



Non-equivalence of Dynamical Ensembles and Emergent Non-ergodicity

Hadrien Vroylandt¹ · Gatien Verley¹

Received: 29 June 2018 / Accepted: 25 October 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Dynamical ensembles have been introduced to study constrained stochastic processes. In the microcanonical ensemble, the value of a dynamical observable is constrained to a given value. In the canonical ensemble a bias is introduced in the process to move the mean value of this observable. The equivalence between the two ensembles means that calculations in one or the other ensemble lead to the same result. In this paper, we study the physical conditions associated with ensemble equivalence and the consequences of non-equivalence. For continuous time Markov jump processes, we show that ergodicity guarantees ensemble equivalence. For non-ergodic systems or systems with emergent ergodicity breaking, we adapt a method developed for equilibrium ensembles to compute asymptotic probabilities while caring about the initial condition. We illustrate our results on the infinite range Ising model by characterizing the fluctuations of magnetization and activity. We discuss the emergence of non-ergodicity by showing that the initial condition can only be forgotten after a time that scales exponentially with the number of spins.

Keywords Large deviations theory · Equivalence of ensembles · Non-ergodicity · Ising model

1 Introduction

Ensemble equivalence offers a convenient way of computing equilibrium potentials (entropy, free energy, grand potential, etc.) by choosing the ensemble in which calculations are easier. Then, a Legendre transformation provides the appropriated potential according to the environmental constraints on the system (isolated, thermostated, chemostated, etc.) [10]. Ensemble equivalence holds in many cases, but it breaks down for systems with long range interactions between system constituents [3,7,11] or in the presence of a phase transition

✉ Hadrien Vroylandt
hadrien.vroylandt@th.u-psud.fr

Gatien Verley
gatien.verley@th.u-psud.fr

¹ Laboratoire de Physique Théorique (UMR8627), CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay, France

treated at the mean field level [15,19]. The equivalence of equilibrium ensembles has been studied in detail [3,6,7,11,54], including in the framework of large deviation theory [50,52].

More recently, different dynamical ensembles have been used by many authors to study stochastic processes [9,13,14,23,31,36,40,47,55] and the question of their equivalence has been an active field of research. Equilibrium and dynamical ensembles differ qualitatively in nature: the first one is made of states and the second of succession of states (either continuous or discrete) called trajectories. Then, the probabilities on the ensembles are also different, a state probability for the former and a probability of trajectories for the latter. Dynamical ensembles must be used when the considered variables are time-integrated observables (like work, irreversible heat exchanges or activity) because their statistics can only be computed using trajectory probabilities. Based on a choice of dynamical observable and of a stochastic process defining bare trajectory probabilities, one can define two dynamical ensembles: the microcanonical and the canonical ensembles. Within the first ensemble, the trajectories are filtrated on the value of the chosen dynamical observable. Within the second ensemble, the probability of trajectories are exponentially biased. This canonical ensemble is also called the *s-ensemble* [25], the *driven, biased* or *tilted* ensemble [31] or *Esscher transform* [20]. The last ensemble has been used to study glass transition [5,12,24,25,29,32,33,39,48] or as a numerical tool to evaluate rare events probabilities [21,43]. As indicated by their names, the way one introduces dynamical ensembles is in clear analogy with the way ensembles are defined in equilibrium statistical physics.

When two ensembles are equivalent, the mean value of the dynamical observable is the same in the two ensembles. Like for equilibrium ensembles, it is more convenient to make calculations (or simulations) within the canonical ensemble motivating the determination of the conditions for ensemble equivalence. Large deviation theory allows to conclude on this point from the convexity of the large deviation function (LDF) or from the differentiability of the cumulant generating function (CGF) [13,14]. Whenever ensembles are equivalent, LDF and CGF encode the same information and are related by Legendre transformation.

In this paper, we aim in the first place at exploring the physical constraints associated with equivalence of microcanonical and canonical dynamical ensembles. In the second place, when ensemble equivalence does not hold, we provide a method to compute non-convex LDFs from the (moment) generating functions.

We identify that ergodicity and additivity are crucial for the ensemble equivalence. Ergodicity guaranties that the system can switch from any state to another in a reasonable time enabling to concatenate pieces of trajectories. Thanks to additivity, the probability of the dynamical observable on a trajectory made of two pieces is connected to a weighted average of the pieces probabilities (based on their duration) constraining the dynamical observable probability to ensure ensemble equivalence. In Sect. 2, we review results on ensemble equivalence and show for Markov jump processes that ergodicity implies the equivalence of dynamical ensembles for any additive observable. The contraposition says that non-ergodicity is required (but may not be sufficient) to break ensemble equivalence, and thus to obtain non-convex LDF. In Sect. 3, we illustrate our results with a simple non-ergodic system and derive the non-convex LDF for the system activity. For this system, non-ergodicity is imposed by construction because the transition rate matrix is reducible. Non-ergodicity impacts the LDF since the latter may change according to the probability of the initial state. Based on our understanding of this simple model, we look at the magnetization and activity of an infinite range Ising model for which ergodicity is broken in the thermodynamic limit. We show how to compute the non-convex LDF from the propagator of the generating function for the chosen observable(s) by adapting a method developed by Touchette in Ref. [51,53] for equilibrium ensembles. Applying this method to the Ising model, we notice that the non equivalence of

equilibrium ensembles is tightly connected to the non equivalence of dynamical ensembles. Since non-ergodicity of the Ising model is an emergent feature, the order of the thermodynamic and long time limits determines which of the convex or non-convex LDF is more appropriated. In Sect. 4, we prove that the time to relax from the initial condition grows exponentially with the number of spins. In practice, this means that for times shorter than this relaxation time the system is not ergodic and one must use the non-convex LDF. For longer times of observation, the system is ergodic and one must use the convex LDF. Such arguments are conventional in the framework of equilibrium ensembles [6,41]: we adapt them for dynamical ensembles and illustrate them precisely using the Ising model.

2 Microcanonical and Canonical Dynamical Ensembles

In this section, we review for time-homogeneous jump processes some results on dynamical ensemble equivalence, see Refs. [13,14] and references therein for more general Markov processes.

2.1 Definitions: Stochastic Processes and Dynamical Ensembles

The Markov jump process X_t takes values in the finite state space Ω at all time $t \in [0, T]$. The probability that $X_t = x \in \Omega$ at time t , denoted $p(x, t)$, is the solution of the time-homogeneous master equation

$$\frac{\partial}{\partial t} p(x, t) = \sum_y K(x, y) p(y, t) \tag{1}$$

where K is the Markov operator with off-diagonal components $K(x, y)$ giving the probability per unit time of the transition from y to x and diagonal components equal to minus the total escape rates from each state. The Markov jump process is said to be irreducible if for any pair of states $x \neq y$ it exists a path that connects x to y (and y to x), i.e. every transitions along the path have non-vanishing transition rates. On the opposite, the process is reducible when at least two subsets of states in Ω cannot be connected. Then, the process is non-ergodic. As a consequence, the ensemble average does not produce the same result as the time average since the process cannot explore the whole state space.

Alternatively, one may use the path-integral formalism to characterize the process X_t . In this case, we denote by $P[x]$ the path probability, where $[x]$ is the short notation for a realization of the process X_t over the time interval $[0, T]$.

We aim at computing the statistics of a dynamical observable \mathcal{O} . The mean value of this observable \mathcal{O} follows from

$$\langle \mathcal{O} \rangle = \int \mathcal{D}[x] P[x] \mathcal{O}[x], \tag{2}$$

where $\int \mathcal{D}[x]$ denotes the sum over all possible paths. For simplicity, we consider an observable of the form

$$\mathcal{O} \equiv \frac{1}{T} \int_0^T f(X_t) dt + \frac{1}{T} \sum_{0 \leq t \leq T: \Delta X_t \neq 0} g(X_{t+}, X_{t-}) \tag{3}$$

where $f(x)$ and $g(x, y)$ are arbitrary functions and the discrete sum is on the time t at which the system changes of state, i.e. $\Delta X_t \equiv X_{t+} - X_{t-} \neq 0$. The f dependent part enables to

study the time average of a state function, and the g dependent part the time average of jump quantities, like currents. For example, choosing $f(x) = 0$ and $g(x, y) = 1$ will lead to the global activity of the system, i.e. the number of jumps per unit time in the trajectory.

The *microcanonical ensemble* is defined by filtering the trajectory ensemble with a condition on an observable, e.g. $\mathcal{O} = \sigma$ where σ is a possible value of the stochastic variable \mathcal{O} such that its probability density function verifies $P_T(\mathcal{O} = \sigma) > 0$. We write the path probability of these trajectories

$$P^\sigma[x] = \frac{P[x, \mathcal{O} = \sigma]}{P_T(\mathcal{O} = \sigma)}, \tag{4}$$

where $P^\sigma[x]$ is the conditional probability of the path $[x]$ given that for this path $\mathcal{O} = \sigma$. The corresponding joint probability is denoted $P[x, \mathcal{O} = \sigma]$. In the microcanonical path ensemble, the observable \mathcal{O} does not fluctuate and always achieves the same value for all trajectories in the ensemble.

The *canonical ensemble* is defined by fixing the mean value of the observable \mathcal{O} only. The path probability for this ensemble can be computed via tilting (or biasing) the process as follows:

$$P_\gamma[x] = \frac{e^{T\gamma\mathcal{O}} P[x]}{\langle e^{T\gamma\mathcal{O}} \rangle}. \tag{5}$$

This path probability is normalized by construction.

2.2 Path Ensemble Equivalence

To discuss the dynamical ensemble equivalence, one needs to define when two path probabilities P and Q are asymptotically equivalent. They do when [14]

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \ln \frac{P[x]}{Q[x]} = 0 \tag{6}$$

almost everywhere with respect to P . This means that P and Q are equal up to sub-exponential terms in T for almost all paths. As a consequence, the mean value at large time of any observable will be the same if computed with any one of the two equivalent path ensembles.

Assuming a large deviation principle for \mathcal{O} [52], the LDF function $I(\sigma)$ provides the rate of decay of the probability density function $P_T(\mathcal{O} = \sigma)$ in the limit $T \rightarrow +\infty$ such that

$$P_T(\mathcal{O} = \sigma) \underset{T \rightarrow +\infty}{\simeq} e^{-TI(\sigma)} \tag{7}$$

A LDF indicates the rate at which the probability density function of a time extensive observable concentrates on its most likely value when time goes by.

Based on large deviation principle, Chetrite and Touchette have shown that the convexity of $I(\sigma)$ determines whether the microcanonical and canonical path ensembles are equivalent [14], in such a case P^σ and P_γ satisfy Eq. (6). They distinguished three cases:

Equivalence If $I(\sigma)$ is a strictly convex function at σ , then there exists a unique $\gamma \in \mathbb{R}$ such that P^σ and P_γ are equivalent.

Non-equivalence If $I(\sigma)$ is a non-convex function at σ , then there are no $\gamma \in \mathbb{R}$ such that P^σ and P_γ are equivalent.

Partial equivalence If $I(\sigma)$ is a convex function but not strictly convex function at σ , then numerous values of σ correspond to the same γ : It may correspond to linear parts in a convex function or to set of points at which the slope is the same.

The convexity of the LDF $I(\sigma)$ is connected to the differentiability of the CGF $\phi(\gamma)$ that is defined as

$$\phi(\gamma) \equiv \lim_{T \rightarrow +\infty} \frac{1}{T} \ln \langle e^{T\gamma\mathcal{O}} \rangle. \tag{8}$$

The CGF is the highest eigenvalue of the titled matrix \mathbf{K}_γ of elements [35]

$$K_\gamma(x, y) \equiv \begin{cases} K(x, y)e^{\gamma g(x, y)} & \text{if } x \neq y \\ K(x, x) + \gamma f(x) & \text{if } x = y \end{cases}. \tag{9}$$

The CGF is also the Legendre–Fenchel transformation of the LDF

$$\phi(\gamma) = \max_{\sigma} (\gamma\sigma - I(\sigma)). \tag{10}$$

Conversely, we define the cLDF $I^{**}(\sigma)$ as the Legendre–Fenchel transformation of the CGF

$$I^{**}(\sigma) \equiv \max_{\gamma} (\gamma\sigma - \phi(\gamma)) = \gamma(\sigma)\sigma - \phi(\gamma(\sigma)), \tag{11}$$

with $\gamma(\sigma)$ the value of γ realizing the maximum. The Gärtner–Ellis theorem states that if $\phi(\gamma)$ is a differentiable function then $I(\sigma) = I^{**}(\sigma)$. The properties of the Legendre–Fenchel transformation implies that $I^{**}(\sigma)$ is a convex function. The strict convexity of the LDF and the differentiability of the CGF are then dual conditions. When the CGF is not differentiable, we can compute the LDF from the CGF in case of partial equivalence. In case of non-equivalence, it is not possible since we only obtain the convex hull of the LDF. Hence, we consider convexity (strict or not) as the criterion to determine whether ensemble equivalence holds, having in mind that we shall be able to compute the LDF from the CGF when it holds. For this reason, we merge equivalence and partial equivalence together.¹

In Ref. [14], the authors conjectured a connection between non-equivalences and non-ergodicity. We support this idea in the following by confirming that non-ergodicity is *required* to observe a non-equivalence of ensemble, but it is not a sufficient condition.

