N-1)/2 + LN(N-1)$  parameters to be fitted simultaneously. In comparison, in GGD, we first fit  $N(N-1)/2$  equal time correlations in the maximum entropy learning step, then for each of the  $N$  components, we fit the dynamics to a parametrized memory kernel with a chosen number of parameters (three in the case we considered). By separating the learning procedure into first learning the static distribution, and then independently the dynamics for each component, we have much fewer parameters to learn. In contrast to generalized linear models, it is guaranteed to agree with the empirical steady-state distribution, and is immune to problems of blow-up divergences that plague inferred state-transition models (see Refs. [48–51]; see also SM, Fig. S11). Thus, GGD introduces memory while guaranteeing the steady state.

GGD is a special case of the generalized master equation, as the transition rate is a function of the current state and past states through a memory kernel, and latent variables can be encoded in autoregressively generated noise in the transition rate. Compared to other GME models, the key ingredient of the GGD is the non-Markovian fluctuation-dissipation relation between the correlation function of the colored noise and the memory kernel, which is essential in restoring the same steady-state distribution for a wide range of possible dynamics. The choice of the memory kernels can be very general and is only constrained by the reversibility constraint imposed by the fluctuation-dissipation relation. The GGD can capture rich dynamics, such as the effect of individual memory, inertia on discrete dynamics, and dynamics which depends on other individuals in the system. While we have focused on continuous-time Glauber dynamics for concreteness, the approach can be extended to any continuous or discrete dynamics (e.g., Metropolis-Hastings). Possible extensions and future applications of GGD include inferring dynamical models for interacting spiking neurons.

By applying GGD to the colocalization pattern of groups of socially interacting mice, we show how GGD can be used on real biological data, where both dynamics and steady-state collective behavior are of interest, and by which it exposes the success and limitations of the current GGD method. For the Eco-HAB mice data, GGD reproduces the collective dynamics, as recapitulated by the waiting time distribution, which occurs on timescales of minutes. The transition times are well predicted both on average and at the distribution level. We also take note of GGD limitations. Its dynamics are reversible by construction, precluding out-of-equilibrium effects. On the Eco-HAB data, a GLM model that does not enforce detailed balance shows asymmetry in the inferred interaction matrix [SM, Fig. S12(d) [53]], suggesting that the Eco-HAB mice are indeed out of equilibrium. Nonetheless, we find a strong correlation between the GLM interaction matrix and the

GGD interaction matrix, with correlation coefficient of 0.70 for the asymmetric GLM interaction matrix, and 0.68 for the symmetrized GLM interaction matrix [SM, Figs. S12(b) and S12(c)], and the local fields  $h$  also correlate with a coefficient of 0.69 [SM, Fig. S12(a)]. Despite these nonequilibrium effects, GGD reproduces the relevant dynamical observables well, and only fails at capturing some behavior at very short timescales (around few seconds, such as active chasing between pairs of mice) compared to the total time of the experiments (around hours).

Because the memory kernel  $\Gamma(t)$  is also an autocorrelation function, it cannot take arbitrary forms. For example, an abrupt suppression to simulate refractory period in neurons would not be possible. In its current form, the constraint on the memory kernel imposes a limit on the maximum roaming effect possible, which is less than what the Eco-HAB mice exhibit. On the technical side, we did not manage to reliably infer the effect of memory between individuals  $G_{ij}(t)$  from the data. This may be due to the large number of parameters to consider, which scales with the number of pairs. Another difficulty is that the memory of histories of other individuals may happen on a shorter timescale than self-memory (as suggested by the chaser-chased dynamics [52]), confusing the inference procedure. Interestingly, the autocorrelation of the box occupation number decays more slowly than predicted by the model, suggesting that these effect may play an important role (SM, Fig. S13). In addition, because the GGD integrates all memory from the past, it is unable to describe a full memory reset. Finally, unlike the maximum entropy model, GGD is not a minimal construction that one can use to build dynamics with increasing complexity. One possible solution is to consider GGD with memory kernels built using a complete basis of functions. Such an extension is likely to be very useful, since it could capture phenomena on different timescales, which we saw is relevant for behavior in the Eco-HAB.

