Eur. Phys. J. Special Topics **224**, 2351–2367 (2015)

DOI: 10.1140/epjst/e2015-02416-9

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

Effective interactions and large deviations in stochastic processes

R.L. Jack[1] and P. Sollich[2]

[1] Department of Physics, University of Bath, Bath BA2 7AY, UK
[2] Department of Mathematics, King's College London, Strand, London WC2R 2LS, UK

Received 19 March 2015 / Received in final form 5 May 2015
Published online 22 June 2015

Abstract. We discuss the relationships between large deviations in stochastic systems, and "effective interactions" that induce particular rare events. We focus on the nature of these effective interactions in physical systems with many interacting degrees of freedom, which we illustrate by reviewing several recent studies. We describe the connections between effective interactions, large deviations at "level 2.5", and the theory of optimal control. Finally, we discuss possible physical applications of variational results associated with those theories.

1 Introduction

Rare events are important in many physical settings: classic examples include phase transformation, protein-folding, and chemical reactions [1–4]. In those cases, a system makes a transition between two distinct states, and a variety of analytical and computational tools are available [4–8]. Here, we focus on a different class of rare events, where systems behave in an unusual fashion over an extended period of time. Specifically, we consider the probability of trajectories in which time-averaged quantities remain far from their typical (equilibrium) values. If the system is ergodic, the probabilities of such events decay to zero as the length of trajectory goes to infinity: the rate of this decay is described by the mathematical theory of large deviations [9]. Recent studies of these large deviations have provided insights into fluctuation theorems [10,11], glassy systems [12–15], protein-folding [16–18], chaotic dynamical systems [19,20] and interacting particle models [21–25].

It turns out that the rare trajectories of interest in these systems can be characterised as typical trajectories for a certain modified system [26–31], which we refer to here as the "auxiliary model". The auxiliary model inherits many of its important properties from the original system of interest: if the original model has the Markov property then so does the auxiliary model. In many cases, the auxiliary model inherits the symmetries of the original model, and other properties like kinetic constraints are also preserved [30].

The existence of this auxiliary model raises important questions for the characterisation of rare events. In particular, it means that by adding a particular set of interactions to the original model, one may drive the system to realise these rare events. In fact, these interactions can be shown to be the "optimal" ones for realising

the rare events of interest, in a certain precise sense [32–35] (see Sect. 4.1, below). It is therefore of great interest to characterise these interactions. For example, in glassy systems, they can stabilise "amorphous solid states" [14,15] that are otherwise only metastable – the nature of the interactions required to achieve this is a long-standing question in the field. In protein-folding systems, effective interactions might stabilise the native state, or they might favour misfolded states [17,18]: understanding how these states can be characterised (and suppressed) is of vital importance in that context.

In this paper, we survey some key results that are related to the existence and nature of these auxiliary models, and the effective interactions that they encode. Our aim is to draw together ideas from several different contexts and to give a (non-rigorous) presentation that highlights the central outstanding questions, and possible routes to solving them. In Sect. 2, we describe the setting for our main results. Section 3 illustrates the kinds of phenomena that we are interested in, through a summary of some recent numerical results. Then, in Sect. 4, we describe some theoretical results, including the relationship to large deviations at "level-2.5" [36] and to optimal control theory [32–35]. These results have not yet been exploited very far in the physics context – we highlight possibilities for future progress along these directions. Section 5 gives a brief summary and outlook.

2 Basic theory

In this section, we collect some key results related to large deviations in stochastic processes. Many of these results have been derived independently in different contexts and by different groups. Here we follow the presentation of [13,30,37]; further details and references can be found in those works.