2.3 Proof of Ensemble Equivalence for Ergodic Processes

The proof of ensemble equivalence amounts to determine the convexity of a LDF function. In this section, we show how the LDF for observable \mathcal{O} can be determined from the convex LDF of more general observables, and how the convexity property may be inherited in the process.

Whenever a stochastic variable depends on a set of other stochastic variables, one can determine the LDF of the former from the LDF of the latter. This operation is called a *contraction*. For instance, the observable \mathcal{O} of Eq. (3) is a function of the empirical occupation ratio \mathbf{R} and the jump rate \mathbf{C}

$$\mathcal{O} = \mathcal{O}(\mathbf{R}, \mathbf{C}) = \sum_x f(x)R(x) + \sum_{x, y} C(x, y)g(x, y). \tag{12}$$

The occupation ratio is defined as the relative time spent in each state, say x here

$$R(x) \equiv \frac{1}{T} \int_0^T \delta(X_t = x) dt, \tag{13}$$

¹ The alternative viewpoint is to consider the non-differentiability of the CGF as the criterion for determining ensemble equivalence (merging together partial equivalence and non-equivalence).

where δ is an indicator function being 1 if the condition is satisfied and 0 otherwise. The jump rate $C(x, y)$ counts the number of transitions between two states (from y to x) per unit time

$$C(x, y) \equiv \frac{1}{T} \sum_{0 \leq t \leq T: \Delta X_t \neq 0} \delta(X_{t+} = x) \delta(X_{t-} = y). \tag{14}$$

The occupation ratios and the jump rates have a joint probability for $\mathbf{R} = \mathbf{r}$ and $\mathbf{C} = \mathbf{c}$ that satisfies a large deviation principle

$$P_T(\mathbf{R} = \mathbf{r}, \mathbf{C} = \mathbf{c}) \underset{T \rightarrow +\infty}{\simeq} e^{-TI(\mathbf{r}, \mathbf{c})}, \tag{15}$$

defining the large deviation function $I(\mathbf{r}, \mathbf{c})$, also called the level 2.5 LDF [2,4,38]. Then, the contraction

$$I(\mathbf{o}) = \min_{\mathbf{r}, \mathbf{c} \text{ s.t. } \mathcal{O}(\mathbf{r}, \mathbf{c}) = \mathbf{o}} I(\mathbf{r}, \mathbf{c}) \tag{16}$$

leads to the LDF for the observable \mathcal{O} . In general terms, a contraction corresponds to the minimization of a multivariate LDF under a condition encoding how the stochastic variables are related [52].

Since ensemble equivalence for a stochastic variable relies on the convexity of the corresponding LDF, it is crucial to determine (i) whether the LDF $I(\mathbf{r}, \mathbf{c})$ is convex and (ii) whether the convexity can be inherited upon contraction. The following theorem appearing in Ref. [8] provides an answer to point (ii) when the new variable is additive:

Theorem (Contraction) *Let $h(\mathbf{x}, \mathbf{z})$ be a convex function and $U(\mathbf{z})$ an additive function, i.e. a function verifying*

$$U(\alpha \mathbf{z}_1 + (1 - \alpha) \mathbf{z}_2) = \alpha U(\mathbf{z}_1) + (1 - \alpha) U(\mathbf{z}_2), \tag{17}$$

then

$$d(\mathbf{y}) = \min_{\mathbf{x} \in \mathcal{C}, \mathbf{z} \in \mathcal{C}_{\mathbf{y}}} h(\mathbf{x}, \mathbf{z}) \text{ with } \mathcal{C} \text{ convex, and } \mathcal{C}_{\mathbf{y}} = \{\mathbf{z} \mid U(\mathbf{z}) = \mathbf{y}\} \tag{18}$$

is a convex function.

Proof We consider $(\mathbf{x}_1^*, \mathbf{z}^*(\mathbf{y}_1))$ the couple of variables realizing the minimum in Eq. (18) for \mathbf{y}_1 , and similarly $(\mathbf{x}_2^*, \mathbf{z}^*(\mathbf{y}_2))$ for \mathbf{y}_2 . The convexity of \mathcal{C} implies that $\alpha \mathbf{x}_1^* + (1 - \alpha) \mathbf{x}_2^* \in \mathcal{C}$ when $\alpha \in [0, 1]$. Moreover, the additivity of U implies that $\alpha \mathbf{z}^*(\mathbf{y}_1) + (1 - \alpha) \mathbf{z}^*(\mathbf{y}_2) \in \mathcal{C}_{\alpha \mathbf{y}_1 + (1 - \alpha) \mathbf{y}_2}$. Hence, we have

$$\begin{aligned} d(\alpha \mathbf{y}_1 + (1 - \alpha) \mathbf{y}_2) &= \min_{\mathbf{x} \in \mathcal{C}, \mathbf{z} \in \mathcal{C}_{\alpha \mathbf{y}_1 + (1 - \alpha) \mathbf{y}_2}} h(\mathbf{x}, \mathbf{z}) \\ &\leq h(\alpha \mathbf{x}_1^* + (1 - \alpha) \mathbf{x}_2^*, \alpha \mathbf{z}^*(\mathbf{y}_1) + (1 - \alpha) \mathbf{z}^*(\mathbf{y}_2)), \\ &\leq \alpha h(\mathbf{x}_1^*, \mathbf{z}^*(\mathbf{y}_1)) + (1 - \alpha) h(\mathbf{x}_2^*, \mathbf{z}^*(\mathbf{y}_2)), \\ &\leq \alpha d(\mathbf{y}_1) + (1 - \alpha) d(\mathbf{y}_2), \end{aligned}$$

where we get the third line by using the convexity of h , and the fourth line using our knowledge of the minimizers of Eq. (18) for both \mathbf{y}_1 and \mathbf{y}_2 . □

We now address point (i) about the convexity of the LDF $I(\mathbf{r}, \mathbf{c})$. This level 2.5 LDF is explicitly known for ergodic Markov jump processes in continuous-time [2,4,38]

$$I(\mathbf{r}, \mathbf{c}) = \sum_{x, y \neq x} \left(r(y) K(x, y) - c(x, y) + c(x, y) \ln \frac{c(x, y)}{r(y) K(x, y)} \right). \tag{19}$$

It is convex since it writes as the sum of a linear part $\sum_{x,y \neq x} [r(y)K(x, y) - c(x, y)]$ and a Kullback–Leibler divergence between \mathbf{c} and $\mathbf{K}\mathbf{r}$

$$D(\mathbf{c} || \mathbf{K}\mathbf{r}) \equiv \sum_{x,y \neq x} c(x, y) \ln \frac{c(x, y)}{K(x, y)r(y)}. \tag{20}$$

This Kullback–Leibler divergence is convex as a consequence of the log-sum inequality (or Jensen’s inequality) [16]. Using the previous theorem with $\mathbf{z} = (\mathbf{r}, \mathbf{c})$, the convexity of the level 2.5 LDF and the additivity of \mathcal{O} for both the occupation ratio and the jump rate, we conclude that the contracted LDF $I(\mathfrak{o})$ is also convex: The ensemble equivalence holds for ergodic Markov jump processes. In particular, the ensemble equivalence holds for irreducible finite size Markov jump processes since they are always ergodic.

Alternatively, one can prove heuristically the ensemble equivalence using the differentiability of the CGF, that is the dual condition with respect to the LDF convexity. To determine the CGF differentiability, one considers the CGF as the highest eigenvalue of the tilted matrix \mathbf{K}_γ defined in Eq.(65). From Perron–Fröbenius theorem for irreducible finite size matrices like \mathbf{K}_γ , the highest (real) eigenvalue is unique: Its multiplicity is always one independently of the value of the counting field and no crossing between the two highest eigenvalues can occur. Moreover, the components of the tilted matrix are differentiable yielding that the CGF is itself differentiable. By duality, the LDF is strictly convex.

From this analysis, we conclude that the non-equivalence between the microcanonical and canonical ensembles based on observables like \mathcal{O} may only happen when the Markov operator is reducible or of infinite dimension [17,18]. Then, the LDF of the variable \mathcal{O} may not be convex.

We emphasize that we made no assumption on the definition of dynamical rates, hence the system may be in or out of equilibrium. We used simple graph Markov processes, but the discussion can be transposed to include multiple transition channels.

3 Explicit LDF Calculation for Non-ergodic Systems

In this section, we adapt the method developed in Ref. [53] for equilibrium ensembles to compute non-convex LDFs from their corresponding generating functions. First, we explain how to compute a non-convex LDF from the propagator of the generating function for a non-ergodic system chosen for its simplicity and for its similarity with the mean field Ising model. We then apply the method to the infinite range Ising model for which non-ergodicity emerges in the thermodynamic limit.

3.1 A Four State Model with Non-convex LDF of Activity

Let’s consider a four state system made of two subsystems with two states (1, 2) and (3, 4). This four state system evolves according to the master equation of Eq. (1) with Markov operator

$$\mathbf{K} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -2 & 2 \\ 0 & 0 & 2 & -2 \end{pmatrix}, \tag{21}$$

that is reducible yielding a non-ergodic process.

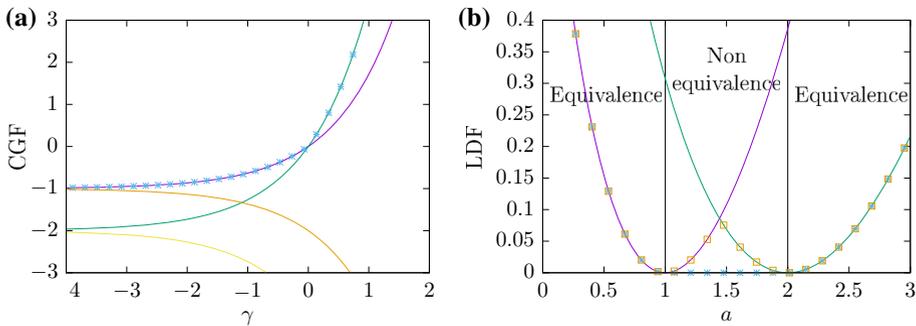


Fig. 1 **a** Four eigenvalues of the biased operator (24) (solid line) and corresponding CGF (cross). **b** Partial LDFs (solid lines) corresponding to the Legendre transformation of the two highest eigenvalues, cLDF (cross) and LDF (squares)

We aim at computing the system activity assuming that one cannot distinguish in which subsystem a transition occurs. The activity rate A is the total number of jumps per unit time: It is given by Eq. (3) when $f(x) = 0$ and $g(x, y) = 1$ for all (x, y) . By definition of the Markov operator of Eq. (21), the activity probability distribution of each subsystem is a Poisson distribution of mean value T for the subsystem (1, 2) and $2T$ for the subsystem (3, 4), because there is 1 and 2 jumps per unit time in each subsystem respectively. The probability distribution of the total number of jumps is the sum of Poisson distributions weighted by the initial probability $p_i(x)$ to start in state x . The probability of having aT jumps after a time T is

$$P_T(A = a) = [p_i(1) + p_i(2)] \frac{(T)^{aT} e^{-T}}{(aT)!} + [p_i(3) + p_i(4)] \frac{(2T)^{aT} e^{-2T}}{(aT)!} \quad (22)$$

For T large enough, $P_T(A = a)$ will be bimodal, and the LDF reads

$$I(a) = - \lim_{T \rightarrow +\infty} \frac{1}{T} \ln P_T(A = a) = \begin{cases} a \ln a - a + 1 & \text{if } a < 1/\ln 2 \\ a \ln \frac{a}{2} - a + 2 & \text{if } a \geq 1/\ln 2 \end{cases}, \quad (23)$$

where we have used Stirling formula and chosen the minimum of the two LDF corresponding to each Poisson distribution. The above LDF of activity is shown in Fig. 1b. It is crucial to note that the initial condition plays a fundamental role here: if the system never starts in states 1 or 2, i.e. $p_i(1) = p_i(2) = 0$, the LDF will include the branch corresponding to the second line of Eq. (23) only. The non-ergodicity impacts the long time statistics of the dynamical observables through the choice of initial conditions.

