Non equivalence of dynamical ensembles and emergent non ergodicity

Hadrien Vroylandt and Gatien Verley
Laboratoire de Physique Théorique (UMR8627), CNRS,
Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay, France
(Dated: July 2, 2018)

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.
I. 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 transform provides the appropriated potential according to the environmental constraints on the system (isolated, thermostatted, chemostated, etc) [1]. Ensemble equivalence holds in many cases, but it breaks down for systems with long range interactions between system constituents [2–4] or in the presence of a phase transition treated at the mean field level [5, 6]. The equivalence of equilibrium ensemble has been studied in detail [2–4, 7, 8], including in the framework of large deviation theory [9, 10].

More recently, different dynamical ensembles have been used by many authors to study stochastic processes [11–18] 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 [19], the driven, biased or tilted ensemble [13] or Esscher transform [20]. The last ensemble has been used to study glass transition [19, 21–28] or as a numerical tool to evaluate rare event probabilities [29, 30]. 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) [11, 12]. Whenever ensembles are equivalent, LDF and CGF encode the same information and are related by Legendre transform.

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 guarantees that the system can switch from any state to another in a reasonable time allowing 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 section II, 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 section III, 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. The non-ergodicity impacts the LDF since it may change according to the probability of the initial state. Beyond this case, 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. [31, 32] for equilibrium ensembles. Applying this to the Ising model, we notice that the non equivalence of equilibrium ensembles is tightly connected to the non equivalence of the dynamical ensembles. Since the 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 section IV, we prove that the time to relax from the initial condition grows exponentially with the number of spins. In practice, this means that for time 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: we adapt them for dynamical ensembles and illustrate them precisely using the Ising model.
II. MICROCANONICAL AND CANONICAL DYNAMICAL ENSEMBLES

In this section, we review for time-homogeneous jump processes some results on dynamical ensembles equivalence, see Refs. [11, 12] and references therein for more general Markov processes.

A. Definitions: stochastic process and dynamical ensembles

The Markov jump process $X_t$ takes values in the finite state space $\Omega$ 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

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

where $K(x, y)$ are the components of the Markov operator $K$, i.e. the probability per unit time of the transition from $y$ to $x$. 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 $\Omega$ 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 $O$. The mean value of this observable $O$ follows from

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

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

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

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. $O = o$ where $o$ is a possible value of the stochastic variable $O$ such that its probability verifies $P_T(O = o) > 0$. We write the path probability of these trajectories

$$P^o[x] = \frac{P[x, O = o]}{P_T(O_T = o)} \quad (4)$$

where $P^o[x]$ is the conditional probability of the path $[x]$ given that for this path $O = o$. The corresponding joint probability is denoted $P[x, O = o]$. In the microcanonical path ensemble, the observable $O$ 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 $O$ only. The path probability for this ensemble can be computed via tilting (or biasing) the process as follows:

$$P_o[x] = \frac{e^{T \gamma O} P[x]}{\langle e^{T \gamma O} \rangle}. \quad (5)$$

This path probability is normalized by construction.
B. Path ensemble equivalence

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

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

(6)

almost everywhere with respect to $P$. This means that $P$ and $Q$ are equal up to subexponential 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 $O \, [10]$, the LDF function $I(o)$ provides the rate of decay of the probability density function $P_T(O = o)$ in the limit $T \to +\infty$ such that

$$P_T(O = o) \underset{T \to +\infty}{\sim} e^{-TI(o)}$$

(7)

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

Assuming a large deviation principle, Chetrite and Touchette have shown that the convexity of $I(o)$ determines whether the microcanonical and canonical path ensembles are equivalent [11], in such a case $P^o$ and $P_g$ satisfy Eq. (6). They distinguished three cases:

(1) **Equivalence** If $I(o)$ is a stricly convex function at $o$, then there exists a unique $\gamma \in \mathbb{R}$ such that $P^o$ and $P_\gamma$ are equivalent.

(2) **Non equivalence** If $I(o)$ is a non-convex function at $o$, then there are no $\gamma$ such that $P^o$ and $P_\gamma$ are equivalent.

(3) **Partial equivalence** If $I(o)$ is a convex function but not strictly convex function at $o$, then numerous values of $o$ correspond to the same $\gamma$: It may correspond to linear parts in a convex function or to a point of a non-convex function at which the slope is not unique.

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

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

The CGF is the highest eigenvalue of the titled matrix $K_\gamma$ of elements [33]

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

(9)

The CGF is also the Legendre-Fenchel transform of the LDF

$$\phi(\gamma) = \max_o (\gamma o - I(o)).$$  

(10)

Conversely, we define $I^**(o)$ as the Legendre-Fenchel transform of the CGF

$$I^**(o) \equiv \max_\gamma (\gamma o - \phi(\gamma)) = \gamma(o)o - \phi(\gamma(o)).$$  

(11)

with $\gamma(o)$ the value of $\gamma$ realizing the maximum. The Gärtner-Ellis theorem states that if $\phi(\gamma)$ is a differentiable function then $I(o) = I^{**}(o)$. The properties of the Legendre-Fenchel transform implies that $I^{**}(o)$ is a convex function. The strict convexity of the LDF and the differentiability of the CGF are then dual conditions. Therefore we have the three following cases:

(1) **Equivalence** If $\phi(\gamma(o))$ is differentiable at $\gamma(o)$, then $I(o) = I^{**}(o)$ from the Gärtner-Ellis theorem, and moreover $o = \partial_\gamma \phi(\gamma)$. $I(o)$ is then a stricly convex function at $o$ and the equivalence holds.

(2) **Non equivalence** If $\phi(\gamma(o))$ is not differentiable at $\gamma(o)$ and $I(o) \neq I^{**}(o)$, then $I(o)$ is a non-convex function at $o$ and we have a non equivalence between the microcanonical and canonical ensembles.

(3) **Partial Equivalence** If $\phi(\gamma(o))$ is not differentiable at $\gamma(o)$ and $I(o) = I^{**}(o)$, then $I(o)$ is a convex function but not strictly convex function at $o$ and we have a partial equivalence between the microcanonical and canonical ensembles.