Living systems are often strongly out of equilibrium, while GGD assumes microscopic reversibility by construction. How widely applicable is GGD to biological system beyond the case of social mice? Previous work has demonstrated the usefulness of equilibriumlike approaches for extracting effective interactions in strongly nonequilibrium systems. For example, the dynamics that govern the evolution of protein families are strongly out of equilibrium, yet simple equilibrium models have shown great power at predicting contacts between residues [34]. In the collective dynamics of bird flocks, it was shown that while the overall dynamics is out of equilibrium, the local dynamics when one looks at shorter timescales appears to be well described by equilibrium dynamics [36]. In flocking models, nonequilibrium effects were shown to be dominant only around the ordering transition, but vanish deep in the ordered and disordered phases [61]. In our social mice example, nonequilibrium effects do exist, but

seem to be restricted to short-time chasing behavior. We thus expect GGD to be of use in a variety of nonequilibrium contexts where the relevant observables are effectively in equilibrium. One hallmark of GGD is the inclusion of memory effects. Such memory is expected to emerge in partially observed systems, where the marginalization over latent variables generates long-term memory, even in equilibrium systems. The existence of such latent variables is ubiquitous in biology, where only part of the system (molecules in a biochemical network, neurons in the brain) can be simultaneously measured. GGD provides a framework to describe such interacting systems with implicit memory.

In summary, GGD provides an essential step to bridging the steady-state distribution with the collective dynamics of living systems. Our model does not solve the complete inference problem for all dynamical systems but fills a niche of interpretable dynamical models that link to the steady-state landscape. As a proof of principle that it is possible to place the dynamics of models of interacting systems on a probability landscape, GGD opens the door to connecting static and dynamic inference of non-Markovian interacting systems.

## IV. MATERIALS AND METHODS

### A. Eco-HAB mice data

The Eco-HAB system and the appropriate data analysis tools are described in Ref. [11]. The particular experiment and data used for analysis and methods development in this manuscript are published in Ref. [52].

### B. Compute normalized chasing rate and rate of sequential flip

In a lead-and-follow pair, e.g., both the chaser-chased pairs in the Eco-HAB system and the sequential spin flips in multiple interacting Ising spins, we compute the normalized following rate. For a given pair of mice (spins)  $(i, j)$  and a fixed time period, we count the number of consecutive transitions where mouse  $j$  leads mouse  $i$  to leave the same box and to enter the same box, separated by a time difference of  $\Delta t$ . We then divide this count by an expected null value, computed by cyclically shuffling the time series of all mice, to obtain the normalized lead-and-follow rate.

### C. Learning the static maximum entropy model

The static maximum entropy model is learned by gradient descent, where at each optimization step, the parameters  $J_{ij}$  and  $h_{i,\alpha}$  are updated by the difference between the empirical observable and Monte Carlo sampled observables at the current estimation for each parameter [62]. The initial condition is for  $h_{i,\alpha}$  to be that of the independent model and all  $J_{ij}$  set to zero. The stop condition is when the square difference between the data



and the simulation is less than the data variation, computed by bootstrapping the data. Monte Carlo sampling of the model and computing the mean and correlation are performed using the UGM MATLAB package [63], while all other steps are performed by customized MATLAB codes.

#### D. Simulate generalized Glauber dynamics

To simulate generalized Glauber dynamics with a given memory kernel  $\Gamma(t)$ , we first generate noise  $\xi(t)$  whose correlation is  $\langle \xi(t)\xi(t') \rangle = \Gamma(t-t')$  using methods of Fourier transforms [64]. For systems with higher dimensions, the noise is first independently generated in the eigenbasis of the memory kernel, then transformed back to the standard basis. Then, the dynamics is simulated by discretizing time and using parallel updating. The time step is chosen to be small to make sure at most one spin transitions at any given time step.

#### E. Inference of dynamical parameters

The inference of the dynamical parameters is performed with two methods. The first is an expectation-maximization algorithm developed in detail in SM [53], and implemented with customized MATLAB code. In the single Ising spin example, we chose the stopping criterion such that the absolute change in all three parameters ( $\mu, a, \sigma_\epsilon^2$ ) must be less than a threshold value of 0.01 over the last 100 iterations of the EM algorithm.

Alternatively, heuristic optimizations use two consecutive grid searches followed by a Nelder-Mead algorithm provided by the built-in MATLAB function `patternsearch` to find the optimum.