2.1 Models and master equations

We consider a Markov process in continuous time, on a (finite) discrete state space with configurations $\mathcal{C}$. For example, one can consider a lattice of Ising spins, or a simple particle model such as the asymmetric exclusion process. In practical settings, one is often interested in the thermodynamic limit, where the size of the state space is taken to infinity, for example by considering spins on lattices of increasing size. Alternatively, one may consider diffusive processes described by Langevin equations (stochastic differential equations). These may typically be obtained from lattice models by a continuum limit: one defines a process on a discrete lattice and then takes the lattice spacing to zero, rescaling time in an appropriate way to ensure diffusive behaviour. Our restriction to finite state spaces means that the following analysis may not always be valid on taking thermodynamic or continuum limits – in typical cases we expect our results to remain valid in such limits, but this is not guaranteed. (In the thermodynamic limit, the most serious problems arise in cases of phase transitions. In the continuum limit, the difficulties are mostly technical – one expects the results here to apply as long the system is ergodic, and particles' probability distributions decay to zero at large distances. We provide a few comments on these points in later sections.)

The transition rates between configurations of the system are $W(\mathcal{C}' \leftarrow \mathcal{C})$. Let $P(\mathcal{C}, t)$ be the probability that the system is in configuration $\mathcal{C}$ at time t: this quantity evolves by a Master equation

$$\partial_t P(\mathcal{C}, t) = -r(\mathcal{C})P(\mathcal{C}, t) + \sum_{\mathcal{C}'} W(\mathcal{C} \leftarrow \mathcal{C}')P(\mathcal{C}', t) \tag{1}$$

where $r(\mathcal{C}) = \sum_{\mathcal{C}'} W(\mathcal{C}' \leftarrow \mathcal{C})$ is the "escape rate" from configuration $\mathcal{C}$. We assume that the process is irreducible, which ensures ergodicity, since the state space is finite. It is also useful to identify the subclass of these models that obey detailed balance. For these models, there exists a "potential" $E_{\mathcal{C}}$ such that

$$W(\mathcal{C} \leftarrow \mathcal{C}')\mathrm{e}^{-E_{\mathcal{C}'}} = W(\mathcal{C}' \leftarrow \mathcal{C})\mathrm{e}^{-E_{\mathcal{C}}}, \tag{2}$$

for all $\mathcal{C}$ and $\mathcal{C}'$. Models with this property have time-reversal symmetric ("equilibrium") steady states, in which the probability distribution over configurations is $p(\mathcal{C}) \propto \mathrm{e}^{-E_{\mathcal{C}}}$.

2.2 Large deviations and biased ensembles

The rare events that we consider are defined by the choice of an observable, which may be one of two types. A trajectory of the system consists of the (ordered) set of states which the system visits, and the times at which transitions (jumps) between states take place. The first type of observable takes the general form

$$A = \sum_{\text{jumps } \mathcal{C} \to \mathcal{C}'} \alpha(\mathcal{C}' \leftarrow \mathcal{C}) \tag{3}$$

where the sum runs over all transitions within the trajectory and the $\alpha(\mathcal{C}' \leftarrow \mathcal{C})$ are a given set of numbers. For example, if $\alpha = 1$ for all pairs of configurations then A is the total number of configuration changes in the trajectory. The second type of observable is the time integral of a state-dependent quantity

$$B = \int_0^{t_{\mathrm{obs}}} \mathrm{d}t\, b(\mathcal{C}(t)). \tag{4}$$

where $\mathcal{C}(t)$ is the configuration of the system at time t. For large t_{obs}, the probability distribution of B generically has a large deviation form:

$$p(B) \sim \exp[-t_{\mathrm{obs}}\phi(B/t_{\mathrm{obs}})] \tag{5}$$

where $\phi(b)$ is known as a rate function. A similar expression holds for the distribution of A. [Here, p is a probability density function and the precise meaning of (5) is that $\lim_{t_{\mathrm{obs}}\to\infty} t_{\mathrm{obs}}^{-1} \ln p(B = bt_{\mathrm{obs}}) = -\phi(b)$; the "$\sim$" symbol is used in this sense throughout this article.] The main question of interest in the following is: what kinds of dynamical trajectory dominate the distribution $p(B)$ when B is not equal to its typical (steady-state) value?