For the four state model, it is straightforward to determine the probability density function of activity, but for other systems or observables this task may be more challenging: one must often compute the CGF instead. Let's derive the result of Eq. (23) in this way using the propagator for the generating function of the activity defined by $G(x_f, x_i, \gamma) = \langle e^{T\gamma A} \rangle_{x_f, x_i}$, where the subscripts x_i and x_f denote respectively the initial and final states of the trajectories appearing in the average. Using standard approach [52,56], the propagator is obtained from the biased operator

$$\mathbf{K}_\gamma = \begin{pmatrix} -1 & e^\gamma & 0 & 0 \\ e^\gamma & -1 & 0 & 0 \\ 0 & 0 & -2 & 2e^\gamma \\ 0 & 0 & 2e^\gamma & -2 \end{pmatrix} \quad (24)$$

as the components of its matrix exponential $[\exp(T\mathbf{K}_\gamma)](x_f, x_i) = G(x_f, x_i, \gamma)$. Using eigenvalues and eigenvectors decomposition of \mathbf{K}_γ , we make explicit the matrix exponential

$$\begin{aligned} \exp [T\mathbf{K}_\gamma] = & e^{T(-1+e^\gamma)} \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \\ 0 \end{pmatrix} \cdot (1 \ 1 \ 0 \ 0) + e^{2T(-1+e^\gamma)} \begin{pmatrix} 0 \\ 0 \\ 1/2 \\ 1/2 \end{pmatrix} \cdot (0 \ 0 \ 1 \ 1) \\ & + e^{-T(1+e^\gamma)} \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \\ 0 \end{pmatrix} \cdot (1 \ -1 \ 0 \ 0) + e^{-2T(1+e^\gamma)} \begin{pmatrix} 0 \\ 0 \\ 1/2 \\ -1/2 \end{pmatrix} \cdot (0 \ 0 \ 1 \ -1). \end{aligned} \tag{25}$$

The orthogonal basis of eigenvectors has normalized right eigenvectors, and the scalar products of the left and right eigenvectors associated to the same eigenvalue are all equal to 1. We notice that the eigenvectors separate in two sets whose supports are disjoint and correspond to each subsystem respectively. Then, the activity LDF is recovered from this propagator by first summing over the initial and final states

$$\langle e^{T\gamma A} \rangle = \sum_{x_f, x_i} G(x_f, x_i, \gamma) p_i(x_i), \tag{26}$$

second, applying an inverse Laplace transformation, and finally by taking the limit $T \rightarrow \infty$. For ergodic systems, this procedure leads to the same LDF whatever the order of these operations. On the contrary, the order matters for non-ergodic systems.

For our 4 state model, the generating function writes

$$\langle e^{T\gamma A} \rangle = [p_i(1) + p_i(2)]e^{T(-1+e^\gamma)} + [p_i(3) + p_i(4)]e^{2T(-1+e^\gamma)}, \tag{27}$$

because the terms in $e^{-T(1+e^\gamma)}$ and $e^{-2T(1+e^\gamma)}$ disappear when summing on x_i and x_f . The inverse Laplace transform yields the probability density function of activity given in Eq. (22). However, first computing the long time limit of the generating function leads to the CGF

$$\phi(\gamma) = \lim_{T \rightarrow +\infty} \frac{1}{T} \ln \sum_{x_f, x_i} G(x_f, x_i, \gamma) p_i(x_i) = \begin{cases} e^\gamma - 1 & \text{if } \gamma < 0, \\ 2(e^\gamma - 1) & \text{if } \gamma \geq 0 \end{cases}, \tag{28}$$

as long as $p_i(x) > 0$ for all x . Noticing that the limit $T \rightarrow +\infty$ enables to use a saddle-point method to approximate the inverse Laplace transformation into a Legendre–Fenchel transformation, the asymptotic probability of activity follows from its corresponding cLDF

$$I^{**}(a) = \max_\gamma [a\gamma - \phi(\gamma)] = \begin{cases} a \ln a - a + 1 & \text{if } a < 1 \\ 0 & \text{if } a \in [1, 2] \\ a \ln \frac{a}{2} - a + 2 & \text{if } a > 2 \end{cases}, \tag{29}$$

that is not the one of Eq. (23). The above cLDF is convex because a Legendre–Fenchel transformation only yields convex functions by definition, while the LDF of Eq. (23) is not convex.

Alternatively, one may obtain the LDF by taking the long time limit on the propagator of the generating function of Eq. (25), and not on the generating function itself, yielding

$$\phi_{x_i}(\gamma) = \lim_{T \rightarrow +\infty} \frac{1}{T} \ln G(x_f, x_i, \gamma) = \begin{cases} (e^\gamma - 1) & \text{if } x_i = 1, 2 \\ 2(e^\gamma - 1) & \text{if } x_i = 3, 4 \end{cases}. \tag{30}$$

The Legendre–Fenchel conjugate of these two branches associated to different initial states are precisely the two branches of the LDF in Eq. (23):

$$I_{x_i}(a) = \max_{\gamma} [a\gamma - \phi_{x_i}(\gamma)] = \begin{cases} a \ln a - a + 1 & \text{if } x_i = 1, 2 \\ a \ln \frac{a}{2} - a + 2 & \text{if } x_i = 3, 4 \end{cases} \tag{31}$$

The summation over initial and final conditions is now carried out using an asymptotic approximation, which can be written heuristically as

$$e^{-TI(a)} \simeq \sum_{x_i} p_i(x_i) e^{-TI_{x_i}(a)} \simeq \exp\left(-T \min_{x_i} I_{x_i}(a)\right) \tag{32}$$

explaining why a minimum on the branches of Eq. (31) appears in the final LDF of Eq. (23). To summarize, from the Legendre–Fenchel conjugate of all the eigenvalues appearing in the biased operator of Eq. (24), we determined the partial LDFs. The minimum among these partial LDFs produces the LDF. We illustrate this procedure in Fig. 1. From this figure, we conclude on the ensemble equivalence for this model: it holds for $a \in [0, 1[\cup]2, +\infty[$, but the equivalence is partial at $a = 1, 2$ and there is no equivalence for $a \in]1, 2[$.

Remark that when the mean activities of each subsystem are the same, the two branches of the LDF merge toward a convex shape. Therefore, in this case non-ergodicity is not sufficient to observe ensemble non-equivalence associated to non-convexity of the LDF.

3.2 General Framework

In the above example, we have seen that one can determine the asymptotic fluctuations of a physical observable by switching from one dynamical ensemble to another as long as the LDFs are piecewise convex. In this section, we develop this approach for the more general framework of Sect. 2. In practice, this amounts to express the inverse Laplace transformation as a Legendre–Fenchel transformation, using the saddle point method and taking care of the initial condition appropriately.

By definition, for $b \in \mathbb{R}$, the LDF for \mathcal{O} writes

$$I(o) \equiv \lim_{T \rightarrow +\infty} \frac{-1}{T} \ln \int_{b-i\infty}^{b+i\infty} d\gamma e^{-T\gamma o} \sum_{x_f, x_i} G(x_f, x_i, \gamma) p_i(x_i), \tag{33}$$

in term of the initial probability $p_i(x_i)$ and of the propagator $G(x_f, x_i, \gamma) = \langle e^{T\gamma\mathcal{O}} \rangle_{x_f, x_i}$ that generates the moments of \mathcal{O} under given initial and final conditions. The argument of the logarithm in Eq. (33) is exactly the probability distribution function of \mathcal{O} . Since solely the most probable events contribute to the LDF, we can focus on the initial conditions leading to the minimal value of the LDF (as seen in Sect. 3.1):

$$I(o) = \min_{x_f, x_i} \lim_{T \rightarrow +\infty} \frac{-1}{T} \ln \int_{b-i\infty}^{b+i\infty} d\gamma e^{-T\gamma o} G(x_f, x_i, \gamma) p_i(x_i). \tag{34}$$

Finally, the complex integral for the inverse Laplace conjugate follows from the saddle point method:

$$I(o) = \min_{x_f, x_i} \max_{\gamma} \left(\gamma o - \lim_{T \rightarrow +\infty} \frac{1}{T} \ln G(x_f, x_i, \gamma) p_i(x_i) \right). \tag{35}$$

Alternatively, the convex hull of $I(\mathfrak{o})$ is the Legendre–Fenchel conjugate of the CGF for the observable \mathcal{O} , namely

$$I^{**}(\mathfrak{o}) \equiv \max_{\gamma} \left(\gamma \mathfrak{o} - \lim_{T \rightarrow +\infty} \frac{1}{T} \ln \sum_{x_f, x_i} G(x_f, x_i, \gamma) p_i(x_i) \right). \tag{36}$$

As seen in Sect. 3.1, the propagator G may have an eigendecomposition for which the terms contributing to the CGF in the limit $T \rightarrow \infty$ depend on the initial or final conditions. Hence, a maximum on x_f and x_i must appear

$$I^{**}(\mathfrak{o}) = \max_{\gamma} \left(\gamma \mathfrak{o} - \max_{x_f, x_i} \lim_{T \rightarrow +\infty} \frac{1}{T} \ln G(x_f, x_i, \gamma) p_i(x_i) \right), \tag{37}$$

to select the dominant asymptotic behavior in the limit $T \rightarrow \infty$. In view of comparing with Eq. (35), the maximization is modified into a minimization through the commutation with the minus sign:

$$I^{**}(\mathfrak{o}) = \max_{\gamma} \min_{x_f, x_i} \left(\gamma \mathfrak{o} - \lim_{T \rightarrow +\infty} \frac{1}{T} \ln G(x_f, x_i, \gamma) p_i(x_i) \right). \tag{38}$$

In the end, the difference between the LDF $I(\mathfrak{o})$ and its convex hull $I^{**}(\mathfrak{o})$ comes from the non commutation of \max_{γ} and \min_{x_f, x_i} as a consequence of the dependence on the initial conditions, i.e. of the non-ergodicity. Of course, if the LDF were convex, the ordering of these extremization would not matter.

The eigendecomposition of the propagator G involves the eigenvalues of a biased matrix [such as Eq. (65)]. For non-ergodic systems or when the state space is infinite, the assumption of the Perron–Frobenius theorem does not hold and several eigenvalues may cross each other. To compute exactly the LDF, one needs all the branches corresponding to each eigenvalue that becomes the highest eigenvalue for at least one γ . Only those branches matters, and other eigenvalues will not contribute, so as can be understood from the following argument: Let’s take two eigenvalues $\phi_1(\gamma)$ and $\phi_2(\gamma)$ such that $\phi_1(\gamma) > \phi_2(\gamma)$ for all γ . Since $\gamma \mathfrak{o} - \phi_1(\gamma) < \gamma \mathfrak{o} - \phi_2(\gamma)$, we have

$$I_1(\mathfrak{o}) = \max_{\gamma} \{ \gamma \mathfrak{o} - \phi_1(\gamma) \} < I_2(\mathfrak{o}) = \max_{\gamma} \{ \gamma \mathfrak{o} - \phi_2(\gamma) \}. \tag{39}$$

The last minimization in Eq. (35) on the partial LDFs will withdraw the contribution coming from the eigenvalue $\phi_2(\gamma)$ if it is smaller that $\phi_1(\gamma)$.

Physically speaking, we study a rare event in a system that has several independent subparts (also called ergodic components). We assume that the initial probability cannot be zero in all states of a subsystem, otherwise this subsystem shall be ignored. Each subpart of the system has its own probability to realize the rare event at stake. The subpart for which the event is the most likely will determine the event probability. This will be so if the rare event corresponds to a fluctuation of a time average quantity on a sufficiently long time so as to neglect the role of the initial state probability.

Mathematically speaking, when dealing with reducible biased operators whose highest eigenvalue is the CGF of interest, we must divide the operator into irreducible sub-operators for which holds the ensemble equivalence. For every sub-operators we proceed normally using the ensemble equivalence to determine the partial LDFs of the chosen observable from the Legendre–Fenchel transformation of the highest eigenvalue of the sub-operator. The final LDF for the total system is then given by the minimum over all partial LDFs. This explains why the final LDF is piecewise convex.

3.3 Mean-Field Ising Model

Using the results of Sect. 3.2, we study the activity and magnetization of an infinite range Ising model. This model is ergodic when considering a finite number of spins, but breaks ergodicity in the thermodynamic limit. This model is famous for its equilibrium phase transition at the thermodynamics [27] and dynamical level [26,36].

In the following, we first introduce the model and its mean field (MF) treatment. Second, we provide the propagator of the generating function for magnetization and activity, and next use it to determine the CGFs, LDFs and their convex hull.