#### F. Learning the generalized linear model

For comparison with the generalized Glauber dynamics, we train a generalized linear model on the Eco-HAB data. The model follows [45] and writes the transition probability in  $\Delta t$  as

$$P_{i,\alpha \rightarrow \beta} = \frac{\tilde{W}_{i,\alpha \rightarrow \beta} \Delta t}{Z_{i,\alpha}},$$

$$P_{i,\alpha \rightarrow \alpha} = \frac{1}{Z_{i,\alpha}},$$

where the transition rate is

$$\tilde{W}_{i,\alpha \rightarrow \beta} = \tilde{\mu}_i \exp \left[ \tilde{h}_{i\beta} - \tilde{h}_{i\alpha} - \sum_j \tilde{J}_{ij} \delta_{\sigma_i \sigma_j} + \int_0^t dt' \tilde{\Gamma}(t') \sigma_i(t') \right]$$

for  $\alpha \neq \beta$  with a normalization factor

$$Z_{i,\alpha} = 1 + \sum_{\beta \neq \alpha, \beta \text{ accessible}} \tilde{W}_{i,\alpha \rightarrow \beta} \Delta t.$$

For direct comparison with the GGD model, the memory kernel is chosen to be parametrized by single exponential,  $\tilde{\Gamma}_i(t) = \tilde{A}_i e^{-t/\tilde{\tau}_i}$ . The parameters  $\tilde{\mu}_i, \tilde{h}_{i\alpha}, \tilde{J}_{ij}, \tilde{A}_i, \tilde{\tau}_i$  are estimated using maximum likelihood.

#### ACKNOWLEDGMENTS

This work was partially supported by the European Research Council Consolidator Grant No. 724208, “BRAINCITY—Centre of Excellence for Neural Plasticity and Brain Disorders” project of the Polish Foundation for Science, and the National Science Center Grant No. 2020/39/D/NZ4/01785. This work was also supported by the Bettencourt Schueller Foundation. The authors are grateful for the discussions and suggestions from Giulio Biroli.

- [1] C. Lee, W. H. Rohrer, and D. L. Sparks, *Population coding of saccadic eye movements by neurons in the superior colliculus*, *Nature (London)* **332**, 357 (1988).
- [2] G. Tkačik *et al.*, *Searching for collective behavior in a large network of sensory neurons*, *PLoS Comput. Biol.* **10**, e1003408 (2014).
- [3] B. B. Averbeck, P. E. Latham, and A. Pouget, *Neural correlations, population coding and computation*, *Nat. Rev. Neurosci.* **7**, 358 (2006).
- [4] A. Puścian, H. Benisty, and M. J. Higley, *NMDAR-dependent emergence of behavioral representation in primary visual cortex*, *Cell Rep.* **32**, 107970 (2020).
- [5] T. Vicsek and A. Zafeiris, *Collective motion*, *Phys. Rep.* **517**, 71 (2012).
- [6] M. Ballerini *et al.*, *Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1232 (2008).
- [7] I. D. Couzin, *Collective cognition in animal groups*, *Trends Cognit. Sci.* **13**, 36 (2009).
- [8] D. H. Kelley and N. T. Ouellette, *Emergent dynamics of laboratory insect swarms*, *Sci. Rep.* **3**, 1 (2013).
- [9] A. Sanchez and J. Gore, *Feedback between population and evolutionary dynamics determines the fate of social microbial populations*, *PLoS Biol.* **11**, e1001547 (2013).
- [10] M. Ballerini *et al.*, *Empirical investigation of starling flocks: A benchmark study in collective animal behaviour*, *Anim. Behav.* **76**, 201 (2008).
- [11] A. Puścian, S. Łęski, G. Kaspruwicz, M. Winiarski, J. Borowska, T. Nikolaev, P. M. Boguszewski, H.-P. Lipp, and E. Knapska, *Eco-HAB as a fully automated and ecologically relevant assessment of social impairments in mouse models of autism*, *eLife* **5**, e19532 (2016).
- [12] R. Sarfati, J. C. Hayes, É Sarfati É, and O. Peleg, *Spatio-temporal reconstruction of emergent flash synchronization in firefly swarms via stereoscopic 360-degree cameras*, *J. R. Soc. Interface* **17**, 20200179 (2020).
- [13] J. P. Nguyen, F. B. Shipley, A. N. Linder, G. S. Plummer, M. Liu, S. U. Setru, J. W. Shaevitz, and A. M. Leifer, *Whole-brain calcium imaging with cellular resolution in freely*