To obtain information about these trajectories, it is convenient to write a biased probability distribution over the possible trajectories of the model:

$$\mathcal{P}[\mathcal{C}(t); s] = \mathcal{P}[\mathcal{C}(t); 0] \cdot \frac{\mathrm{e}^{-sB[\mathcal{C}(t)]}}{Z(s, t_{\mathrm{obs}})} \tag{6}$$

where $\mathcal{P}[\mathcal{C}(t); 0]$ is the unbiased (steady-state) probability distribution over trajectories $\mathcal{C}(t)$, the notation $B = B[\mathcal{C}(t)]$ indicates functional dependence on the trajectory $\mathcal{C}(t)$, the parameter s sets the strength of the bias, and $Z = \langle \mathrm{e}^{-sB} \rangle_0$ resembles a partition function. (We note that $\mathcal{P}[\mathcal{C}(t); s]$ is a probability density function in the space of trajectories: see for example [13,37] for an explicit construction of these objects. By contrast, probabilities such as $P(\mathcal{C}, t)$ in (1) are distributions over the discrete

configuration space, at a fixed time t.) Hence, the average of any observable O within the generalised ensemble defined by (6) is

$$\langle O\rangle_s = \frac{\langle O\mathrm{e}^{-sB}\rangle_0}{Z(s,t_{\mathrm{obs}})}. \tag{7}$$

It may be shown [38] that averages within this biased ensemble are the same as those in an ensemble in which the value of A (or B) is constrained to a particular value. [Note that this equivalence is assured only in systems with finite state spaces, in which case the free energy $\psi(s)$ is analytic and convex. In systems with infinite state spaces, dynamical phase transitions [12,37,39] may mean that trajectories which are representative of some values of A (or B) cannot be obtained within biased ensembles of the form given in (6).]

To analyse these biased ensembles, one considers the probability that a system is in configuration $\mathcal{C}$ at time t, and that the observable B has a particular value $\tilde{B}$ associated with the trajectory up to time t [11,13,37]. (The analysis for observables of type A is similar.) If the probability that B is between $\tilde{B}$ and $\tilde{B}+\mathrm{d}\tilde{B}$ is $P(\mathcal{C},\tilde{B},t)\mathrm{d}\tilde{B}$ then we define $P(\mathcal{C},s,t)=\int \mathrm{d}\tilde{B}\, P(\mathcal{C},\tilde{B},t)\mathrm{e}^{-s\tilde{B}}$. This quantity evolves by an equation which is formed from (1) by replacing $P(\mathcal{C},t)$ with $P(\mathcal{C},s,t)$, and adding a term $-sb(\mathcal{C})P(\mathcal{C},s,t)$ to the right hand side. The resulting equation is linear in P so it is useful to write it formally as

$$\partial_t|P\rangle = \mathbb{W}(s)|P\rangle \tag{8}$$

where $\mathbb{W}(s)$ is an operator (matrix) with diagonal elements $-r(\mathcal{C})-sb(\mathcal{C})$ and off-diagonal elements $W(\mathcal{C}'\leftarrow\mathcal{C})$. In the case of type-$A$ observables, the parameters $\alpha(\mathcal{C}'\leftarrow\mathcal{C})$ appear in the off-diagonal elements via multiplicative factors $\mathrm{e}^{-s\alpha}$ [13,37]. Note that (8) resembles a master equation, but it does not conserve probability (in the sense that $\sum_{\mathcal{C}} P(\mathcal{C},s,t)$ is not constant under the time evolution).

2.3 Connection between type-A and type-B observables

We note at this point that the operator $\mathbb{W}(s)$ fully specifies the probability distribution in (6), up to possible boundary terms that we will neglect in the following (see also Sect. 2.4, below). This means that if two processes have the same initial condition and are associated with the same operator $\mathbb{W}(s)$, then they have the same behaviour. It follows that ensembles defined by type-A observables can be given alternative definitions in terms of type-B observables, but for a different underlying stochastic model.