3.3.1 Model Description and Thermodynamic Limit

We consider the fully connected Ising model made of N interacting spins $\{s\} = (s_1, \dots, s_N)$. Each spin s_i can hop between states $+1$ and -1 by exchanging heat with a thermostat at inverse temperature β . The interaction energy between two spins is V/N when the spins are not aligned and vanishes for parallel spins. The interaction energy is independent of the distance between spins. Beside the spin-spin interaction, each spin has a potential energy $-s_i E$ due to the presence of an external magnetic field. We introduce the free energy $F(n) \equiv U(n) - S^{\text{int}}(n)/\beta$ of the mesostate $n = \sum_{i=1}^N s_i$ in term of the total energy given (up to a constant) by

$$U(n) \equiv -\frac{V}{N} \sum_{1 \leq i \leq j \leq N} s_i s_j - E \sum_{1 \leq i \leq N} s_i = -n^2 \frac{V}{2N} - nE \tag{40}$$

and of the internal entropy

$$S^{\text{int}}(n) \equiv \ln N! / \left[\left(\frac{N+n}{2} \right)! \left(\frac{N-n}{2} \right)! \right]. \tag{41}$$

We chose the transition rate from n to $n + 2\epsilon$ (with $\epsilon = \pm 1$) to be

$$K(n + 2\epsilon, n) \equiv \Gamma(N - \epsilon n) e^{\frac{\beta}{2}((2\epsilon n + 2)V/N + 2\epsilon E)}. \tag{42}$$

In the following we chose $\Gamma = 1$ and $\beta = 1$ to set the time and energy scales respectively. The system is in thermal equilibrium and the transition rates satisfy the detail balance equation

$$\ln \frac{K(n + 2\epsilon, n)}{K(n, n + 2\epsilon)} = -\beta(F(n + 2\epsilon) - F(n)), \tag{43}$$

The probability of state n at time t , denoted $p_n = p_n(t)$ evolves according to the master equation

$$\dot{p}_n = \sum_{\epsilon=\pm 1} [K(n, n + 2\epsilon)p_{n+2\epsilon} - K(n + 2\epsilon, n)p_n], \tag{44}$$

where \dot{p}_n is the time derivative of p_n . The stationary probability $\pi(n)$ is the equilibrium probability

$$\pi(n) \equiv \frac{1}{Z_{\text{eq}}} e^{-F(n)} \quad \text{with} \quad Z_{\text{eq}} \equiv \sum_{n=0}^N e^{-F(n)}. \tag{45}$$

The time-averaged stochastic magnetization is obtained from Eq. (3) by choosing the state dependent function $f(n) = n/N$ and a vanishing function $g(n, n') = 0$ for all (n, n') :

$$M \equiv \frac{1}{NT} \int_0^T dt n(t), \tag{46}$$

where $n(t)$ is the mesostate at time t . We denote by m some real value that can be achieved by the stochastic variable M . The mean magnetization in the stationary state writes

$$\langle M \rangle = \frac{1}{N} \sum_{n=0}^N n\pi(n). \tag{47}$$

The activity rate is obtained from Eq. (3) by choosing $g(n, n') = 1/N$ and $f(n) = 0$ for all (n, n') :

$$A \equiv \frac{1}{NT} \sum_{0 \leq t \leq T: \Delta n_t \neq 0} 1 \tag{48}$$

The activity rate is thus a time-symmetric observable: The number of spin flips per unit time and per spin are identical in a trajectory and its time reversal. We denote by a some real value that can be achieved by the stochastic variable A . The mean activity in the stationary state writes

$$\langle A \rangle = \frac{1}{N} \sum_{n=0}^N \sum_{\epsilon=\pm 1} \pi(n) K(n + 2\epsilon, n). \tag{49}$$

In the thermodynamic limit, when taking the continuous limit for the mesostate $x \equiv n/N \in [-1, 1]$, the system energy changes by

$$\lim_{N \rightarrow \infty} [U(n + 2\epsilon) - U(n)] = 2\epsilon(Vx + E) \tag{50}$$

for a transition from n to $n + 2\epsilon$. Accordingly, the transition rates are in the same limit

$$J_\epsilon(x) \equiv \lim_{N \rightarrow +\infty} \frac{K(xN + 2\epsilon, xN)}{N} = (1 - \epsilon x) e^{\epsilon[Vx + E]}. \tag{51}$$

In this case, the master equation Eq. (44) can be transformed into an evolution equation for x in the mean field (MF) approximation

$$\langle \dot{x} \rangle = \sum_{\epsilon=\pm 1} \epsilon J_\epsilon(\langle x \rangle). \tag{52}$$

The steady state solution of this equation is the mean field magnetization $x = m^{\text{MF}}$ verifying

$$J_-(m^{\text{MF}}) = J_+(m^{\text{MF}}). \tag{53}$$

Using Eq. (51), the previous equation is equivalent to the transcendental equation

$$m^{\text{MF}} = \tanh(Vm^{\text{MF}} + E). \tag{54}$$

The MF activity follows from the mean-field magnetization:

$$a^{\text{MF}} = J_-(m^{\text{MF}}) + J_+(m^{\text{MF}}) = [1 - m^{\text{MF}}] e^{Vm^{\text{MF}} + E} + [1 + m^{\text{MF}}] e^{-Vm^{\text{MF}} - E}. \tag{55}$$

The MF magnetization and activity are shown in the bifurcation diagram of Fig. 2. At a critical value of the interaction energy, three MF magnetizations appear instead of a unique one, due to the well known ferromagnetic transition of the Ising model. This bifurcation also affects the system activity as shown on Fig. 2b and as expected from Eq. (55) since a^{MF} is a function of m^{MF} . We remark that for $E = 0$ the system activity is an even function of the magnetization and hence the bifurcation diagram for activity has two branches only. We also notice that the MF activity is higher for the branch that does not break the system symmetry.

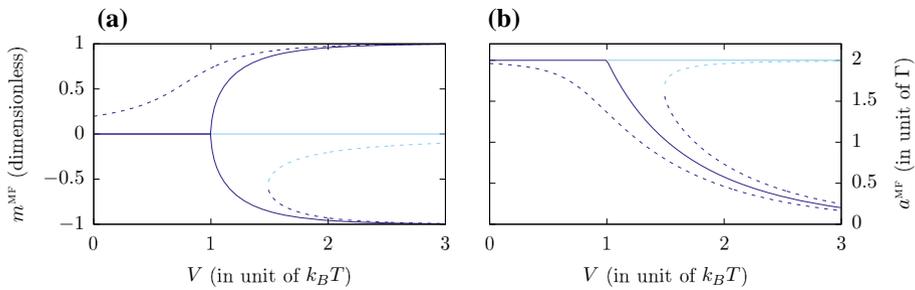


Fig. 2 **a** Stable (dark blue) and unstable (light blue) mean field steady state densities m^{MF} versus interaction energy V . **b** Stable (dark blue) and unstable (light blue) mean field steady state activity a^{MF} versus interaction energy V . $E = 0$ for solid lines and $E = 0.2$ for dashed lines (Color figure online)

3.3.2 Propagator of the Generating Function for Magnetization and Activity

Like in the four state model of Sect. 3.1, we look for the propagator of the generating function for the activity and magnetization:

$$G(x_f, x_i, \kappa, \gamma) = \langle e^{NT(\kappa m + \gamma a)} \rangle_{x_f, x_i}, \tag{56}$$

where as before the subscripts indicate a conditioning on the initial and final states, respectively $x_i = n_i/N$ and $x_f = n_f/N$. From a path integral approach, see Appendices A and B, an asymptotic expression of the propagator reads

$$G(x_f, x_i, \kappa, \gamma) \simeq_{N \rightarrow +\infty} \exp \left(N \int_0^T dt' \mathcal{L}(x_{t'}, \dot{x}_{t'}, \kappa, \gamma) \right), \tag{57}$$

where \mathcal{L} is the Lagrangian

$$\mathcal{L}(x, \dot{x}, \kappa, \gamma) = \frac{\dot{x}}{2} \ln \left(\frac{-\dot{x} + \sqrt{\dot{x}^2 + \varphi(x, \gamma)}}{2J_-(x)e^\gamma} \right) - \sum_{\epsilon=\pm 1} J_\epsilon(x) + \sqrt{\dot{x}^2 + \varphi(x, \gamma)} + \kappa x, \tag{58}$$

with

$$\varphi(x, \gamma) = 4 \prod_{\epsilon=\pm 1} J_\epsilon(x)e^\gamma = 4(1+x)(1-x)e^{2\gamma}. \tag{59}$$

The propagator of Eq. (57) is almost explicit: the path $[x]$ starting in x_i and ending in x_f must be determined using the Euler–Lagrange equation

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{x}} \right). \tag{60}$$

The propagator of the generating function is hence determined by the initial and final conditions.

As explained in Sect. 3.2, we now vary initial and final conditions to obtain the dominant contributions to the generating function. As the large size and long time limit of the propagator is evaluated with large size limit first and then long time limit, the system is non-ergodic. Due to this non-ergodicity, the dominant contributions are obtained from stationary trajectories, i.e. element of the propagator $G(x^*, x^*, \kappa, \gamma)$ such that

$$\frac{\partial \mathcal{L}}{\partial x}(x^*, 0, \kappa, \gamma) = 0. \tag{61}$$

and we denote $x^*(\kappa, \gamma)$ the various solutions of this equation. These constant trajectories are the only ones that will dominate for at least one value of (κ, γ) . Following Sect. 3.2, we can restrain the extremization over initial and final conditions to these stationary trajectories.

We provide a heuristic argument for the choice of stationary trajectories: Dominant trajectories correspond to long-time behavior of the tilted system. We expect the tilted system to be in a stationary state at long-time, as we study an equilibrium system. Neglecting the boundary terms due to long time limit, we restrain to trajectories starting and ending into a stationary state. The system being non-ergodic, the state space is divided into various subpart each associated to different x^* and we forbid trajectories that start and end into different subpart of the state space. This approach is confirmed by numerical computation of the CGF in the next section.

3.3.3 CGF of Magnetization and Activity

We now use the propagator of Eq. (57) to derive the CGF of magnetization and activity. This CGF proceeds from the leading elements of the propagator of the generating function

$$\begin{aligned} \phi(\kappa, \gamma) &= \lim_{N, T \rightarrow \infty} \frac{1}{NT} \ln \langle e^{NT\kappa m + NT\gamma a} \rangle, \\ &= \max_{x_f, x_i} \lim_{N, T \rightarrow \infty} \frac{1}{NT} \ln G(x_f, x_i, \kappa, \gamma) p_i(x_i). \end{aligned} \tag{62}$$

As explained above, we can focus on stationary trajectories that solve Eq. (61). Then, solving for x amounts to find the extrema of $\mathcal{L}(x, 0, \kappa, \gamma)$. Assuming that $p_i(x) > 0$ for all x , one can safely drop the initial probability p_i in Eq. (62) ending with

$$\phi(\kappa, \gamma) = \max_x \mathcal{L}(x, 0, \kappa, \gamma), \tag{63}$$

$$= \sqrt{\varphi(x^*(\kappa, \gamma), \gamma)} - \sum_{\epsilon=\pm 1} J_\epsilon(x^*(\kappa, \gamma)) + \kappa x^*(\kappa, \gamma). \tag{64}$$

Unfortunately, the determination of $x^*(\kappa, \gamma)$ involves a transcendental equation. We solved this equation numerically to provide the CGFs before and after the bifurcation in Fig. 3c and d respectively. Cross-sections of the CGF in the plane $\kappa = 0$ and $\gamma = 0$ are shown in Fig. 3a and b. After the bifurcation for $V = 1.5$, the CGF is clearly not differentiable. The left and right partial derivatives at $(\kappa, \gamma) = (0, 0)$ lead to different mean magnetization and activity in agreement with the bifurcation diagram of Fig. 2 of the mean-field framework. We notice on Fig. 3d that before the transition the CGF has a non differentiability not located at the origin of the (γ, κ) plane. Hence, the mean magnetization and activity are unique, but their fluctuations are impacted by the phase transition. We confirm our results by computing numerically the CGF as the highest eigenvalue of the biased matrix

$$\mathbf{K}_{\gamma, \kappa}(x, y) = \begin{cases} K(x, y)e^{\gamma/N} & \text{if } x \neq y \\ K(x, x) + \kappa x/N & \text{if } x = y \end{cases}, \tag{65}$$

for systems with $N = 10$ and $N = 25$ spins. As explained in Sect. 2.3, the CGFs of finite size systems are everywhere differentiable. Then, they only approach gradually the non-differentiable CGF when $N \rightarrow \infty$.