- behaving Caenorhabditis elegans*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E1074 (2016).
- [14] A. L. Juavinett, G. Bekheet, and A. K. Churchland, *Chronically implanted neuropixels probes enable high-yield recordings in freely moving mice*, *eLife* **8**, e47188 (2019).
- [15] E. Schneidman, M. J. Berry II, R. Segev, and W. Bialek, *Weak pairwise correlations imply strongly correlated network states in a neural population*, *Nature (London)* **440**, 1007 (2006).
- [16] S. Cocco, S. Leibler, and R. Monasson, *Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14058 (2009).
- [17] U. Ferrari, T. Obuchi, and T. Mora, *Random versus maximum entropy models of neural population activity*, *Phys. Rev. E* **95**, 042321 (2017).
- [18] X. Chen, F. Randi, A. M. Leifer, and W. Bialek, *Searching for collective behavior in a small brain*, *Phys. Rev. E* **99**, 052418 (2019).
- [19] 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. U.S.A.* **109**, 4786 (2012).
- [20] Y. Shemesh, Y. Sztainberg, O. Forkosh, T. Shlapobersky, A. Chen, and E. Schneidman, *High-order social interactions in groups of mice*, *eLife* **2**, e00759 (2013).
- [21] B. J. Cole, *Fractal time in animal behaviour: The movement activity of drosophila*, *Anim. Behav.* **50**, 1317 (1995).
- [22] G. J. Stephens, M. Bueno de Mesquita, W. S. Ryu, and W. Bialek, *Emergence of long timescales and stereotyped behaviors in *Caenorhabditis elegans**, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7286 (2011).
- [23] V. Alba, G. J. Berman, W. Bialek, and J. W. Shaevitz, *Exploring a strongly non-Markovian animal behavior*, *arXiv:2012.15681*.
- [24] G. J. Berman, *Measuring behavior across scales*, *BMC Biol.* **16**, 23 (2018).
- [25] J. C. Marques, M. Li, D. Schaak, D. N. Robson, and J. M. Li, *Internal state dynamics shape brainwide activity and foraging behaviour*, *Nature (London)* **577**, 239 (2020).
- [26] H. S. Kaplan, O. S. Thula, N. Khoss, and M. Zimmer, *Nested neuronal dynamics orchestrate a behavioral hierarchy across timescales*, *Neuron* **105**, 562 (2020).
- [27] W. Bialek and R. Ranganathan, *Rediscovering the power of pairwise interactions*, *arXiv:0712.4397*.
- [28] Y. Roudi, S. Nirenberg, and P. E. Latham, *Pairwise maximum entropy models for studying large biological systems: When they can work and when they can't*, *PLoS Comput. Biol.* **5**, e1000380 (2009).
- [29] R. R. Stein, D. S. Marks, and C. Sander, *Inferring pairwise interactions from biological data using maximum-entropy probability models*, *PLoS Comput. Biol.* **11**, e1004182 (2015).
- [30] L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, and W. Bialek, *Collective behavior of place and non-place neurons in the hippocampal network*, *Neuron* **96**, 1178 (2017).
- [31] A. Tang *et al.*, *A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro*, *J. Neurosci.* **28**, 505 (2008).
- [32] I. E. Ohiorhenuan, F. Mechler, K. P. Purpura, A. M. Schmid, Q. Hu, and J. D. Victor, *Sparse coding and high-order correlations in fine-scale cortical networks*, *Nature (London)* **466**, 617 (2010).
- [33] A. Attanasi *et al.*, *Information transfer and behavioural inertia in starling flocks*, *Nat. Phys.* **10**, 691 (2014).
- [34] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, *Identification of direct residue contacts in protein-protein interaction by message passing*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 67 (2009).
- [35] S. Wolf, G. Le Goc, G. Debrégeas, S. Cocco, and R. Monasson, *Emergence of time persistence in a data-driven neural network model*, *eLife* **12**, e79541 (2023).
- [36] T. Mora, A. M. Walczak, L. D. Castello, F. Ginelli, S. Melillo, L. Parisi, M. Viale, A. Cavagna, and I. Giardina, *Local equilibrium in bird flocks*, *Nat. Phys.* **12**, 1153 (2016).
- [37] A. Frishman and P. Ronceray, *Learning force fields from stochastic trajectories*, *Phys. Rev. X* **10**, 021009 (2020).
- [38] F. Ferretti, V. Chardès, T. Mora, A. M. Walczak, and I. Giardina, *Building general Langevin models from discrete datasets*, *Phys. Rev. X* **10**, 031018 (2020).
- [39] A. C. Costa, T. Ahamed, D. Jordan, and G. Stephens, *Maximally predictive states: From partial observations to long timescales*, *Chaos* **33**, 023136 (2023).
- [40] E. H. Thiede, D. Giannakis, A. R. Dinner, and J. Weare, *Galerkin approximation of dynamical quantities using trajectory data*, *J. Chem. Phys.* **150**, 244111 (2019).
- [41] O. Marre, S. El Boustani, Y. Frégnac, and A. Destexhe, *Prediction of spatiotemporal patterns of neural activity from pairwise correlations*, *Phys. Rev. Lett.* **102**, 138101 (2020).
- [42] J. C. Vasquez, O. Marre, A. G. Palacios, M. J. Berry II, and B. Cessac, *Gibbs distribution analysis of temporal correlations structure in retina ganglion cells*, *J. Physiol.* **106**, 120 (2012).
- [43] A. Cavagna, I. Giardina, F. Ginelli, T. Mora, D. Piovani, R. Tavarone, and A. M. Walczak, *Dynamical maximum entropy approach to flocking*, *Phys. Rev. E* **89**, 042707 (2014).
- [44] T. Mora, S. Deny, and O. Marre, *Dynamical criticality in the collective activity of a population of retinal neurons*, *Phys. Rev. Lett.* **114**, 078105 (2015).
- [45] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli, *Spatio-temporal correlations and visual signalling in a complete neuronal population*, *Nature (London)* **454**, 995 (2008).
- [46] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, *Principles of maximum entropy and maximum caliber in statistical physics*, *Rev. Mod. Phys.* **85**, 1115 (2013).
- [47] H. Vroylandt, L. Goudenège, P. Monmarché, F. Pietrucci, and B. Rotenberg, *Likelihood-based non-Markovian models from molecular dynamics*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117586119 (2022).
- [48] I. M. Park, E. W. Archer, N. Priebe, and J. W. Pillow, *Spectral methods for neural characterization using generalized quadratic models*, *Advances in Neural Information Processing Systems 26 (NIPS 2013)* (2013), [https://proceedings.neurips.cc/paper\\_files/paper/2013/hash/a8240-cb8235e9c493a0c30607586166c-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2013/hash/a8240-cb8235e9c493a0c30607586166c-Abstract.html).