Hence, the non-equivalence of ensembles prevents to compute the LDF from the CGF since $I^{**}(o)$ is only the convex hull of $I(o)$. In Ref. [11], 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.
C. Proof of ensemble equivalence for ergodic process

The proof of ensemble equivalence amounts to determine the convexity of a LDF function. In this section, we show how the LDF for observable \( O \) 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 infer the LDF of the former from the LDF of the latters. This operation is called a contraction. For instance, the observable \( O \) of Eq. (48) is a function of the empirical occupation ratio \( R \) and the jump rate \( C \)

\[
O = O(R, C) = \sum_x f(x) R(x) + \sum_{x,y} C(x, y) g(x, y).
\]

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

\[
R(x) = \frac{1}{T} \int_0^T \delta(X_t = x) dt,
\]

where \( \delta \) 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

\[
C(x, y) = \frac{1}{T} \sum_{0 \leq t \leq T, \Delta X_t \neq 0} \delta(X_t+ = x) \delta(X_t- = y).
\]

The occupation ratios and the jump rates have a joint probability for \( R = r \) and \( C = c \) that satisfies a large deviation principle

\[
P_T(R = r, C = c) \sim e^{-TI(r,c)},
\]

defining the large deviation function \( I(r,c) \), also called the level 2.5 LDF[34–36]. Then, the contraction

\[
I(o) = \min_{r,c \text{ s.t. } O(r,c) = o} I(r,c)
\]

leads to the LDF for the observable \( O \). In general terms, a contraction corresponds to the minimization of a multi-variate LDF under a condition encoding how the stochastic variables are related [10].

Since ensemble equivalence for a stochastic variable relies on the convexity of the corresponding LDF, it is crucial to determine (i) whether the LDF \( I(r,c) \) is convex and (ii) whether the convexity can be inherited upon contraction. The following theorem appearing in Ref. [37] provides an answer to point (ii) when the new variable is additive:

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

\[
U(\alpha z_1 + (1-\alpha) z_2) = \alpha U(z_1) + (1-\alpha) U(z_2),
\]

then

\[
d(y) = \min_{x \in C, z \in C_y} h(x, z) \text{ with } C \text{ convex, and } C_y = \{z \mid U(z) = y\}
\]

is a convex function.

**Proof.** We consider \((x_1^*, z^*(y_1))\) the couple of variables realizing the minimum in Eq. (18) for \( y_1 \), and similarly \((x_2^*, z^*(y_2))\) for \( y_2 \). The convexity of \( C \) implies that \( \alpha x_1^* + (1-\alpha) x_2^* \in C \) when \( \alpha \in [0,1] \). Moreover, the additivity of \( U \) implies that \( \alpha z^*(y_1) + (1-\alpha) z^*(y_2) \in C_{\alpha y_1 + (1-\alpha) y_2} \). Hence, we have

\[
d(\alpha y_1 + (1-\alpha) y_2) = \min_{x \in C, z \in C_{\alpha y_1 + (1-\alpha) y_2}} h(x, z)
\]

\[
\leq h(\alpha x_1^* + (1-\alpha) x_2^*, \alpha z^*(y_1) + (1-\alpha) z^*(y_2)),
\]

\[
\leq \alpha h(x_1^*, z^*(y_1)) + (1-\alpha) h(x_2^*, z^*(y_2)),
\]

\[
\leq \alpha d(y_1) + (1-\alpha) d(y_2),
\]

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 \( y_1 \) and \( y_2 \). \( \square \)
We now address the point (i) about the convexity of the LDF \( I(r,c) \). This level 2.5 LDF is explicitly known for \textit{ergodic} Markov jump processes in continuous-time \cite{34-36}

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

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 \( c \) and \( Kr \)

\[
D(c||Kr) = \sum_{x,y\neq x} c(x,y)\ln \frac{c(x,y)}{K(x,y)r(y)}. \tag{20}
\]

This Kullback-Leibler divergence is convex as a consequence of the log-sum inequality (or Jensen’s inequality) \cite{38}. Using the previous theorem with \( z = (r,c) \), the convexity of the level 2.5 LDF and the additivity of \( O \) for both the occupation ratio and the jump rate, we conclude the contracted LDF \( I(\sigma) \) 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 consider the CGF as the highest eigenvalues of the titled matrix \( K_I \), defined in Eq.(65). From Perron-Frobenius theorem for irreducible finite size matrices like \( K_I \), 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 titled 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 \( O \) may only happen when the Markov operator is reducible or of infinite dimension. Then, the LDF of the variable \( O \) 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 channel transitions.

III. EXPLICIT LDF CALCULATION FOR NON-ERGODIC SYSTEMS

In this section, we adapt the method developed in Ref. \cite{31} for equilibrium ensembles to compute non-convex LDF from their corresponding generating function. First, we explain how to compute a non-convex LDF from the propagator of the generating function for a simple non-ergodic system. We then apply the same method to infinite range Ising model for which the non-ergodicity emerges in the thermodynamic limit.

A. 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 \((3,4)\). The four state system evolves according to the master equation

\[
\frac{d}{dt} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -2 & 2 \\ 0 & 0 & 2 & -2 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix}. \tag{21}
\]

The Markov operator is reducible yielding to a non ergodic process. 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. (48) when \( f(x) = 0 \) and \( g(x,y) = 1 \) for all \((x,y)\). By definition of the Markov operator in the right hand side 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 \( p_{i,n} \) to start in state \( n \). The probability of having \( aT \) jumps after a time \( T \) is

\[
Pr(A = a) = (p_{i,1} + p_{i,2})(T)^{aT}e^{-T}(aT)! + (p_{i,3} + p_{i,4})(2T)^{aT}e^{-2T}(aT)! \tag{22}
\]
For $T$ large enough, $P_T(A = a)$ will be bimodal, and the microcanonical LDF reads

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

where we have used Stirling formula and chosen the minimum of the two LDF corresponding to each Poisson distribution. The above microcanonical 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 $G(x_1, x_1, \gamma) = \langle e^{T\gamma A} \rangle_{x_1, x_1}$, where the subscripts $x_1$ and $x_i$ denote respectively the initial and final states of the trajectories appearing in the average. Using standard approach [10, 39], the propagator is obtained from the biased operator