For example, suppose that a model has transition rates $W(\mathcal{C}'\leftarrow\mathcal{C})$ and is biased by an observable of type A. Then, the same operator $\mathbb{W}(s)$ can be obtained by considering a different model with transition rates $\tilde{W}(\mathcal{C}'\leftarrow\mathcal{C})=W(\mathcal{C}'\leftarrow\mathcal{C})\mathrm{e}^{-s\alpha(\mathcal{C}'\leftarrow\mathcal{C})}$, biased by an observable $\tilde{B}=s^{-1}\int \mathrm{d}t[r(\mathcal{C}_t)-\tilde{r}(\mathcal{C}_t)]$ where $\tilde{r}(\mathcal{C})=\sum_{\mathcal{C}'}\tilde{W}(\mathcal{C}'\leftarrow\mathcal{C})$: see for example [13, Appendix B]. A similar transformation means that any B-biased process can always be re-written as an A-biased one. (This requires that $r(\mathcal{C})+sb(\mathcal{C})>0$ for all configurations, which in finite state spaces can always be achieved by including an appropriate constant shift in $b(\mathcal{C})$.) Hence, in the following, we sometimes state results either for type-A or type-B observables, since the results for the other type can always be derived by an appropriate transformation.

2.4 Auxiliary models

Given a model [specified by rates $W(\mathcal{C}'\leftarrow\mathcal{C})$] and an observable [specified by the $\alpha(\mathcal{C}'\leftarrow\mathcal{C})$ or $b(\mathcal{C})$], one may always define an auxiliary model whose steady state

distribution of trajectories is close to (6). [A precise characterisation of this "closeness" is given in (14) below]. For observables of type B, the transition rates of the auxiliary model are [27,28,30,31]

$$W^{\mathrm{aux}}(\mathcal{C}' \leftarrow \mathcal{C}) = u_{\mathcal{C}'} W(\mathcal{C}' \leftarrow \mathcal{C}) u_{\mathcal{C}}^{-1} \tag{9}$$

where the $u_{\mathcal{C}}$ are obtained by solving an eigenvalue equation for the operator $\mathbb{W}(s)$. Specifically, $\langle u|$ is the left eigenvector associated with the smallest eigenvalue of $-\mathbb{W}(s)$:

$$\langle u|(-\mathbb{W}(s)) = \psi(s)\langle u|. \tag{10}$$

Here $\psi(s)$ is a dynamical free energy, related to the dynamical partition function by $Z(s, t_{\mathrm{obs}}) \sim \mathrm{e}^{-t_{\mathrm{obs}}\psi(s)}$. The matrix $-\mathbb{W}(s)$ obeys the necessary conditions of the Perron-Frobenius theorem, so the eigenvector $\langle u|$ is unique and has strictly positive elements (the original Markov process was assumed to be irreducible). We note the connection of (9) to Doob's h-transform [40], which is one of the earliest results connecting rare events to auxiliary models of this kind. Similar results also appear in other kinds of biased rare-event problems [29,34,41], and may also be generalised to quantum systems [42].

Equation (9) motivates us to define an "effective potential"

$$\Delta V_{\mathcal{C}} = -2 \ln u_{\mathcal{C}}. \tag{11}$$

With this definition, $W^{\mathrm{aux}}(\mathcal{C}' \leftarrow \mathcal{C}) = W(\mathcal{C}' \leftarrow \mathcal{C})\mathrm{e}^{(\Delta V_{\mathcal{C}} - \Delta V_{\mathcal{C}'})/2}$, which can be interpreted as a modification of the original transition rates according to the change of the effective potential in a transition. For type-A observables, the analogue of (9) is

$$W^{\mathrm{aux}}(\mathcal{C}' \leftarrow \mathcal{C}) = u_{\mathcal{C}'} W(\mathcal{C}' \leftarrow \mathcal{C})\mathrm{e}^{-s\alpha(\mathcal{C}' \leftarrow \mathcal{C})} u_{\mathcal{C}}^{-1}. \tag{12}$$

To see the relation between W^{aux} and $\mathbb{W}(s)$, we define an operator $\mathbb{W}^{\mathrm{aux}}$ whose off-diagonal elements are the $W^{\mathrm{aux}}(\mathcal{C}' \leftarrow \mathcal{C})$ and whose diagonal elements are $-r^{\mathrm{aux}}(\mathcal{C})$, with escape rates $r^{\mathrm{aux}}(\mathcal{C}) = \sum_{\mathcal{C}'} W^{\mathrm{aux}}(\mathcal{C}' \leftarrow \mathcal{C})$. If we also define $\hat{u}$ to be a diagonal operator whose elements are the $u_{\mathcal{C}}$, it follows [30] that

$$\mathbb{W}^{\mathrm{aux}} = \hat{u}\mathbb{W}(s)\hat{u}^{-1} + \psi, \tag{13}$$

which holds for both type-A and type-B observables. Denoting the elements of the dominant right eigenvector of $-\mathbb{W}(s)$ by $v_{\mathcal{C}}$, one has that the steady-state distribution of configurations in the auxiliary model is $p^{\mathrm{aux}}(\mathcal{C}) \propto u_{\mathcal{C}} v_{\mathcal{C}}$ [30].