- [49] D. Hocker and I. M. Park, *Multistep inference for generalized linear spiking models curbs runaway excitation*, in *Proceedings of the 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2017 (IEEE, 2017), pp 613–616.
- [50] F. Gerhard, M. Deger, and W. Truccolo, *On the stability and dynamics of stochastic spiking neuron models: Nonlinear Hawkes process and point process GLMs*, *PLoS Comput. Biol.* **13**, e1005390 (2017).
- [51] G. Mahuas, G. Isacchini, O. Marre, U. Ferrari, and T. Mora, *A new inference approach for training shallow and deep generalized linear models of noisy interacting neurons*, *Adv. Neural Inf. Process. Syst.* **33**, 5070 (2020), <https://proceedings.neurips.cc/paper/2020/file/356dc40642abeb3a-437e7e06f178701c-Paper.pdf>.
- [52] M. Winiarski *et al.*, *Social learning about rewards—How information from others helps to adapt to changing environment*, *bioRxiv*, 10.1101/2021.03.09.434563.
- [53] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevX.13.041053> for detailed derivation of generalized Glauber dynamics for Ising and Potts spins, the expectation-maximization algorithm for inferring dynamical parameters for GGD, and Supplemental figures.
- [54] T. Mora, A. Walczak, W. Bialek, and C. Callan, *Maximum entropy models for antibody diversity*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5405 (2010).
- [55] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).
- [56] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, *J. R. Stat. Soc. Ser. B* **39**, 1 (1977).
- [57] A. C. Smith and E. N. Brown, *Estimating a state-space model from point process observations*, *Neural Comput.* **15**, 965 (2003).
- [58] J. E. Kulkarni and L. Paninski, *Common-input models for multiple neural spike-train data*, *Network Comput. Neural Syst.* **18**, 375 (2007).
- [59] K. Yuan and M. Niranjan, *Estimating a state-space model from point process observations: A note on convergence*, *Neural Comput.* **22**, 1993 (2010).
- [60] K. Yuan, *Inference and learning in state-space point process models: Algorithms and applications*, Ph.D. thesis, University of Southampton, 2013.
- [61] F. Ferretti, S. Grosse-Holz, C. Holmes, J. L. Shivers, I. Giardina, T. Mora, and A. M. Walczak, *Signatures of irreversibility in microscopic models of flocking*, *Phys. Rev. E* **106**, 034608 (2022).
- [62] G. Tkacik, E. Schneidman, M. J. Berry II, and W. Bialek, *Ising models for networks of real neurons*, *arXiv:q-bio/0611072*.
- [63] M. Schmidt, *UGM: Matlab code for undirected graphical models*, <https://www.cs.ubc.ca/~schmidtm/Software/UGM.html>.
- [64] J. Tüeckmantel, *Digital generation of noise-signals with arbitrary constant or time-varying spectra (a noise generation software package and its application)*, CERN, Report No. LHC-PROJECT-Report-1055, 2008.