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

as the components of its matrix exponential $[\exp(TK_\gamma)] (x_1, x_i) = G(x_1, x_i, \gamma)$. Using eigenvalues and eigenvectors decomposition of $K_\gamma$, we make explicit the matrix exponential

$$\exp [TK_\gamma] = e^{T(-1+e^\gamma)} \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\
 0 & 1 & 0 & 0 \\
 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 1 \end{pmatrix} \cdot (0 \quad 0 \quad 1 \quad 1)$$

$$+ e^{-T(1+e^\gamma)} \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\
 0 & 1 & 0 & 0 \\
 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 1 \end{pmatrix} \cdot (0 \quad 0 \quad 1 \quad -1), \quad (25)$$

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. We remark also that the above propagator should be norm conserving when $\gamma = 0$, but the two terms in the second line of Eq. (25) do not fulfill this requirement. We do not consider then as physical and keep only the two first terms in Eq. (25). Then, the activity LDF is recovered from this propagator by first summing over the initial and final states

$$\langle e^{T\gamma A} \rangle = \sum_{x_1, x_i} G(x_1, x_i, \gamma) p_t(x_i), \quad (26)$$

second, applying an inverse Laplace transform, and finally by taking the limit $T \to \infty$. 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

$$\langle e^{T\gamma A} \rangle = (p_{i,1} + p_{i,2}) e^{T(-1+e^\gamma)} + (p_{i,3} + p_{i,4}) e^{2T(-1+e^\gamma)}. \quad (27)$$

Its 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

$$\phi(\gamma) = \lim_{T \to +\infty} \frac{1}{T} \ln \sum_{x_1, x_i} G(x_1, x_i, \gamma) p_t(x_i) = \begin{cases} e^{\gamma} - 1 & \text{if } \gamma < 0 \\
 2(e^{\gamma} - 1) & \text{if } \gamma \geq 0 \end{cases}, \quad (28)$$

as long as $p_t(x) > 0$ for all $x$. Noticing that the limit $T \to +\infty$ enables to use a saddle-point method to approximate the inverse Laplace transform into a Legendre-Fenchel transform, the asymptotic probability of activity follows from
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 transform of the two highest eigenvalues, canonical LDF (cross) and microcanonical LDF (squares).

its corresponding LDF

\[
I_c(a) = \max_{\gamma} [a\gamma - \phi_{x_i}(\gamma)] = \begin{cases} 
  a \ln a - a + 1 & \text{if } a < 1 \\
  0 & \text{if } a \in [1, 2] \\
  a \ln \frac{a}{2} - a + 2 & \text{if } a > 2
\end{cases}, \tag{29}
\]

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

Alternatively, one may obtain the microcanonical 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

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

The Legendre-Fenchel conjugate of these two branches associated to different initial states are precisely the two branches of the microcanonical 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

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

explaining why a minimum on the branches of Eq. (31) appears in the final microcanonical 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 microcanonical 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.
B. 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 section II. In practice, this amounts to express the inverse Laplace transform as a Legendre-Fenchel transform, using the saddle point method and taking care of the initial condition appropriately.

By definition, for \( b \in \mathbb{R} \), the microcanonical LDF for \( O \) writes

\[
I_{mc}(\phi) \equiv \lim_{T \to +\infty} -\frac{1}{T} \ln \int_{b-\infty}^{b+i\infty} d\gamma e^{-T\gamma \phi} \sum_{x_i, x_i} G(x_i, x_i, \gamma) p_i(x_i),
\]

in term of the initial probability \( p_i(x_i) \) and of the propagator \( G(x_i, x_i, \gamma) = \langle e^{T\gamma O} \rangle_{x_i, x_i} \) that generates the moments of \( O \) under given initial and final conditions. The argument of the logarithm in Eq. (33) is exactly the probability distribution function of \( O \). 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 section III A):

\[
I_{mc}(\phi) = \min_{x_i, x_i} \lim_{T \to +\infty} -\frac{1}{T} \ln \int_{b-\infty}^{b+i\infty} d\gamma e^{-T\gamma \phi} G(x_i, x_i, \gamma) p_i(x_i).
\]

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

\[
I_{mc}(\phi) = \min \max_{x_i, x_i} \left( \gamma \phi - \lim_{T \to +\infty} \frac{1}{T} \ln G(x_i, x_i, \gamma) p_i(x_i) \right).
\]

Alternatively, the convex hull of \( I_{mc}(\phi) \) is the Legendre-Fenchel conjugate of the CGF for the observable \( O \), namely

\[
I_c(\gamma) \equiv \max_{\phi} \left( \gamma \phi - \lim_{T \to +\infty} \frac{1}{T} \ln \sum_{x_i, x_i} G(x_i, x_i, \gamma) p_i(x_i) \right).
\]

As seen in section III A, the propagator \( G \) may have an eigendecomposition for which the terms contributing to the CGF in the limit \( T \to \infty \) depend on the initial or final condition. Hence, a maximum on \( x_i \) and \( x_i \) must appear

\[
I_c(\gamma) = \max_{\phi} \left( \gamma \phi - \max_{x_i, x_i} \lim_{T \to +\infty} \frac{1}{T} \ln G(x_i, x_i, \gamma) p_i(x_i) \right),
\]

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

\[
I_c(\gamma) = \max \min_{x_i, x_i} \left( \gamma \phi - \lim_{T \to +\infty} \frac{1}{T} \ln G(x_i, x_i, \gamma) p_i(x_i) \right).
\]

In the end, the difference between the LDF \( I_{mc}(\phi) \) and its convex hull \( I_c(\gamma) \) comes from the non commutation of \( \max_{\phi} \) and \( \min_{x_i, x_i} \) 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 the biased matrix of 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 microcanonical LDF, one needs all the branches corresponding to each eigenvalue that becomes the highest eigenvalue for at least one \( \gamma \). Only those branches matters, and other eigenvalues will not contribute, so as can be understood from the following argument: Be two eigenvalues \( \phi_1(\gamma) \) and \( \phi_2(\gamma) \) such that \( \phi_1(\gamma) > \phi_2(\gamma) \) for all \( \gamma \). Since \( \gamma \phi - \phi_1(\gamma) < \gamma \phi - \phi_2(\gamma) \), we have

\[
I_1(\phi) = \min_{\gamma} \{ \gamma \phi - \phi_1(\gamma) \} < I_2(\phi) = \min_{\gamma} \{ \gamma \phi - \phi_2(\gamma) \}.
\]

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. We assume that the initial probability cannot be zero in all states of a subsytem, 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 transform 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.