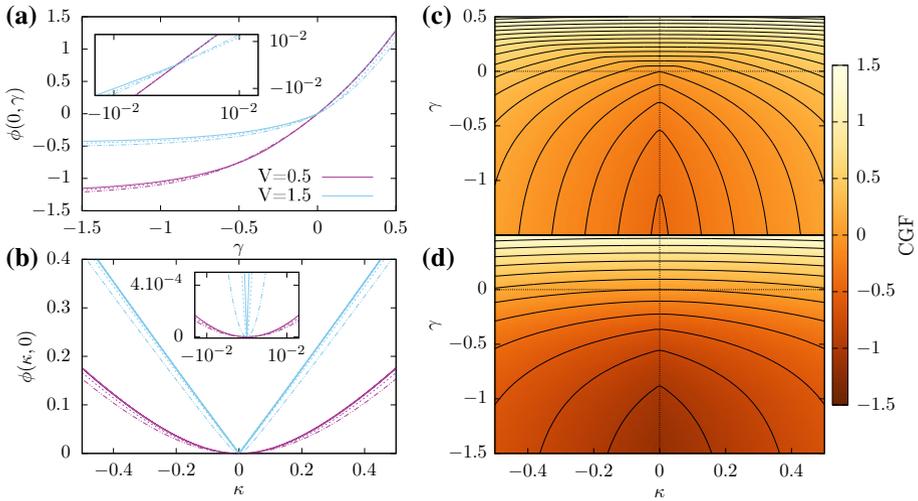


Fig. 3 Cross-sections of the CGF $\phi(\kappa, \gamma)$ along **a** $\kappa = 0$ and **b** $\gamma = 0$ (solid lines), and corresponding CGFs for the finite size system with $N = 10$ (dot-dashed lines) and $N = 25$ particles (dashed lines). Insets: Zoom on the non-differentiable point of the light blue line for which $V = 1.5$. **c** CGF $\phi(\kappa, \gamma)$ and level lines for $V = 1.5$, **d** CGF $\phi(\kappa, \gamma)$ and level lines for $V = 0.5$. For all figures $E=0$ (Color figure online)

3.3.4 Legendre–Fenchel Transformation of the CGF of Magnetization and Activity

The cLDF $I^{**}(m, a)$ shown in Fig. 4c and d is the Legendre–Fenchel conjugate of the CGF

$$I^{**}(m, a) = \max_{\kappa, \gamma} \{ \kappa m + \gamma a - \phi(\kappa, \gamma) \}. \tag{66}$$

At low interaction energy V , we observe a smooth function whose unique minimum is given by the mean-field solution of Eqs. (54)–(55). However, at higher interaction energy, a plateau appears in the cLDF between the MF solutions. This plateau, in association with the non-differentiability of the CGF, indicates a phase transition. As emphasized before, we always have ergodicity within the canonical ensemble, and this plateau is the result of a “temporal coexistence” of MF states: the system spends most of its time into the various MF states leading to a time averaged magnetization and activity belonging to the convex area defined by the MF solutions, here a triangle.

The contracted cLDF for activity $I^{**}(a)$ and magnetization $I^{**}(m)$ defined by

$$I^{**}(m) \equiv \min_a I^{**}(m, a) = \max_{\kappa} \{ \kappa m - \phi(\kappa, 0) \} \tag{67}$$

$$I^{**}(a) \equiv \min_m I^{**}(m, a) = \max_{\gamma} \{ \gamma a - \phi(0, \gamma) \}, \tag{68}$$

are shown in Fig. 4a and b, together with the LDFs for the finite size systems obtained from the Legendre transformation of their corresponding CGFs in Fig. 3a and b. For $V = 1.5$, the latter LDFs converge towards the plateau with a speed that is lower for the activity LDF than for the magnetization LDF in agreement with the fact that the plateau for the LDF of activity lies between a stable MF solution and an unstable MF solution, while the plateau for the magnetization lies between two stable solutions.

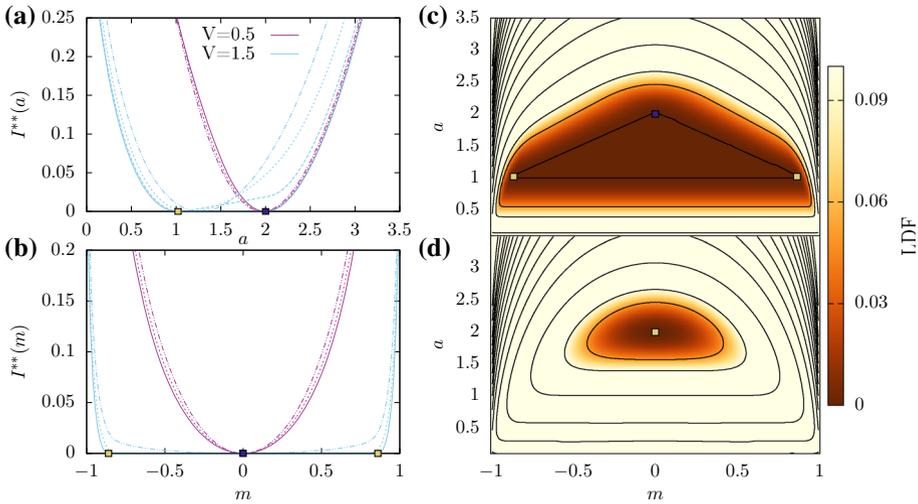


Fig. 4 **a** cLDF $I^{**}(a)$ (solid lines) and finite size LDF for $N = 10$ (dot-dashed lines), $N = 25$ (dashed lines) and $N=100$ (long-dashed lines). **b** cLDF $I^{**}(m)$ (solid lines) and finite size LDF for $N = 10$ (dot-dashed lines) and $N = 25$ (dashed lines). Parameters are for **(a, b)**: $V = 0.5$ (magenta lines) and $V = 1.5$ (light blue lines). **c–d** cLDF $I^{**}(m, a)$ of activity and magnetization and level lines **(c)** for $V = 1.5$ and **(d)** for $V = 0.5$. Beige squares indicate the location of the stable solutions of Eqs. (54)–(55) whereas dark-blue squares are for unstable solutions. For all figures $E = 0$. For figure **(c–d)**, the color map replaces the level lines for low values of the LDFs (Color figure online)

3.3.5 LDF of Magnetization and Activity

As explained in Sect. 3.2, the LDF follows from the Legendre–Fenchel transformation of the elements of the propagator of the generating function. Therefore, Eq. (35) yields

$$I(m, a) = \min_{x_f, x_i} \max_{\kappa, \gamma} \left[\kappa m + \gamma a - \lim_{N, T \rightarrow \infty} \frac{1}{NT} \ln G(x_f, x_i, \kappa, \gamma) p_i(x_i) \right] \tag{69}$$

Each stationary solution of Eq. (61) denoted x^* defines a partial LDF

$$I_{x^*}(m, a) = \max_{\kappa, \gamma} \{ \kappa m + \gamma a - \mathcal{L}(x^*(\kappa, \gamma), 0, \kappa, \gamma) \}. \tag{70}$$

If the initial magnetization is in the same subpart of the state space than x^* , the system’s fluctuations are best described by $I_{x^*}(m, a)$. When ensemble averaging on the initial condition, we look for the minimum on the stationary trajectories to obtain the LDF

$$I(m, a) = \min_{x \in \{x^*\}} I_x(m, a). \tag{71}$$

From the fact that I^{**} is the convex hull of I , we get the following inequality between the two LDFs:

$$I(m, a) \geq I^{**}(m, a). \tag{72}$$

The LDF $I(m, a)$ is shown after the bifurcation on Fig. 5c, We also provide in Fig. 5a and b the partial LDF for activity and magnetization (after contraction) and their convex hulls. As expected, the LDF is not convex: the ensemble equivalence does not hold (in a specific interval of magnetization and activity) for our model in the thermodynamic limit, due to non-ergodicity.

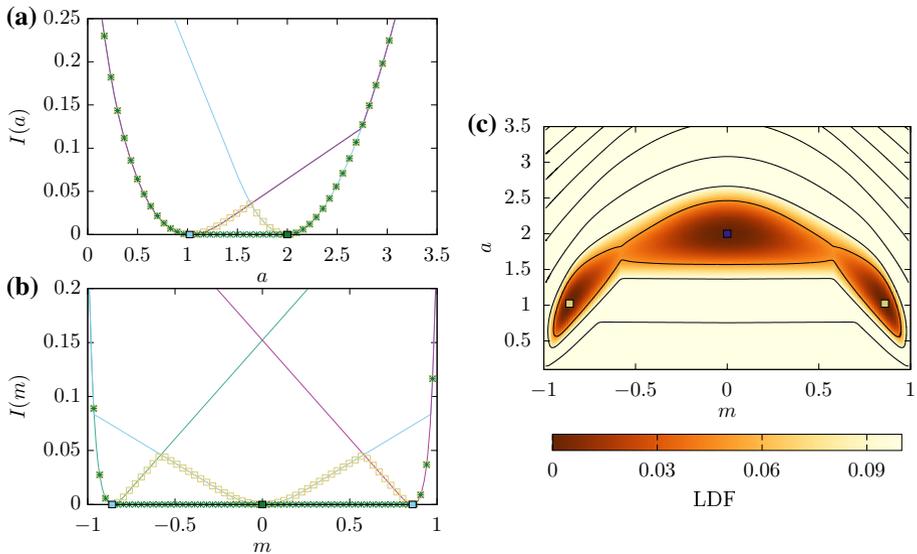


Fig. 5 Branches of the LDF of (a) activity and (b) magnetization (solid lines). Lines with symbols represent the cLDF (crosses) and the LDF (open squares). c Level lines of the LDF of activity and magnetization. For all figures, parameters are: $E = 0$ and $V = 1.5$. For figure (c), the color map replaces the level lines for low values of the LDF (Color figure online)

Comparing the cLDFs obtained from Eqs. (67)–(68) and the LDFs, we notice that the formers are as expected the convex hull of the latters.

4 Non-convex LDF and Divergent Mixing Time

In the previous sections, we have obtained the LDF of activity and magnetization for the canonical and microcanonical ensembles. In the thermodynamic limit, the ensemble equivalence is broken at high interaction even though it holds at finite size. We now explore the transition from equivalence to non-equivalence when increasing the size of the system, putting the emphasis on the order of the large size and long time limits in the computation of the statistics of magnetization and activity.

In this section, we point out the existence of a mixing time t_{mix} that depends on the system size. For systems of finite size, the mixing time governs the fluctuation regime. First, we define the mixing time from the spectral gap of the finite size operator \mathbf{K} . Second, we give an estimate of the mixing time for the Ising model and prove that it diverges when $N \rightarrow \infty$. This means that it is impossible to fully relax from the initial condition when the system includes too many spins, leading to an ergodicity breaking. Finally, we explore the different regimes of fluctuations at finite size and time with numerical simulations, enlightening the coherence of our previous results.

For the numerical simulations of this section, the magnetic field is non-zero ($E = 0.2$) in order to break the up-down symmetry of the Ising model. Thanks to the magnetic field, two different stable magnetizations exist associated to different stable activities. Then, the activity probability density function will be bimodal as it is beyond numerical reach (at large N) to detect the third unstable magnetization and its associated activity. Without the magnetic field,

just the unique stable activity would appear on the numerical simulations and the activity probability density function would be unimodal. For magnetization, we would observe only bimodal probability density function for symmetry reasons.

4.1 Gap at Finite Size

We come back to the continuous-time process of Eq. (1) of Markov operator K . We assume that K is diagonalizable. The eigenvalues are $\{\lambda_i\}_{i=1,N}$, where $\lambda_1 = 0$ and all other eigenvalues have negative real parts. Its set of eigenvectors is $\{l_i, r_i\}_{i=1,N}$ where l_1 is a uniform vector and $r_1 = \pi$ is the stationary probability. The spectral gap $\Delta\lambda$ of the operator K is the difference between the largest eigenvalue λ_1 and the real part of its second eigenvalues λ_2 .

Considering the initial probability p_i , the probability

$$p(x, t) = \left(e^{tK} p_i \right) (x) = \pi(x) + \sum_i e^{t\lambda_i} r_i(x) \sum_x l_i(x) p_i(x) \tag{73}$$

is a formal solution to Eq. (1). Since the spectral gap $\Delta\lambda$ of the operator K is positive, the probability $p(x, t)$ converges towards π . The mixing time $t_{\text{mix}}(\varepsilon)$ is used to quantify the time that the system takes to relax to the stationary probability [37]. By definition, the mixing time $t_{\text{mix}}(\varepsilon)$ is the minimal time for which starting from any initial probability the system is at most at a distance ε of the stationary probability. Formally, the mixing time is

$$t_{\text{mix}}(\varepsilon) \equiv \inf \left\{ t \geq 0 : \max_{p_i} \| e^{tK} p_i - \pi \|_{TV} \leq \varepsilon \right\}, \tag{74}$$

where the total variation distance is defined as $\|u\|_{TV} = \sup_x u(x)$. A common choice for ε is e^{-1} , such that $t_{\text{mix}} \equiv t_{\text{mix}}(e^{-1})$. The mixing time quantifies the time needed to reach stationary probability whatever the initial probability. In particular, an infinite mixing time is a feature of non-ergodic systems. In practice, when considering large deviation statistics, we must formally take a long time limit, i.e. a time much longer than the mixing time.