With these definitions, the trajectory measure for the steady state of the auxiliary model, $\mathcal{P}[\mathcal{C}(t); \mathrm{aux}]$, is related to the biased ensemble (6) as

$$\mathcal{P}[\mathcal{C}(t); s] = \mathcal{P}[\mathcal{C}(t); \mathrm{aux}] \cdot \frac{\mathrm{e}^{[\Delta V_{\mathcal{C}(t_{\mathrm{obs}})} - \Delta V_{\mathcal{C}(0)}]/2}}{Z^{\mathrm{aux}}} \cdot \frac{p_0(\mathcal{C}(0))}{p^{\mathrm{aux}}(\mathcal{C}(0))} \tag{14}$$

where Z^{aux} is a normalisation constant, and the final factor on the rhs is the ratio of the probability of the initial configuration $\mathcal{C}(0)$ in the original process $[p_0(\mathcal{C}(0))]$, and the probability of the same configuration in the steady state of the auxiliary process $[p^{\mathrm{aux}}(\mathcal{C}(0))]$. (The Perron-Frobenius property of $-\mathbb{W}(s)$ ensures that $p^{\mathrm{aux}}(\mathcal{C}) > 0$ for all $\mathcal{C}$ so this ratio always exists.) Equation (14) is most easily derived via direct

construction of the various $\mathcal{P}[\mathcal{C}(t)]$. For example, if the biasing obervable is of type A then we have

$$\mathcal{P}[\mathcal{C}(t),s] = \left[\prod_{k=1}^{K} \mathrm{e}^{-(t_k-t_{k-1})r(\mathcal{C}_{k-1})}\mathrm{e}^{-s\alpha(\mathcal{C}_k\leftarrow\mathcal{C}_{k-1})}W(\mathcal{C}_k\leftarrow\mathcal{C}_{k-1})\right]$$

$$\times\mathrm{e}^{-(t_{\mathrm{obs}}-t_K)r(\mathcal{C}_K)}p_0(\mathcal{C}_0)\frac{1}{Z(s,t_{\mathrm{obs}})} \tag{15}$$

where the trajectory is composed of configurations $\mathcal{C}_0,\mathcal{C}_1,\ldots,\mathcal{C}_K$, with configuration changes at times $t_1,t_2,\ldots t_K$, we define $t_0=0$, and $p_0(\mathcal{C})$ is the probability of finding configuration $\mathcal{C}$ in the steady state of the original (unbiased) model. A similar construction of the analogous probability density for the auxiliary process then yields (14). For a detailed analysis, see [31], which also includes an analysis of models with continuous state spaces.

Note that $\mathcal{P}[\mathcal{C}(t);\mathrm{aux}]$ in (14) is defined specifically as the steady-state probability distribution over trajectories, while $\mathcal{P}[\mathcal{C}(t);s]$ is defined in terms of a general distribution of initial conditions. The reason for this distinction is to emphasise that while $\mathcal{P}[\mathcal{C}(t);s]$ is defined in terms of the distribution p_0, which specifies the initial condition $\mathcal{C}(0)$ for the unbiased process, the bias s affects the actual initial distribution of $\mathcal{C}(0)$ in the biased ensemble. In fact, since the differences between the auxiliary and biased ensembles in (14) depend only on the initial and final states, one expects that for large t_{obs} then $\mathrm{Prob}[\mathcal{C}(t);s]$ and $\mathrm{Prob}[\mathcal{C}(t);\mathrm{aux}]$ will differ only through initial and final "transient" regimes. In this case, it may be seen from (14) that the distribution of the initial configuration $\mathcal{C}(0)$ in the biased ensemble is proportional to $\mathrm{e}^{-\Delta V_{\mathcal{C}(0)}/2}p_0(\mathcal{C}_0)$, which does indeed depend on the bias s (through $\Delta V_{\mathcal{C}(0)}$). The initial and final transient regimes are discussed in more detail in [13] and also in [31], where it was shown how a set of time-dependent auxiliary rates can lead to exact correspondence between the auxiliary and biased processes.