C. Mean-field Ising model

Using the results of sec. III B, 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 thermodynamics limit. 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 CGF and both the canonical and microcanonical LDFs.

1. Model description and thermodynamics limit

We consider the fully connected Ising model made of \( N \) interacting spins \( \{ s \} = (s_1, \ldots, s_N) \). Each spin \( s_i \) can hop between states \( +1 \) and \( -1 \) by exchanging heat with a thermostat at inverse temperature \( \beta \). 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

\[
U(n) \equiv -\frac{V}{N} \sum_{1 \leq i \leq j \leq N} s_is_j - E \sum_{1 \leq i \leq N} s_i = -n^2 \frac{V}{2N} - nE
\]  

and of the internal entropy

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

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

\[
K(n + 2\epsilon, n) \equiv \Gamma (N - cn) e^{\beta ((2\epsilon n + 2) V/N + 2\epsilon E)}. 
\]

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

\[
\ln \frac{K(n + 2\epsilon, n)}{K(n, n + 2\epsilon)} = -\beta (F(n + 2\epsilon) - F(n)),
\]

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

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

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

\[
\pi(n) \equiv \frac{1}{Z_{\text{eq}}} e^{-F(n)} \quad \text{with} \quad Z_{\text{eq}} \equiv \sum_{n=0}^{N} e^{-F(n)}.
\]
The time-averaged stochastic magnetization is obtained from Eq. (48) by choosing the state dependent function \( f(n) = n/N \) and a vanishing function \( g(n, n') = 0 \) for all \((n, n')\):

\[
M \equiv \frac{1}{NT} \int_0^T dt n(t),
\]

(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

\[
\langle M \rangle = \frac{1}{N} \sum_{n=0}^N n \pi(n).
\]

(47)

The activity is obtained from Eq. (48) by choosing \( g(n, n') = 1/N \) and \( f(n) = 0 \) for all \((n, n')\):

\[
A \equiv \frac{1}{NT} \sum_{0 \leq t \leq T: \Delta n \neq 0} 1.
\]

(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

\[
\langle A \rangle = \frac{1}{N} \sum_{n=0}^N \sum_{\epsilon = \pm 1} \pi(n) K_{n+2\epsilon,n}.
\]

(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

\[
\lim_{N \to \infty} [U(n + 2\epsilon) - U(n)] = 2\epsilon (Vx + E)
\]

(50)

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

\[
J_\epsilon(x) \equiv \lim_{N \to +\infty} \frac{K_{xN + 2\epsilon,xN}}{N} = (1 - \epsilon x) e^{[Vx + E]}.
\]

(51)

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

\[
\langle \dot{x} \rangle = \sum_{\epsilon = \pm 1} \epsilon J_\epsilon (\langle x \rangle).
\]

(52)

The steady state solution of this equation is the mean field magnetization \( x = m_{\text{MF}} \) verifying

\[
J_- (m_{\text{MF}}) = J_+ (m_{\text{MF}}).
\]

(53)

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

\[
m_{\text{MF}} = \tanh (V m_{\text{MF}} + E).
\]

(54)

The MF activity follows from the mean-field magnetization:

\[
a_{\text{MF}} = J_- (m_{\text{MF}}) + J_+ (m_{\text{MF}}) = [1 - m_{\text{MF}}] e^{V m_{\text{MF}} + E} + [1 + m_{\text{MF}}] e^{-V m_{\text{MF}} - E}.
\]

(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 studied ferromagnetic transition in the Ising model. This bifurcation also affects the system activity as shown on the Fig. 2b and as expected from Eq. (55) since \( a_{\text{MF}} \) is a function of \( m_{\text{MF}} \). 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.
2. Propagator of the generating function for magnetization and activity

Like in the four state model of section III A, we look for the propagator of the generating function for the activity and magnetization:

\[ G(x_f, x_i, \kappa, \gamma) = \langle e^{N(T(\kappa m + \gamma a))_{x_i, x_f}} \rangle, \]

(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

\[ G(x_f, x_i, \kappa, \gamma) \simeq N \to +\infty \exp \left( N \int_0^T dt' L(x_{t'}, \dot{x}_{t'}, \kappa, \gamma) \right), \]

(57)

where \( L \) is the Lagrangian

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

(58)

with

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

(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

\[ \frac{\partial L}{\partial x} = \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{x}} \right). \]

(60)

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

As explained in sec. III B, 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

\[ \frac{\partial L}{\partial x}(x^*, 0, \kappa, \gamma) = 0. \]

(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 \((\kappa, \gamma)\). Following sec. III B, we can restrain the extremization over initial and final conditions to these stationary trajectories.
We provide an 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. 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

$$\phi(\kappa, \gamma) = \lim_{N,T \to \infty} \frac{1}{NT} \ln \left(e^{NT\kappa m + NT\gamma a}\right),$$

$$= \max_{x_{i}, x_{f}, N,T \to \infty} \frac{1}{NT} \ln G(x_{i}, x_{f}, \kappa, \gamma)p_{i}(x_{i}). \quad (62)$$

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

$$\phi(\kappa, \gamma) = \max_{x} \mathcal{L}(x, 0, \kappa, \gamma),$$

$$= \sqrt{\varphi(x^*(\kappa, \gamma), \gamma)} - \sum_{\epsilon=\pm1, \nu=1,2} J_{\nu}^*(x^*(\kappa, \gamma)) + \kappa x^*(\kappa, \gamma). \quad (63)$$

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 3d and 3c respectively. Cross-sections of the CGF in the plane $\kappa = 0$ and $\gamma = 0$ are shown in Fig 3a-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 $(\gamma, \kappa)$ 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

$$K_{\gamma, \kappa}(x, y) = \begin{cases} k(x, y)e^{\gamma/N} & \text{if } x \neq y \\ k(x, x) + \kappa x/N & \text{if } x = y \end{cases}, \quad (64)$$