The mixing time of Eq. (74) is a maximization over all initial probabilities, therefore looking at $\| e^{tK} p_i - \pi \|_{TV}$ for a uniform initial probability $p_i = 1$ underestimates the mixing time. We plot on Fig. 6c the evolution of this distance for various sizes. First, we observe an exponential scaling of the total variation distance with time. When comparing with Eq. (73), we expect the mixing time to be connected with the spectral gap. Indeed, we have for Markov processes [37]

$$\frac{1}{\Delta\lambda} \leq t_{\text{mix}} \leq -\frac{\log p_i^{\min}}{\Delta\lambda} \tag{75}$$

where $p_i^{\min} = \min_x p_i(x)$ and $\Delta\lambda$ is the spectral gap. Second, at very short times, the evolution of total variation distance has a different scaling. This short time behavior is due to the other eigenvalues whose influence on the probability $p(x, t)$ disappears quickly. Finally, the evolution of the total variation distance with system size reveals the strong dependencies of the mixing time on the system size allowing for longer and longer transient behavior.

At finite time, the fluctuations critically differ if $T \gg t_{\text{mix}}$ or if $T \ll t_{\text{mix}}$. On Fig. 6a and b we show the empirical density probability of activity and magnetization for a system of $N = 60$ spins and for different durations T , with $t_{\text{mix}} \simeq 100$. And on Fig. 6d, we plot the empirical probability density of activity for a system size $N = 200$ where $t_{\text{mix}} \simeq 10^5$.

For $T \gg t_{\text{mix}}$, the long time probabilities of magnetization and activity are those of an ergodic system (the system has enough time to switch between MF states). The asymptotic probability is then given by the cLDF of Sect. 3.3.4. The convergence toward the plateau

for these cLDFs is discussed in the next section. Notice that the added small magnetic field is responsible for the unimodality of the probability density functions, see Fig. 6a and b at $T = 300$.

Otherwise when the second eigenvalue is well separated from the others, i.e. when $\Delta\lambda \ll |\lambda_0 - \lambda_3|$, the probability $p(x, t)$ is well approximated for $|\lambda_0 - \lambda_3|^{-1} \ll T \ll t_{\text{mix}}$ by

$$p(x, t) \simeq \pi(x) + e^{-t\Delta\lambda} r_1(x) \sum_y l_1(y) p_i(y). \tag{76}$$

It contains a term coming from the second eigenvalue. This second term induces the secondary peak on the probability density function. Therefore, for intermediate times before the mixing time, the system behaves as an effective non-ergodic system with fluctuations around each MF states (the system lacks of time to switch between states). We have a transient behavior where the probabilities of magnetization and activity are then bimodal, see Fig. 6d, and are linked with the non-convex LDF of Sect. 3.3.5.

4.2 Estimation of the Gap for Systems with Detailed Balance

We now determine the scaling of the mixing time with the system size. To do so, we find an upper bound on the spectral gap of systems with detailed balance which, combined with the inequality of Eq. (75), estimates the mixing time scaling.

For Markov processes with detailed balance, we can estimate the spectral gap from the Cheeger bound [34]. Indeed, in this case the Markov operator is symmetrizable, and we can use known results on the gap of symmetric matrices. We introduce the matrix $D_{\sqrt{\pi}}$ as the diagonal matrix with elements $D_{\sqrt{\pi}}(n, n) \equiv \sqrt{\pi_n}$, then the matrix $D_{\sqrt{\pi}} K D_{\sqrt{\pi}}^{-1}$ is symmetric if the Markov matrix K respects detailed balance. We introduce the Cheeger constant as

$$\Phi = \inf_{\mathcal{E} \subset \Omega, 0 < \pi(\mathcal{E}) \leq 1/2} \frac{Q(\mathcal{E}^c, \mathcal{E})}{\pi(\mathcal{E})}, \tag{77}$$

where \mathcal{E} is any subset of the set of state of our system Ω such that $\pi(\mathcal{E}) = \sum_{x \in \mathcal{E}} \pi(x) < 1/2$ and,

$$Q(\mathcal{E}^c, \mathcal{E}) = \sum_{x, y | x \in \mathcal{E}, y \in \mathcal{E}^c} \pi(x) K(y, x) \tag{78}$$

is the sum of probability flow from the subset \mathcal{E} to the complementary of \mathcal{E} , denoted \mathcal{E}^c . From Ref. [34]

$$\Delta\lambda \leq 2\Phi. \tag{79}$$

Computing the Cheeger constant is difficult in general, but it is possible to bound it. For a subset \mathcal{E} , the probability flow from \mathcal{E} to \mathcal{E}^c is bounded by

$$Q(\mathcal{E}^c, \mathcal{E}) \leq N_b \pi(\acute{x}) K(\acute{y}, \acute{x}) \tag{80}$$

where N_b is the number of edge connecting \mathcal{E} and \mathcal{E}^c and

$$(\acute{x}, \acute{y}) = \arg \max_{x \in \mathcal{E}, y \in \mathcal{E}^c} \pi(x) K(y, x), \tag{81}$$

the edge supporting the biggest probability flow from \mathcal{E} to \mathcal{E}^c . Denoting then \tilde{x} the most probable state in \mathcal{E} , we have

$$\Phi \leq \frac{N_b \pi(\acute{x}) K(\acute{y}, \acute{x})}{\pi(\tilde{x})}. \tag{82}$$

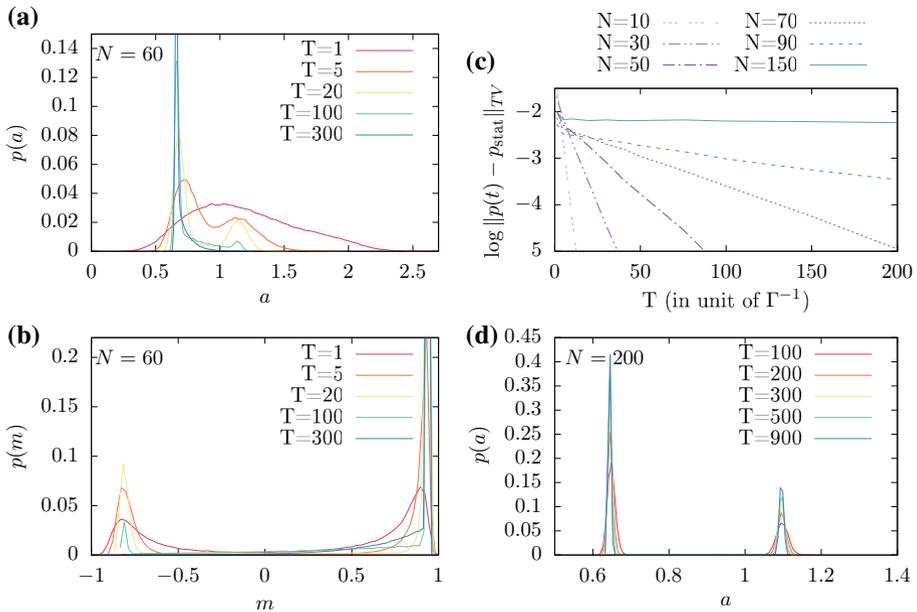


Fig. 6 **a** Activity and **b** magnetization probability density functions obtained from numerical simulations of various duration T for $N = 60$. **c** Logarithm of the total variation distance between $e^{t\mathbf{K}} p_1$ and π as a function of the duration of the evolution for various system size N . We simulate the evolution of N spins using the Gillespie algorithm. For each size and time, a total of 15×10^4 trajectories are drawn. The initial condition of the trajectory is draw from uniform distribution. The probability density functions are computed from a histogram of 75 bins between the minimum and maximum values. The total variation distance is the maximum of the difference between an histogram of the final value of the trajectory and an histogram of 15×10^4 points drawn from probability of Eq. (45). **d** Activity probability density functions obtained from numerical simulations of various duration T for $N = 200$. The parameters are: $V = 1.7, E = 0.2$

Since we are mainly interested in the large size limit of the mixing time, we now assume that the stationary probability satisfies a large deviation principle

$$\pi(x) \simeq_{N \rightarrow +\infty} e^{-N\mathcal{I}(x)}. \tag{83}$$

We consider a connected subset of states \mathcal{E} such that the state x^{MF} with $\mathcal{I}(x^{\text{MF}}) = 0$ is not in \mathcal{E} . For large enough N , the probability of \mathcal{E} is surely less than $1/2$. Using the bound (82) on the Cheeger constant, we find a large deviation estimate as

$$\Phi \leq \frac{N_b \pi(\hat{x}) K(\hat{y}, \hat{x})}{\pi(\tilde{x})} \simeq N_b K(\hat{y}, \hat{x}) e^{N(\mathcal{I}(\tilde{x}) - \mathcal{I}(\hat{x}))}. \tag{84}$$

Therefore, if it exists \mathcal{E} such that the LDF $\mathcal{I}(x)$ has a local minimum $\tilde{x} \in \mathcal{E}$ that is not a global minimum, we will have $\mathcal{I}(\tilde{x}) - \mathcal{I}(\hat{x}) < 0$. Then, if the product $N_b K(\hat{y}, \hat{x})$ is not diverging exponentially, we have bounded the spectral gap by something going to 0 as $N \rightarrow +\infty$. Hence, the mixing time diverges with the system size, and the divergence is even exponential.

In our case, we consider the subset of state $\mathcal{E}^- = \{x | x \leq 0\}$, if $E \geq 0$ it has a stationary probability less than $1/2$, otherwise if $E < 0$, we consider the subset $\mathcal{E}^+ = \{x | x \geq 0\}$. We have then $N_b = 1, \hat{x} = 0^-$ and $\hat{y} = 0^+$, the lower and upper closest states to 0. The stationary probability respects a large deviation principle, with a local minimum appearing

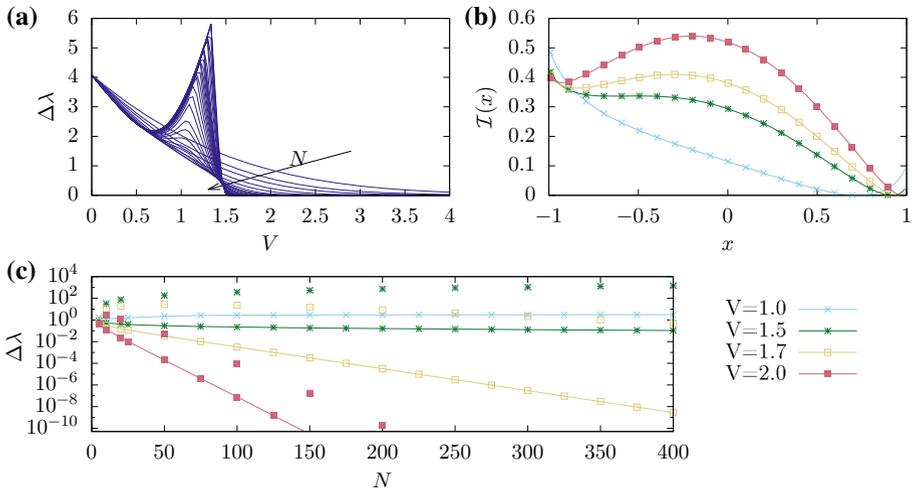


Fig. 7 **a** Spectral gap as a function of the interaction energy V for various system sizes. **b** Large size LDF of x defined by $\mathcal{I}(x) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \pi(x)$ for several interaction energies V . **c** Spectral Gap (lines with symbols) and bound of Eq. (84) (symbols) as a function of the system size for various interaction energy V . For all figures, we take $E = 0.2$

after the phase transition, see Fig. 7b. Therefore, the mixing time is diverging after the phase transition, but not before.

These results are confirmed by numerical computation of the spectral gap. On Fig. 7a, we plot the spectral gap as a function of the system size N and of the interaction energy V . Before the phase transition, the spectral gap remains finite and so does the mixing time. After the phase transition, the spectral gap goes to zero when increasing N . The speed of convergence of the spectral gap is well caught by the bounds (84) as indicates the exponential decay with increasing N shown on Fig. 7c. The mixing time is so at least diverging exponentially with the system size.

4.3 Mixing Time as a Criterion to Study Fluctuations

From the previous results, we are now able to explain the effective transition from non-ergodicity to ergodicity as the observation time becomes longer than the mixing time. At the level of the dynamical ensembles, this corresponds to a transition from non-equivalence to equivalence of ensembles. The fluctuations differ before and after the mixing time that in addition is diverging with the system size N . In a first place, if we take the long time limit before large size limit, we overcome the mixing time and the system stays ergodic. As the equivalence remains valid, the fluctuations are correctly given by the cLDF. In a second place, if the large size limit is taken first, the mixing time is infinite such that we stay in the regime of fluctuations before the mixing time, the system is not ergodic and the LDF is likely to be non-convex. Large deviation theory aims at describing asymptotic fluctuations, but the limits should not be understood strictly: it is fundamental to distinguish the mixing time and its dependency on the system size to interpret them correctly and use this framework for real experimental systems or numerical simulations.