It is useful to note that (for type-B observables)

$$u(\mathcal{C}) \propto \lim_{t_{\mathrm{obs}}\to\infty}\langle \mathrm{e}^{-sB+t_{\mathrm{obs}}\psi(s)}\rangle_{\mathcal{C},0} \tag{16}$$

where the average is taken with respect to the unbiased dynamics, for a system initialised in configuration $\mathcal{C}$ [27,28,30,61]. The term $t_{\mathrm{obs}}\psi(s)$ in the exponent ensures that the average does not grow or decay exponentially in time, because from the definition of Z and its link to the dynamical free energy one has

$$\mathrm{e}^{-t_{\mathrm{obs}}\psi(s)} \sim Z(s,t_{\mathrm{obs}}) = \langle \mathrm{e}^{-sB}\rangle_0. \tag{17}$$

2.5 Biased ensembles with time-reversal symmetry

In cases where the biased ensembles are symmetric under time-reversal, the eigenvalue problem (10) may be simplified: it reduces to finding the largest eigenvalue of a symmetric matrix. The most common situation in which this occurs is when the unbiased model obeys detailed balance, and the biasing observable is either of type-B, or of type-A with $\alpha(\mathcal{C}'\leftarrow\mathcal{C}) = \alpha(\mathcal{C}\leftarrow\mathcal{C}')$ for all $\mathcal{C}$ and $\mathcal{C}'$. In this case one has simply [13,30]

$$\psi = \min_{|x\rangle}\frac{\langle x|\mathrm{e}^{\hat{E}/2}(-\mathbb{W}(s))\mathrm{e}^{-\hat{E}/2}|x\rangle}{\langle x|x\rangle} \tag{18}$$

where $\hat{E}$ is a diagonal operator whose elements are the energies $E_{\mathcal{C}}$ that appear in the detailed balance relation (2), and the maximisation is over vectors with elements $x_{\mathcal{C}}$. The maximum occurs when $x_{\mathcal{C}} = u_{\mathcal{C}} \mathrm{e}^{-E_{\mathcal{C}}/2}$ so this variational result allows direct estimation of the effective interactions. Generalisations of this result to cases without time-reversal symmetry will be discussed in Sect. 4 below.

3 Illustrative results from model systems

Having introduced the general features of biased ensembles of trajectories, we now return to our original focus on complex systems with many interacting degrees of freedom. In these cases, it is not usually possible to solve the eigenproblem (10) in order to obtain the $u_{\mathcal{C}}$. Further, even if this eigenvector could be obtained exactly, it typically has such a large dimensionality that it does not provide direct information about the physical nature of effective interactions in the system. To illustrate these physical ideas, we now recall some recent results on the physical features of effective interactions in biased ensembles, for different model systems.

3.1 Glass-forming systems

Kinetically constrained models consist of interacting spins (or particles) in which local rules mean that only a subset of spins are able to flip at any given time step [43,44] These models provide simple descriptions of glass-forming liquids [45]. The "mobile" subset of spins changes with time, and the system is ergodic on long time scales. The dynamical motion in these systems can be complex and co-operative, even if their static (thermodynamic) properties are very simple.

In these systems, it is typically possible to construct configurations in which the subset of mobile spins remains finite in the limit of large system size. In this case, if one considers the large deviations of the total number of spin flips in a trajectory (type-A observable with all $\alpha = 1$), it can be shown from (18) that (i) $\lim_{N\to\infty} \psi/N \leq 0$ where N is the system size, (ii) this bound is saturated for all $s > 0$, and (iii) the effective interaction in this case drives the system into configurations with a finite number of mobile spins. It follows that these systems have dynamical phase transitions at $s = 0$ [12,13]. The dominant feature of the effective interactions for $s > 0$ is a very strong suppression of mobile spins, although the detailed nature of the effective interactions that produce this suppression is not known.