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

4. Canonical LDF of magnetization and activity

The canonical LDF $I_{c}(m, a)$ shown in Fig. 4c-d is the Legendre-Fenchel conjugate of the CGF

$$I_{c}(m, a) = \max_{\kappa, \gamma} \left\{ \kappa m + \gamma a - \phi(\kappa, \gamma) \right\}. \quad (66)$$

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 LDF 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 LDF for activity $I_{c}(a)$ and magnetization $I_{c}(m)$ defined by

$$I_{c}(m) \equiv \min_{a} I_{c}(m, a) = \max_{\kappa} \left\{ \kappa m - \phi(\kappa, 0) \right\}, \quad (67)$$

$$I_{c}(a) \equiv \min_{m} I_{c}(m, a) = \max_{\gamma} \left\{ \gamma a - \phi(0, \gamma) \right\}, \quad (68)$$

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 3d and 3c respectively. Cross-sections of the CGF in the plane $\kappa = 0$ and $\gamma = 0$ are shown in Fig 3a-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 $(\gamma, \kappa)$ 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

$$K_{\gamma, \kappa}(x, y) = \begin{cases} k(x, y)e^{\gamma/N} & \text{if } x \neq y \\ k(x, x) + \kappa x/N & \text{if } x = y \end{cases}, \quad (64)$$

for systems with $N = 10$ and $N = 25$ spins. As explained in Sec II C, the CGFs of finite size systems are everywhere differentiable. Then, they only approach gradually the non-differentiable CGF when $N \to \infty$.
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). Inserts: 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 \( E = 0 \).

are shown in Fig. 4a-b, together with the LDFs for the finite size systems obtained from the Legendre transform of their corresponding CGFs in Fig. 3. 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.

5. Microcanonical LDF of magnetization and activity

As explain in sec. III B, the microcanonical LDF follows from the Legendre-Fenchel transform of the elements of the propagator of the generating function. Therefore Eq. (35) yields

\[
I_{mc}(m, a) = \min_{x_1, x_i} \max_{\kappa, \gamma} \left[ \kappa m + \gamma a - \lim_{N,T \to \infty} \frac{1}{NT} \ln G(x_1, x_i, \kappa, \gamma) p_i(x) \right].
\]  

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

\[
I_{x^*}(m, a) = \max_{\kappa, \gamma} \left\{ \kappa m + \gamma a - \mathcal{L}(x^*(\kappa, \gamma), 0, \kappa, \gamma) \right\}.
\]

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 microcanonical LDF

\[
I_{mc}(m, a) = \min_{x \in \{x^*\}} I_x(m, a).
\]

From the fact that \( I_c \) is the convex hull of \( I_{mc} \), we get the following inequality between the two LDFs:

\[
I_{mc}(m, a) \geq I_c(m, a).
\]
The microcanonical LDF $I_{mc}(m, a)$ is shown after the bifurcation on Fig. 5c. We also provide in Fig. 5a-b the partial LDF for activity and magnetization (after contraction) and their convex hulls. As expected, the microcanonical LDF is not convex: the ensemble equivalence does not hold (in a specific interval of magnetization and activity) for our model in the thermodynamics limit, due to lack of ergodicity.

Comparing the canonical LDF obtained from Eqs. (67) and the microcanonical LDF, we notice that the former is as expected the convex hull of the latter.

IV. 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 $K$. Second, we give an estimate of the mixing time for the Ising model and prove that it diverges when $N \to \infty$. 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 regime 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 there exists two different stable magnetizations 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.

A. Gap at finite size

We come back to the continuous-time process of Eq. (1) of Markov operator \( \mathbf{K} \). We assume that \( \mathbf{K} \) is diagonalizable. The eigenvalues are \( \{\lambda_i\}_{i=1,N} \), where \( \lambda_1 = 0 \) and all other eigenvalues are negative. Its set of eigenvectors is \( \{l_i, r_i\}_{i=1,N} \) where \( l_1 \) is an uniform vector and \( r_1 = \pi \) is the stationnary probability. The spectral gap \( \Delta \lambda \) of the operator \( \mathbf{k} \) is the difference between the largest eigenvalue \( \lambda_1 \) and the real part of its second eigenvalues \( \lambda_2 \).

Considering the initial probability \( p_0 \), the probability

\[
p(x,t) = (e^{tk}p_i)(x) = \pi(x) + \sum_i e^{t\lambda_i} r_i(x) \sum_x l_i(x)p_i(x)
\]

is a formal solution to Eq. (1). Since the spectral gap \( \Delta \lambda \) of the operator \( \mathbf{K} \) is positive, the probability \( p(x,t) \) converges towards \( \pi \). The mixing time \( t_{\text{mix}}(\varepsilon) \) is used to quantify the time that the system takes to relax to the stationary probability \cite{40}. 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 \( \varepsilon \) of the stationary probability. Formally, the mixing time is

\[
t_{\text{mix}}(\varepsilon) \equiv \inf \left\{ t \geq 0 : \max_{p_i} \| e^{t\mathbf{K}}p_i - \pi \|_{TV} \leq \varepsilon \right\}.
\]
where the distance is the total variation distance defined as \( \|u\|_{TV} = \sup_{A} u(A) \). A common choice for \( \varepsilon \) 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 is the initial probability. In particular, an infinite mixing time is a feature of non-ergodic systems. When considering large deviations statistics, we must formally take long time limit. We emphasize that for times longer than the mixing time, the large deviation statistics is approximately valid.

The mixing time of Eq. (74) is a maximization over all initial probability, therefore looking at \( \|e^{tK}p_{i} - \pi\|_{TV} \) for an uniform initial probability \( p_{i} = 1 \) underestimates the mixing time. We plot on the 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. (76), we expect the mixing time to be connected with the spectral gap. Indeed we have for any Markov process [40]

\[
\frac{1}{\Delta \lambda} \leq t_{\text{mix}} \leq -\frac{\log p_{i}^{\text{min}}}{\Delta \lambda} \tag{75}
\]

where \( p_{i}^{\text{min}} = \min_{x} p_{i}(x) \) and \( \Delta \lambda \) is the spectral gap. Second, at very short time, 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 dependences 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 6b we show the empirical density probability for activity and magnetization for a system of \( N = 60 \) spins and for different durations \( T \), with \( t_{\text{mix}} \approx 100 \). And on Fig. 6d, we plot the empirical probability density of activity for a system size \( N = 200 \) where \( t_{\text{mix}} \approx 10^{2} \).