5 Discussion and Conclusion

Like for equilibrium ensembles, ergodicity breaking is fundamental to understand the non-equivalence of dynamical ensembles. Considering the connection between state variables and dynamical observables, equilibrium systems without equivalence of equilibrium ensembles will also exhibit non-equivalence of dynamical ensembles. We have provided general arguments supporting that ergodicity is required, but may not be sufficient to observe non-equivalence between the microcanonical and canonical dynamical ensembles. We have considered a system without ergodicity (by construction) and one with emergent non-ergodicity. In both cases we have obtained LDFs with an explicit dependency in the initial condition, the so-called partial LDFs. From these partial LDFs, the LDF follows from minimizing on the initial condition. For systems with emergent ergodicity breaking we have discussed the physical meaning of both the non-convex and convex LDFs. We did so by comparing the observation time and the mixing time after which the initial condition can be forgotten.

At a different level of large deviation, one may investigate the structure of the trajectories corresponding to the null minima of the LDF and the relative weight of these minima. When the system is not ergodic and the LDF is non-convex, we observe separate zeros of the LDF that can be associated to the most probable values of the dynamical observable. In this case, trajectories remain in a subpart of the state space given by the initial condition. For the fully connected Ising model, the most probable states are the MF magnetizations and their relative probabilities are determined from the initial probability to start within each subpart of the state space.

On the contrary, when the system is ergodic and the LDF convex, we observe a whole continuous set of magnetizations and activities for which the LDF is zero, e.g. the triangle region in Fig. 4c. These values correspond to trajectories where the system spends a fraction of time around one MF state and another fraction around another MF state. Then, the final value of the observable is the time average on these trajectories. The move between MF states is allowed thanks to ergodicity. These trajectories are instantons of the Lagrangian, i.e. trajectories going from a stationary solution of Euler–Lagrange equations to another [22,28,45]. The typical time scale (connected to the mixing time) at which these instantons occur depends on the system size, as explained for driven diffusive channels in Ref. [1]. In the present paper, we did not take instantons into account for the minimization of Eq. (62). However, taking them into account does not modify the cLDF at the considered level of large deviations, but only describes fluctuations inside the linear region (here a plateau) of the cLDF [46].

In our framework, we have assumed that it is always possible to find at least one stationary stochastic process (called driven process) that reproduces a fluctuation of the dynamical observables as a typical event. Ensemble equivalence means that this driven process is unique and does not depend on the initial condition. In our framework, ensemble non-equivalence means that there are several driven processes according to the initial conditions. We did not consider the case where no stationary driven process exists. This may happen for diffusive processes for which the state space is non-compact and the system relaxes while never reaching a stationary state. These cases were associated with a partial equivalence of dynamical ensembles in Refs. [42,44,49].

Model of glasses are known to have large relaxation time scales that are beyond numerical and experimental reach. The characterization of their fluctuations during this relaxation is an

active field of research. Since we characterized fluctuations before and after mixing time, we expect that our framework could be useful to study metastable states of glasses.

Acknowledgements We acknowledge Massimiliano Esposito for his advice on this project that started when GV was a post-doctoral fellow in his research team. We thank V. Lecomte and A. Lazarescu for the insightful discussions in connection with this work.

Appendix A: Derivation of the Propagator for Generating Function

We look for an expression of the propagator of the generating function $G(x_f, x_i, \kappa, \gamma) = \langle e^{NT(\kappa m + \gamma a)} \rangle_{x_f, x_i}$ as a path integral over all trajectories for the system state n . We use the discrete time $t_k = kd$ corresponding to the k 's time step of duration dt , with the final time $T = t_L = Ldt$. We write the system state $n_{t_k} = n_k$ for short. This state at time t_k changes by $n_k - n_{k-1} = 2\epsilon_k = -2, 0$ or 2 when a spin jumps from up to down, doesn't jump or jumps from down to up. The sum over all paths for a given initial condition is

$$\sum_{[\epsilon_k]} = \sum_{\epsilon_0} \sum_{\epsilon_1} \cdots \sum_{\epsilon_{N-1}}. \tag{85}$$

Since jump probabilities depend on the magnetization, the path integration should include a sum on the system state n_k for all k

$$\sum_{[n_k]} = \sum_{n_1} \sum_{n_2} \cdots \sum_{n_{L-1}}, \tag{86}$$

with the condition that one spin at most jumps at each time step. In the above sums, the index n goes from $-N$ to N . All in all, the propagator of the generating function writes

$$G(n_i, n_f, \kappa, \gamma) = \sum_{[n_k, \epsilon_k]} \left[\prod_{k=1}^L p(n_k | n_{k-1}) \delta(n_k - n_{k-1} - 2\epsilon_{k-1}) \right] \times e^{\kappa \left(\sum_{k=1}^L n_{k-1} dt \right) + \gamma \left(\sum_{k=1}^L \epsilon_k \right)}, \tag{87}$$

with $p(n_k | n_{k-1})$ the probability of the transition $n_{k-1} \rightarrow n_k$. Let's use the Laplace representation of the Dirac distribution

$$\delta(n_k - n_{k-1} - 2\epsilon_{k-1}) = \int_{ib_k - \infty}^{ib_k + \infty} \frac{dp_k}{2\pi} e^{ip_k(n_k - n_{k-1} - 2\epsilon_{k-1})}, \quad b_k \in \mathbb{R}, \tag{88}$$

for every integer k between 1 and L to get

$$G(n_i, n_f, \kappa, \gamma) = \sum_{[n_k, \epsilon_k]} \int \left(\prod_{k=1}^L \frac{dp_k}{2\pi} \right) \times \left[\prod_{k=1}^L p(n_k | n_{k-1}) e^{ip_k(n_k - n_{k-1} - 2\epsilon_{k-1}) + \kappa n_{k-1} dt + \gamma} \right]. \tag{89}$$

In the next step, we sum over $[\epsilon_k]$ to obtain

$$G(n_i, n_f, \kappa, \gamma) = \sum_{[n_k]} \int \left(\prod_{k=1}^L \frac{dp_k}{2\pi} e^{ip_k(n_k - n_{k-1})} \right)$$

$$\times \prod_{k=1}^L \left[p(n_{k-1}|n_{k-1}) + \sum_{\epsilon=\pm 1} p(n_{k-1} + \epsilon|n_{k-1}) e^{-2i p_k \epsilon + \gamma} \right] e^{\kappa n_{k-1} dt}. \tag{90}$$

The propagator for staying in the same state during a time step is $p(n_k|n_k) = 1 - \left(\sum_{\epsilon=\pm 1} K(n_k + 2\epsilon, n_k) \right) dt$, and when changing of state writes $p(n_k + \epsilon|n_k) = K(n_k + 2\epsilon, n_k) dt$. Exponentiating the product yields

$$G(n_i, n_f, \kappa, \gamma) = \sum_{[n_k]} \int \left(\prod_{k=1}^L \frac{dp_k}{2\pi} \right) \exp \left(i dt \sum_{k=1}^L p_k \frac{(n_k - n_{k-1})}{dt} \right) \times \exp \left[\sum_{k=1}^L dt \left(\sum_{\epsilon} K(n_{k-1} + 2\epsilon, n_{k-1}) (e^{-2i p_k \epsilon + \gamma} - 1) + \kappa n_{k-1} \right) \right]. \tag{91}$$

The propagator of the generating function becomes in the continuous limit

$$G(x_f, x_i, \kappa, \gamma) = \int \mathcal{D}[x] \int \mathcal{D}[p] \exp \left[N \int_0^T dt i p_t \dot{x}_t - \int_0^T dt \left(\sum_{\epsilon} J_{\epsilon}(x_t) (1 - e^{-2i p_t \epsilon + \gamma}) + \kappa x_t \right) \right] \tag{92}$$

where $\int \mathcal{D}[x]$ (resp. $\int \mathcal{D}[p]$) is a short notation for the integral on the paths of density (resp. momentum) with given initial and final values. Finally, the path integral expression of the generating function is given by

$$G(x_f, x_i, \kappa, \gamma) = \int \mathcal{D}[x] \int \mathcal{D}[p] \exp (N \mathcal{A}_{\kappa, \gamma}[x, p, T]) \tag{93}$$

with the action $\mathcal{A}_{\kappa, \gamma}$ being a functional of the paths $[x, p]$

$$\mathcal{A}_{\kappa, \gamma}[x, p, T] = \int_0^T dt i p_t \dot{x}_t - \int_0^T dt \mathcal{H}(x_t, p_t, \kappa, \gamma). \tag{94}$$

We have introduced the Hamiltonian function

$$\mathcal{H}(x, p, \kappa, \gamma) = \sum_{\epsilon=\pm 1} [1 - \exp(-2i p \epsilon + \gamma)] J_{\epsilon}(x) - \kappa x. \tag{95}$$

This function is the representation in the continuous limit of a tridiagonal Metzler matrix, whose spectrum is real.² As a consequence, $i p$ is a real number.

Appendix B: Wentzel–Kramers–Brillouin Approximation

We now use a saddle point integration to obtain the leading order in N of the propagator of the generating function $G(x_f, x_i, \kappa, \gamma)$ which is asymptotically exact in the large size

² This property is stated in problem 5, page 174 of Ref. [30].

limit. For each integration, we deform the contour so that it goes through the saddle point of $\mathcal{A}_{\kappa,\gamma}[x, \mathbf{p}, t]$. Therefore, the saddle point is the minimum of $\mathcal{A}_{\kappa,\gamma}[x, \mathbf{p}, t]$ solving

$$\frac{\delta}{\delta \mathbf{p}_t} \mathcal{A}_{\kappa,\gamma}[x, \mathbf{p}, T] = 0, \tag{96}$$

$$\frac{\delta}{\delta x_t} \mathcal{A}_{\kappa,\gamma}[x, \mathbf{p}, T] = 0, \tag{97}$$

where $\frac{\delta}{\delta x_t}$ denotes a functional derivative.

In this case, the saddle point calculation is equivalent to the WKB approximation of quantum mechanics. The propagator of the generating function will be given by the paths maximizing the action, i.e. the classical paths starting at x_i and ending at x_f . These paths solve the Hamilton's equations

$$i\dot{x}_t = \partial_{\mathbf{p}} \mathcal{H}(x_t, \mathbf{p}_t, \kappa, \gamma), \tag{98}$$

$$i\dot{\mathbf{p}}_t = -\partial_x \mathcal{H}(x_t, \mathbf{p}_t, \kappa, \gamma). \tag{99}$$

The propagator of the generating function can be written as

$$G(x_f, x_i, \kappa, \gamma) \simeq_{N \rightarrow +\infty} \exp \left[N \int_0^T dt (i\mathbf{p}_t \dot{x}_t - \mathcal{H}(x_t, \mathbf{p}_t, \kappa, \gamma)) \right], \tag{100}$$

where (x, \mathbf{p}) correspond to the solutions of Eqs. (98)–(99) with initial and final conditions x_i and x_f . Moreover the Hamiltonian (95) is time independent, and is hence a conserved quantity along the trajectory. Maupertuis' action $\int_0^T dt i\mathbf{p}_t \dot{x}_t$ is a function of the Hamiltonian and writes

$$\int_0^T dt i\mathbf{p}_t \dot{x}_t = \int_{x_i}^{x_f} dx i\mathbf{p}(x). \tag{101}$$

where $\mathbf{p}(x)$ can be obtained by inverting Eq. (95). We observe that Maupertuis' action is always negative in view of the clockwise motion along the orbits that can be justified by inspection of Eq. (98).

In the large size limit, we can also use a saddle point integration on \mathbf{p} to switch to the Lagrangian framework. To do so, we look for the solution of Eq. (98). We derive a quadratic equation either for $e^{i\mathbf{p}_s}$ or for $e^{-i\mathbf{p}_s}$ whose solutions are

$$e^{\pm 2i\mathbf{p}_s} = \frac{\mp \dot{x}_s + \sqrt{\dot{x}_s^2 + \varphi(x_s, \gamma)}}{2J_{\mp}(x_s)e^{\gamma}}, \tag{102}$$

with $\varphi(x, \gamma)$ defined in Eq. (59). Inserting this into Eq. (93) leads to the propagator of the generating function of Eq. (57) with Lagrangian (58).