Similar phase transitions exist in fully-connected ("mean-field") spin-glass models with large numbers of metastable states [46], and there is also numerical evidence for them in atomistic models of glass-forming liquids [14,15], but the nature of the effective interactions again remains unclear. (In fact, even establishing the existence of phase transitions from numerical simulations is very challenging, since it requires a finite-size scaling analysis in which both system size and observation time t_{obs} are considered together [15,47]. The relevant analysis is not difficult in principle, but obtaining accurate results over a sufficiently large range of length- and time-scales is often difficult with current methods.)

A recent study of a particular kinetically constrained model (the East model [43]) highlights the complex effective interactions that can appear even in simple systems. On biasing this model to low activity, one observes the dynamical phase transition discussed above. However, if one biases instead to high activity, one observes a hierarchy of responses that mirror the "aging" behaviour of the same model [48]. (Aging behaviour occurs when the system is initialised at high temperature followed by dynamical relaxation at low temperature.) The dominant features of these states

are (i) effective interactions that are long-ranged even for weak biases s, and (ii) a hierarchy of length scales associated with different relaxation processes within the system.

3.2 Exclusion processes

There have been many studies of large deviations in exclusion processes, in which particles move on a lattice, with at most one particle per site. Effective interactions in biased ensembles have been considered in relatively few cases; two examples are the limits of maximal dynamical activity or maximal current, where the effective interactions can be found exactly [49]. These interactions are dominated by a long-ranged repulsion between particles: the system can be mapped to a "one-component plasma" of positively-changed particles interacting by Coulomb-like forces. The result of these long-ranged forces is that the system becomes "hyperuniform" [50] – density fluctuations on large length scales are strongly suppressed [51]. Such correlations occur in a variety of non-equilbrium systems [52–54], but they are forbidden in equilibrium systems with short-ranged forces.

Biasing exclusion processes to small activity can also result in phase transitions into inhomogeneous states [23,39,55], although the effective interactions associated with these states have not been investigated in detail. Similar behavior can occur in simple models of heat conduction [56].

3.3 Numerical results

As well as these analytic results, there are several numerical methods that allow large deviations to be investigated. Briefly, *transition path sampling* [7] is a computational method for sampling trajectories of systems according to general path ensembles, including examples such as (6) [14,57]. The method is most easily implemented for processes obeying detailed balance, although generalisations are possible [58]. Alternatively the *cloning* method was developed specifically to study large deviations [19,59,60] and is not restricted to systems with time-reversal symmetry – it involves many copies ("clones") of the system evolving in parallel. Finally, a third method was proposed recently by Nemoto and Sasa [61], which involves direct estimation of the auxiliary rates in (12), in a manner reminiscent of thermodynamic integration. We note that in a system of N spins (or particles), these methods all rely on direct dynamical simulation of trajectories of the system, at a cost that scales linearly with N and with the total time $\mathcal{T}$ to be simulated. The total time $\mathcal{T}$ required to obtain accurate results is not known *a priori*: it depends on the system of interest and the method used, but it also increases strongly as the bias strength $|s|$ increases. For large biases, the cost can quickly become prohibitive. Exact diagonalisation of the generator is also possible in principle: the time required is polynomial in the number of states of the system, so exponential in N. Nevertheless, for small systems and large biases, this method can sometimes be competitive [48].

Methods based on direct simulation have provided a number of interesting insights, especially for models that are not tractable analytically. Examples include model protein-folding systems [17], where biased ensembles are dominated by "misfolded" states, reminiscent of the low-activity states discussed in Sect. 3.1. Similar results can also be obtained in protein systems for which Markovian effective descriptions are available – if the resulting state space is sufficiently small then large deviations can be analysed by exact diagonalisation of the operator $\mathbb{W}(s)$ [16]. One again finds that the effective interactions stabilise misfolded metastable states [18].