For \( T \gg t_{\text{mix}} \), the long time probabilities of magnetization and activity are those of an ergodic systems (the system has enough time to switch between MF states). The asymptotic probability is then given by the convex LDF of sec. III C 4. The convergence toward the plateau for these LDFs is discussed in the next section. Notice that the added small magnetic field is responsible of the unimodality of the probability density functions, see Fig. 6a and 6b 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 \( |\lambda_{0} - \lambda_{3}|^{-1} < T \ll t_{\text{mix}} \) 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 microcanonical LDF of sec. III C 5.

**B. 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 [41]. 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 \( D_{\sqrt{\pi}}(n, n) = \sqrt{\pi_{n}} \), then the matrix \( D_{\sqrt{\pi}}MD_{\sqrt{\pi}}^{-1} \) is symmetric if the Markov matrix \( K \) respects detailed balance. We introduce the Cheeger constant as

\[
\Phi = \inf_{\Xi \subseteq \Omega, 0 < \pi(\Xi) \leq 1/2} \frac{Q(\Xi^{c}, \Xi)}{\pi(\Xi)} \tag{77}
\]

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

\[
Q(\Xi^{c}, \Xi) = \sum_{x, y | x \in \Xi, y \in \Xi^{c}} \pi(x)K(y, x) \tag{78}
\]

is the sum of probability flow from the subset \( \Xi \) to the complementary of \( \Xi \), denoted \( \Xi^{c} \). From Ref. [41]

\[
\Delta \lambda \leq 2\Phi \tag{79}
\]
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 $e^{tK}p_i$ and $\pi$ in 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 $15 \times 10^4$ 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 $15 \times 10^4$ 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: $V = 1.7$, $E = 0.2$.

Computing the Cheeger constant is difficult in general, but it is possible to bound it. For a subset $\Xi$, the probability flow from $\Xi$ to $\Xi'$ is bounded by

$$Q(\Xi', \Xi) \leq N_b \pi(\hat{x}) K(\hat{y}, \hat{x})$$  \hspace{1cm} (80)

where $N_b$ is the number of edge connecting $\Xi$ and $\Xi'$ and

$$(\hat{x}, \hat{y}) = \arg \max_{x \in \Xi, y \in \Xi'} \pi(x) K(y, x),$$  \hspace{1cm} (81)

the edge supporting the biggest probability flow from $\Xi$ to $\Xi'$. Denoting then $\hat{x}$ the most probable state in $\Xi$, we have

$$\Phi \leq N_b \pi(\hat{x}) K(\hat{y}, \hat{x}) \pi(\hat{x}).$$  \hspace{1cm} (82)

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

$$\pi(x) \simeq_{N \to +\infty} e^{-N\mathcal{I}(x)}.$$  \hspace{1cm} (83)

We consider a connected subset of state $\Xi$ such that the state $x^{MF}$ with $\mathcal{I}(x^{MF}) = 0$ is not in $\Xi$. For large enough $N$, the probability of $\Xi$ is surely less than $1/2$. Using the bound (82) on the Cheeger constant, we find a large deviation
estimate as
\[ \Phi \leq \frac{N_b \pi(\hat{x}) K(\hat{y}, \hat{x})}{\pi(\hat{x})} \approx N_b K(\hat{y}, \hat{x}) e^{N(I(\hat{x}) - I(\hat{x}))}. \] (84)

Therefore, if it exists \( \Xi \) such that the LDF \( I(x) \) has a local minimum \( \hat{x} \in \Xi \) that is not a global minimum, we will have \( I(\hat{x}) - I(\hat{x}) < 0 \). 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 \( N \to +\infty \). Hence, the mixing time diverges with the system size, and the divergence is even exponential.

In our case, we consider the subset of state \( \Xi^- = \{ x | x \leq 0 \} \), if \( E \geq 0 \) it has a stationary probability less than \( 1/2 \), otherwise if \( E < 0 \), we consider the subset \( \Xi^+ = \{ x | x \geq 0 \} \). 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 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.

C. Mixing time as a criteria 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 canonical LDF. 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.

V. 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 an emergent non ergodicity. In both cases we have obtained LDFs with an explicit dependency in the initial condition, the so-called partial LDF. From this partial LDF, the microcanonical LDF follows from minimizing on the initial condition. For systems with emergent ergodicity breaking we have discussed the physical meaning of both the microcanonical (non-convex) and canonical (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 corresponds 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 observables 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 [42–44]. In this paper, we did not take instantons into account for the minimization of Eq. (62). However, taking them into account does not modify the canonical LDF, but gives instead another LDF at a different large deviation speed [45].

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 [46, 47].

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.

ACKNOWLEDGEMENT

We acknowledge Massimiliano Esposito for his advice on this project that started when GV was a post-doctoral fellow in his research team.

[1] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, 1985).
[2] Julien Barré, David Mukamel, and Stefano Ruzzo, “Inequivalence of ensembles in a system with long-range interactions,” Phys. Rev. Lett. 87, 030601 (2001).
[3] Alessandro Campa, Thierry Dauxois, and Stefano Ruzzo, “Statistical mechanics and dynamics of solvable models with long-range interactions,” Physics Reports 480, 57–159 (2009).
[4] Freddy Bouchet, Shamik Gupta, and David Mukamel, “Thermodynamics and dynamics of systems with long-range interactions,” Physica A: Statistical Mechanics and its Applications 389, 4389–4405 (2010), proceedings of the 12th International Summer School on Fundamental Problems in Statistical Physics.
[5] Richard S. Ellis, Hugo Touchette, and Bruce Turkington, “Thermodynamic versus statistical nonequivalence of ensembles for the mean-field blume-emery-griffiths model,” Physica A: Statistical Mechanics and its Applications 335, 518–538 (2004).

[6] Marius Costeniuc, Richard S. Ellis, and Hugo Touchette, “Complete analysis of phase transitions and ensemble equivalence for the curie-weiss-potts model,” J. Math. Phys. 46, 063301.2005. (2005), cond-mat/0410744.