References

1. Baek, Y., Kafri, Y., Lecomte, V.: Dynamical phase transitions in the current distribution of driven diffusive channels. *J. Phys. A Math. Theor.* **51**(10), 105001 (2018). <http://stacks.iop.org/1751-8121/51/i=10/a=105001>
2. Barato, A.C., Chetrite, R.: A formal view on level 2.5 large deviations and fluctuation relations. *J. Stat. Phys.* **160**(5), 1154–1172 (2015). <https://doi.org/10.1007/s10955-015-1283-0>
3. Barré, J., Mukamel, D., Ruffo, S.: Inequivalence of ensembles in a system with long-range interactions. *Phys. Rev. Lett.* **87**, 030,601 (2001). <https://doi.org/10.1103/PhysRevLett.87.030601>

4. Bertini, L., Faggionato, A., Gabrielli, D.: From level 2.5 to level 2 large deviations for continuous time Markov chains. *Markov Process. Relat. Fields.* **20**(3), 545–562 (2012)
5. Biroli, G., Garrahan, J.P.: Perspective: the glass transition. *J. Chem. Phys.* **138**(12), 12A301 (2013). <https://doi.org/10.1063/1.4795539>
6. Bouchet, F., Dauxois, T., Mukamel, D., Ruffo, S.: Phase space gaps and ergodicity breaking in systems with long-range interactions. *Phys. Rev. E* **77**(1), 011,125 (2008). <https://doi.org/10.1103/PhysRevE.77.011125>
7. Bouchet, F., Gupta, S., Mukamel, D.: Thermodynamics and dynamics of systems with long-range interactions. *Phys. A Stat. Mech. Appl.* **389**(20), 4389–4405 (2010). <https://doi.org/10.1016/j.physa.2010.02.024>. <http://www.sciencedirect.com/science/article/pii/S0378437110001512>. Proceedings of the 12th International Summer School on Fundamental Problems in Statistical Physics
8. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
9. Budini, A.A., Turner, R.M., Garrahan, J.P.: Fluctuating observation time ensembles in the thermodynamics of trajectories. *J. Stat. Mech. Theory Exp.* **2014**(3), P03012 (2014). <http://stacks.iop.org/1742-5468/2014/i=3/a=P03012>
10. Callen, H.B.: *Thermodynamics and an Introduction to Thermostatistics*. Wiley (1985). <https://www.wiley.com/en-us/Thermodynamics+and+an+Introduction+to+Thermostatistics%2C+2nd+Edition-p-9780471862567>
11. Campa, A., Dauxois, T., Ruffo, S.: Statistical mechanics and dynamics of solvable models with long-range interactions. *Phys. Repo.* **480**(3–6), 57–159 (2009). <https://doi.org/10.1016/j.physrep.2009.07.001>. <http://www.sciencedirect.com/science/article/pii/S0370157309001586>
12. Chandler, D., Garrahan, J.P.: Dynamics on the way to forming glass: bubbles in space-time. *Annu. Rev. Phys. Chem.* **61**(1), 191–217 (2010). <https://doi.org/10.1146/annurev.physchem.040808.090405>. <https://doi.org/10.1146/annurev.physchem.040808.090405>. PMID: 20055676
13. Chetrite, R., Touchette, H.: Nonequilibrium microcanonical and canonical ensembles and their equivalence. *Phys. Rev. Lett.* **111**, 120,601 (2013). <https://doi.org/10.1103/PhysRevLett.111.120601>
14. Chetrite, R., Touchette, H.: Nonequilibrium Markov processes conditioned on large deviations. *Ann. Henri Poincaré* **16**(9), 2005–2057 (2015). <https://doi.org/10.1007/s00023-014-0375-8>
15. Costeniuc, M., Ellis, R.S., Touchette, H.: Complete analysis of phase transitions and ensemble equivalence for the Curie–Weiss–Potts model. *J. Math. Phys.* **46**(6), 063301 (2005). <https://doi.org/10.1063/1.1904507>
16. Csizár, I., Shields, P.C.: *Information Theory and Statistics: A Tutorial*. Now Foundations and Trends (2004). <https://doi.org/10.1561/01000000004>
17. Dinwoodie, I.H.: Identifying a large deviation rate function. *Ann. Probab.* **21**(1), 216–231 (1993). <https://doi.org/10.1214/aop/1176989402>
18. Dinwoodie, I.H., Zabell, S.L.: Large deviations for exchangeable random vectors. *Ann. Probab.* **20**(3), 1147–1166 (1992). <https://doi.org/10.1214/aop/1176989683>
19. Ellis, R.S., Touchette, H., Turkington, B.: Thermodynamic versus statistical nonequivalence of ensembles for the mean-field Blume–Emery–Griffiths model. *Phys. A Stat. Mech. Appl.* **335**(3–4), 518–538 (2004). <https://doi.org/10.1016/j.physa.2003.11.028>. <http://www.sciencedirect.com/science/article/pii/S0378437103011075>
20. Feller, W.: *An Introduction to Probability Theory and Its Applications*, vol. 2. Wiley, Hoboken (2008)
21. Ferré, G., Touchette, H.: Adaptive sampling of large deviations. *J. Stat. Phys.* (2018). <https://doi.org/10.1007/s10955-018-2108-8>
22. Freidlin, M.I., Wentzell, A.D.: *Random Perturbations of Dynamical Systems*. Grundlehren der mathematischen Wissenschaften. Springer, New York (1984). <https://doi.org/10.1007/978-3-642-25847-3>
23. Garrahan, J.P.: Classical stochastic dynamics and continuous matrix product states: gauge transformations, conditioned and driven processes, and equivalence of trajectory ensembles. *J. Stat. Mech. Theory Exp.* **2016**(7), 073208 (2016). <http://stacks.iop.org/1742-5468/2016/i=7/a=073208>
24. Garrahan, J.P., Jack, R.L., Lecomte, V., Pitard, E., van Duijvendijk, K., van Wijland, F.: Dynamical first-order phase transition in kinetically constrained models of glasses. *Phys. Rev. Lett.* **98**(19), 195,702 (2007). <https://doi.org/10.1103/PhysRevLett.98.195702>
25. Garrahan, J.P., Jack, R.L., Lecomte, V., Pitard, E., van Duijvendijk, K., van Wijland, F.: First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories. *J. Phys. A Math. Theor.* **42**(7), 075007 (2009). <http://stacks.iop.org/1751-8121/42/i=7/a=075007>
26. Glauber, R.J.: Time-dependant statistics of the Ising model. *J. Math. Phys.* **4**, 294 (1963). <https://doi.org/10.1063/1.1703954>
27. Goldenfeld, N.: *Lectures on Phase Transitions and the Renormalization Group*. Perseus Books Publishing, L.L.C., New York (1992)

28. Grafke, T., Grauer, R., Schäfer, T.: The instanton method and its numerical implementation in fluid mechanics. *J. Phys. A Math. Theor.* **48**(33), 333001 (2015). <http://stacks.iop.org/1751-8121/48/i=33/a=333001>
29. Hedges, L.O., Jack, R.L., Garrahan, J.P., Chandler, D.: Dynamic order–disorder in atomistic models of structural glass formers. *Science* **323**(5919), 1309–1313 (2009). <https://doi.org/10.1126/science.1166665>. <http://science.sciencemag.org/content/323/5919/1309>
30. Horn, R.A., Johnson, C.R.: *Matrix Analysis*, 2nd edn. Cambridge University Press, Cambridge (2012). <https://doi.org/10.1017/9781139020411>
31. Jack, R.L., Sollich, P.: Large deviations and ensembles of trajectories in stochastic models. *Prog. Theor. Phys. Suppl.* **184**, 304–317 (2010). <https://doi.org/10.1143/PTPS.184.304>. <http://ptps.oxfordjournals.org/content/184/304.abstract>
32. Jack, R.L., Garrahan, J.P., Chandler, D.: Space-time thermodynamics and subsystem observables in a kinetically constrained model of glassy materials. *J. Chem. Phys.* **125**(18), 184509 (2006). <https://doi.org/10.1063/1.2374885>
33. Keys, A.S., Chandler, D., Garrahan, J.P.: Using the s ensemble to probe glasses formed by cooling and aging. *Phys. Rev. E* **92**, 022304 (2015). <https://doi.org/10.1103/PhysRevE.92.022304>
34. Lawler, G.F., Sokal, A.D.: Bounds on the l^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger’s inequality. *Trans. Am. Math. Soc.* **309**(2), 557–580 (1988). <http://www.jstor.org/stable/2000925>
35. Lebowitz, J.L., Spohn, H.: A Gallavotti–Cohen-type symmetry in the large deviation functional for stochastic dynamics. *J. Stat. Phys.* **95**(1), 333–365 (1999). <https://doi.org/10.1023/A:1004589714161>
36. Lecomte, V., Appert-Rolland, C., van Wijland, F.: Thermodynamic formalism for systems with Markov dynamics. *J. Stat. Phys.* **127**(1), 51–106 (2007). <https://doi.org/10.1007/s10955-006-9254-0>
37. Levin, D.A., Yuval Peres, E.L.W.: *Markov Chains and Mixing Times*, vol. 58. American Mathematical Society, Providence (2008)
38. Maes, C., Netočný, K.: Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states. *EPL (Europhys. Lett.)* **82**(3), 30003 (2008). <http://stacks.iop.org/0295-5075/82/i=3/a=30003>
39. Merolle, M., Garrahan, J.P., Chandler, D.: Space-time thermodynamics of the glass transition. *Proc. Natl. Acad. Sci.* **102**(31), 10837–10840 (2005). <https://doi.org/10.1073/pnas.0504820102>. <http://www.pnas.org/content/102/31/10837>
40. Monthus, C.: Non-equilibrium steady states: maximization of the Shannon entropy associated with the distribution of dynamical trajectories in the presence of constraints. *J. Stat. Mech.* **2011**(03), P03008 (2011). <http://stacks.iop.org/1742-5468/2011/i=03/a=P03008>
41. Mukamel, D., Ruffo, S., Schreiber, N.: Breaking of ergodicity and long relaxation times in systems with long-range interactions. *Phys. Rev. Lett.* **95**, 240,604 (2005). <https://doi.org/10.1103/PhysRevLett.95.240604>
42. Nemoto, T.: Zon–Cohen singularity and negative inverse temperature in a trapped-particle limit. *Phys. Rev. E* **85**, 061,124 (2012). <https://doi.org/10.1103/PhysRevE.85.061124>
43. Nemoto, T., Sasa, S.: Computation of large deviation statistics via iterative measurement-and-feedback procedure. *Phys. Rev. Lett.* **112**(9), 090,602 (2014). <https://doi.org/10.1103/PhysRevLett.112.090602>
44. Nyawo, P.T., Touchette, H.: A minimal model of dynamical phase transition. *EPL (Europhys. Lett.)* **116**(5), 50009 (2016). <http://stacks.iop.org/0295-5075/116/i=5/a=50009>
45. Onsager, L., Machlup, S.: Fluctuations and irreversible processes. *Phys. Rev.* **91**(6), 1505–1512 (1953). <https://doi.org/10.1103/PhysRev.91.1505>
46. Pérez-Espigares, C., Carollo, F., Garrahan, J.P., Hurtado, P.I.: Dynamical criticality in driven systems: non-perturbative results, microscopic origin and direct observation. *ArXiv e-prints* (2018)
47. Speck, T.: Thermodynamic formalism and linear response theory for nonequilibrium steady states. *Phys. Rev. E* **94**, 022,131 (2016). <https://doi.org/10.1103/PhysRevE.94.022131>
48. Speck, T., Garrahan, J.P.: Space-time phase transitions in driven kinetically constrained latticemodels. *Eur. Phys. J. B* **79**(1), 1–6 (2011). <https://doi.org/10.1140/epjb/e2010-10800-x>
49. Szavits-Nossan, J., Evans, M.R.: Inequivalence of nonequilibrium path ensembles: the example of stochastic bridges. *J. Stat. Mech. Theory Exp.* **2015**(12), P12008 (2015). <http://stacks.iop.org/1742-5468/2015/i=12/a=P12008>
50. Touchette, H.: Equivalence and nonequivalence of the microcanonical and canonical ensembles : a large deviations study. Ph.D. thesis, PhD Thesis Department of Physics, McGill University (2003). http://digitool.Library.McGill.CA:80/R/-?func=dbin-jump-full&object_id=84851&silolibrary=GEN01
51. Touchette, H.: Simple spin models with non-concave entropies. *Am. J. Phys.* **76**, 26–30 (2008). <https://doi.org/10.1119/1.2794350>

52. Touchette, H.: The large deviation approach to statistical mechanics (2009). <https://doi.org/10.1016/j.physrep.2009.05.002>. <http://www.sciencedirect.com/science/article/pii/S0370157309001410>
53. Touchette, H.: Methods for calculating nonconcave entropies. *J. Stat. Mech.* **05008** (2010). <http://stacks.iop.org/1742-5468/2010/i=05/a=P05008>
54. Touchette, H., Ellis, R.S., Turkington, B.: An introduction to the thermodynamic and macrostate levels of nonequivalent ensembles. *Phys. A Stat. Mech. Appl.* **340**(1–3), 138–146 (2004). <https://doi.org/10.1016/j.physa.2004.03.088>. <http://www.sciencedirect.com/science/article/pii/S037843710400408X>. News and Expectations in Thermostatistics
55. Tsobgni Nyawo, P., Touchette, H.: Large deviations of the current for driven periodic diffusions. *Phys. Rev. E* **94**, 032,101 (2016). <https://doi.org/10.1103/PhysRevE.94.032101>
56. Van Kampen, N.: *Stochastic Processes in Physics and Chemistry* (Third Edition). Elsevier, Amsterdam (2007). <https://www.elsevier.com/books/stochastic-processes-in-physics-and-chemistry/van-kampen/978-0-444-52965-7>