[7] Hugo Touchette, Richard S. Ellis, and Bruce Turkington, “An introduction to the thermodynamic and macrostate levels of nonequivalent ensembles,” Physica A: Statistical Mechanics and its Applications 340, 138–146 (2004), news and Expectations in Thermostatistics.

[8] Freddy Bouchet, Thierry Dauxois, David Mukamel, and Stefano Ruffo, “Phase space gaps and ergodicity breaking in systems with long-range interactions,” Phys. Rev. E 77, 011125 (2008).

[9] H. Touchette, Equivalence and nonequivalence of the microcanonical and canonical ensembles : a large deviations study, Ph.D. thesis, PhD Thesis Department of Physics, McGill University (2003).

[10] Hugo Touchette, “The large deviation approach to statistical mechanics,” Physics Reports 478, 1–69 (2009).

[11] Raphaël Chetrite and Hugo Touchette, “Nonequilibrium markov processes conditioned on large deviations,” Annales Henri Poincaré 16, 2005–2057 (2015).

[12] Raphaël Chetrite and Hugo Touchette, “Nonequilibrium microcanonical and canonical ensembles and their equivalence,” Phys. Rev. Lett. 111, 120601 (2013).

[13] Robert L. Jack and Peter Sollich, “Large deviations and ensembles of trajectories in stochastic models,” Progress of Theoretical Physics Supplement 184, 304–317 (2010), http://ptps.oxfordjournals.org/content/184/304.full.pdf+html.

[14] V. Lecomte, C. Appert-Rolland, and F. van Wijland, “Thermodynamic formalism for systems with markov dynamics,” Journal of Statistical Physics 127, 51–106 (2007).

[15] Thomas Speck, “Thermodynamic formalism and linear response theory for nonequilibrium steady states,” Phys. Rev. E 94, 022131 (2016), 1601.03540.

[16] Juan P. Garrahan, “Classical stochastic dynamics and continuous matrix product states: gauge transformations, conditioned and driven processes, and equivalence of trajectory ensembles,” Journal of Statistical Mechanics: Theory and Experiment 2016, 073208 (2016).

[17] Cécile Monthus, “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, P03008 (2011).

[18] Adrián A Budini, Robert M Turner, and Juan P Garrahan, “Fluctuating observation time ensembles in the thermodynamics of trajectories,” Journal of Statistical Mechanics: Theory and Experiment 2014, P03012 (2014).

[19] Juan P Garrahan, Robert L Jack, Vivien Lecomte, Estelle Pitard, Kristina van Duijvendijk, and Frédéric van Wijland, “First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories,” Journal of Physics A: Mathematical and Theoretical 42, 075007 (2009).

[20] William Feller, An introduction to probability theory and its applications, Vol. 2 (John Wiley & Sons, 2008).

[21] Aaron S. Keys, David Chandler, and Juan P. Garrahan, “Using the $s$ ensemble to probe glasses formed by cooling and aging,” Phys. Rev. E 92, 022304 (2015).

[22] Giulio Biroli and Juan P. Garrahan, “Perspective: The glass transition,” The Journal of Chemical Physics 138, 12A301 (2013), https://doi.org/10.1063/1.4795539.

[23] Mauro Merolle, Juan P. Garrahan, and David Chandler, “Space-time thermodynamics of the glass transition,” Proceedings of the National Academy of Sciences 102, 10837–10840 (2005), http://www.pnas.org/content/102/31/10837.full.pdf.

[24] Robert L. Jack, Juan P. Garrahan, and David Chandler, “Space-time thermodynamics and subsystem observables in a kinetically constrained model of glassy materials,” The Journal of Chemical Physics 125, 184509 (2006), https://doi.org/10.1063/1.2374885.

[25] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, “Dynamical first-order phase transition in kinetically constrained models of glasses,” Phys. Rev. Lett. 98, 195702 (2007).

[26] Lester O. Hedges, Robert L. Jack, Juan P. Garrahan, and David Chandler, “Dynamic order-disorder in atomistic models of structural glass formers,” Science 323, 1309–1313 (2009), http://science.sciencemag.org/content/323/5919/1309.full.pdf.

[27] David Chandler and Juan P. Garrahan, “Dynamics on the way to forming glass: Bubbles in space-time,” Annual Review of Physical Chemistry 61, 191–217 (2010), pMID: 20055676, https://doi.org/10.1146/annurev.physchem.040808.090405.

[28] T. Speck and J. P. Garrahan, “Space-time phase transitions in driven kinetically constrained lattice models,” The European Physical Journal B 79, 1–6 (2011).

[29] Takahiro Nemoto and Shin-ichi Sasa, “Computation of large deviation statistics via iterative measurement-and-feedback procedure,” Phys. Rev. Lett. 112, 090602 (2014).

[30] Grégoire Ferré and Hugo Touchette, “Adaptive sampling of large deviations,” 1803.1117v1.

[31] Hugo Touchette, “Methods for calculating nonconcave entropies,” J. Stat. Mech. P 05008 (2010), 1003.0382.

[32] Hugo Touchette, “Simple spin models with non-concave entropies,” Am. J. Phys. 76., 26–30,2008 (2008), cond-mat/0504202.

[33] Joel L. Lebowitz and Herbert Spohn, “A gallavotti–cohen-type symmetry in the large deviation functional for stochastic dynamics,” Journal of Statistical Physics 95, 333–365 (1999).

[34] C. Maes and K. Netočný, “Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states,” EPL (Europhysics Letters) 82, 30003 (2008).
Appendix A: Derivation of the propagator for generating function

We look for an expression of the propagator of the generating function $G(x_1, x_2, \gamma, \gamma') = \langle e^{N(x_1 \cdot x_2)} \rangle_{x_1, x_2}$ as a path integral over all trajectories for the system state $n$. We use the discrete time $t_k = kdt$ corresponding to the $k$’s time step of duration $d t$, with the final time $T = t_L = Ldt$. We write the system state $n_k = n_k$ 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, does n’t jump or jumps from down to up. The sum over all paths for a given initial condition is

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

(A1)

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

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

(A2)

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

$$G(n_1, n_L, \gamma, \gamma') = \sum_{[n_k, \epsilon_k]} \left[ \prod_{k=1}^{L} p(n_k|n_{k-1}) \delta(n_k - n_{k-1} - 2\epsilon_k) \right] e^{\kappa \left( \sum_{k=1}^{L} n_{k-1} dt + \gamma \left( \sum_{k=1}^{L} \epsilon_k \right) \right)}$$

(A3)

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

$$\delta(n_k - n_{k-1} - 2\epsilon_k) = \int_{ib_k - \infty}^{ib_k + \infty} \frac{d p_k}{2\pi} e^{i p k (n_k - n_{k-1} - 2\epsilon_k)}$$

(A4)

for every integer $k$ between 1 and $L$ to get

$$G(n_1, n_L, \gamma, \gamma') = \sum_{[n_k, \epsilon_k]} \left[ \prod_{k=1}^{L} \frac{d p_k}{2\pi} \right] \left[ \prod_{k=1}^{L} p(n_k|n_{k-1}) e^{i \frac{p_k (n_k - n_{k-1} - 2\epsilon_k)}{2}} + \gamma n_{k-1} dt \right]$$

(A5)
In the next step, we sum over \([\epsilon_k]\) to obtain

\[
G(n_i, n_f, \kappa, \gamma) = \sum_{[n_k]} \int \left( \prod_{k=1}^{L} \frac{dp_k}{2\pi} e^{ip_k(n_k - n_{k-1})} \right) \prod_{k=1}^{L} p(n_{k-1}|n_{k-1}) + \sum_{\epsilon = \pm 1} p(n_{k-1} + \epsilon|n_{k-1})e^{-2ip_k\epsilon + \gamma} e^{\kappa n_{k-1} - idt}. \quad (A6)
\]

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

\[
G(n_i, n_f, \kappa, \gamma) = \sum_{[n_k]} \int \left( \prod_{k=1}^{L} \frac{dp_k}{2\pi} \right) \exp \left( \frac{idt}{2} \sum_{k=1}^{L} p_k \frac{(n_k - n_{k-1})}{dt} \right) \exp \left[ \sum_{k=1}^{L} dt \left( \sum_{\epsilon} K_{n_k-1+2\epsilon,n_k-1} (e^{-2ip_k\epsilon + \gamma} - 1) + \kappa n_{k-1} \right) \right]. \quad (A7)
\]

The propagator of the generating function becomes in the continuous limit

\[
G(x_f, x_i, \kappa, \gamma) = \int \mathcal{D}[x] \int \mathcal{D}[p] \exp \left[ N \int_0^T dt i p \dot{x}_t - \int_0^T dt \left( \sum_{\epsilon} J_{\epsilon}(x_t) \left( 1 - e^{-2ip\epsilon + \gamma} \right) + \kappa x_t \right) \right] \quad (A8)
\]

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

\[
G(x_f, x_i, \kappa, \gamma) = \int \mathcal{D}[x] \int \mathcal{D}[p] \exp \left( N A_{\kappa,\gamma}[x,p,T] \right) \quad (A9)
\]

with the action \(A_{\kappa,\gamma}\) being a functional of the paths \([x,p]\)

\[
A_{\kappa,\gamma}[x,p,T] = \int_0^T dt i p \dot{x}_t - \int_0^T dt \mathcal{H}(x_t, p_t, \kappa, \gamma). \quad (A10)
\]

We have introduced the Hamiltonian function

\[
\mathcal{H}(x,p,\kappa,\gamma) = \sum_{\epsilon = \pm 1} \left[ 1 - \exp (-2ip\epsilon + \gamma) \right] J_{\epsilon}(x) - \kappa x. \quad (A11)
\]

This function is the representation in the continuous limit of a tridiagonal Metzler matrix, whose spectrum is real [48]. As a consequence, \(ip\) is a real number in the following.

**Appendix B: WKB 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 limit. For each integration, we deform the contour so that it goes through the saddle point of \(A_{\kappa,\gamma}[x,p,t]\). Therefore, the saddle point is the minimum of \(A_{\kappa,\gamma}[x,p,t]\) solving

\[
\frac{\delta}{\delta p_t} A_{\kappa,\gamma}[x,p,T] = 0, \quad (B1)
\]

\[
\frac{\delta}{\delta x_t} A_{\kappa,\gamma}[x,p,T] = 0, \quad (B2)
\]

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

\[
\begin{align*}
    i\dot{x}_t &= \partial_p \mathcal{H}(x_t, p_t, \kappa, \gamma), \\
    i\dot{p}_t &= -\partial_x \mathcal{H}(x_t, p_t, \kappa, \gamma).
\end{align*}
\]

(B3) \hspace{1cm} (B4)

The propagator of the generating function can be written as

\[
G(x_f, x_i, \kappa, \gamma) \simeq N \to +\infty \exp \left[ N \int_0^T dt \left( i p_t \dot{x}_t - \mathcal{H}(x_t, p_t, \kappa, \gamma) \right) \right],
\]

(B5)

where \((x, p)\) correspond to the solutions of Eq. (B3-B4) with initial and final conditions \( x_i \) and \( x_f \). Moreover the Hamiltonian (A11) is time independent, and is hence a conserved quantity along the trajectory. Maupertuis’ action \( \int_0^T dt i p_t \dot{x}_t \) is a function of the Hamiltonian and writes

\[
\int_0^T dt i p_t \dot{x}_t = \int_{x_i}^{x_f} dx p(x),
\]

(B6)

where \( p(x) \) can be obtained by inverting Eq. (A11). 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. (B3).

In this large size limit, we can also use a saddle point integration on \( p \) to switch to the Lagrangian framework. To do so, we look for the solution of Eq. (B3). We derive a quadratic equation either for \( e^{ip_s} \) or for \( e^{-ip_s} \) whose solutions are

\[
e^{\pm 2i p_s} = \frac{\mp \dot{x}_s + \sqrt{\dot{x}_s^2 + \varphi(x_s, \gamma)}}{2J_\varnothing(x_s) e^{\gamma}},
\]

(B7)

with \( \varphi(x, \gamma) \) defined in Eq. (59). Inserting this into Eq. (A9) leads to the propagator of the generating function of Eq. (57) with Lagrangian (58).