Numerical methods have also been used to study the competition between chaotic and periodic behaviour in dynamical systems [19,20]. In particular, even if a system's steady state is chaotic, its large deviations may be characterised by periodic trajectories, which allow the system to avoid "equilibration" into an ergodic state.

We emphasise that the path sampling and cloning methods do not provide direct information about effective interactions, and even the method of [61] typically requires an approximate parameterisation of these interactions to be chosen before starting the analysis. However, the methods do yield representative configurations of the biased system, which at least provide qualitative insights into the underlying interactions. We believe that further development of methods in this area is a useful direction for further study.

3.4 General principles

We identify two general principles from the illustrative examples above. Firstly, biased ensembles of trajectories often contain correlations that are very unusual in equilibrium systems. The hyperuniform states found in exclusion processes are stabilised by long-ranged effective interactions [49,50] – these might not have been anticipated given the simple local rules and the simple bias to high activity. Similarly, the long-ranged correlated states found in the East model biased to high activity do not at all resemble the equilibrium state of that system [48], and nor do the periodic (non-chaotic) trajectories found in some dynamical systems [19,20]. We emphasise that biased states are optimised with respect to global observables (A or B) that depend on the whole system, integrated over a long period of time, so there is no general reason to expect effective interactions to be the short-ranged forces that are familiar from equilibrium settings. So one may expect to find new and unusual phenomena on investigating large deviations. Similarly, if one considers a trajectory of a d-dimensional system as a $(d+1)$-dimensional object, and the biased distribution (6) as a Gibbs-like distribution for this $(d+1)$-dimensional system, one does not expect d-dimensional cross-sections (layers) through the larger system to be described by a simple set of short-ranged interactions [62].

Secondly, effective interactions are often linked with underlying metastable states in a system – biasing to low activity often drives the system into "glassy" metastable states, as found in kinetically-constrained models [12,13], atomistic glass-formers [14,15], and proteins [16–18]. Given the variational principle (18), this may not be suprising – the low-lying eigenvalues of the operator $-\mathbb{W}(0)$ are naturally linked with metastable states and phase transitions, so weak perturbations can be expected to lead to hybridisation of these states with the dominant eigenvector. However, the use of large deviation methods to further analyse dynamical metastability and glassy behaviour seems promising. For example, recent work on biased ensembles in quantum systems also highlights the importance of quiescent (inactive) states that couple weakly to their environment [42].

4 Effective interactions without time-reversal symmetry

This section surveys some results, mostly from the mathematical physics literature, which provide variational methods for determining u_C in systems without time-reversal symmetry, so that (18) does not apply. For systems of practical interest, we are proposing that these results could be useful for (i) analytic bounds on dynamical free energies (for example, proving the existence of phase transitions in non-equilibrium systems, following the analysis of the time-reversible case [12,13]);

(ii) variational analyses of effective interactions, as used in [48]; (iii) improved numerical procedures, for example obtaining an approximation to the auxiliary dynamics in order to improve sampling within a computational scheme. Our purpose here is to highlight these opportunities so we mostly quote relevant results, referring to the literature for more detailed analysis and derivations.

4.1 Optimal control theory

We first state a general variational formula for the free energy $\psi(s)$, which may be viewed as a generalisation of (18) for systems lacking time-reversal symmetry. The variation is over sets of transition rates, which should be chosen to reproduce the auxiliary rates (9) as closely as possible. For type-B observables,

$$\psi(s) = \lim_{t_{\rm obs}\to\infty} \left[\min_{\{W^{\rm var}\}} \frac{1}{t_{\rm obs}} \left\langle sB + \sum_{\text{jumps } \mathcal{C}'\leftarrow\mathcal{C}} \frac{L(\mathcal{C}'\leftarrow\mathcal{C})}{W^{\rm var}(\mathcal{C}'\leftarrow\mathcal{C})} \right\rangle_{\rm var} \right] \tag{19}$$

where the variational parameters are (non-negative) rates $W^{\rm var}(\mathcal{C}'\leftarrow\mathcal{C})$, the average is over a dynamical evolution under those rates starting from some arbitrary initial state, and

$$L(\mathcal{C}\leftarrow\mathcal{C}') = W^{\rm var}(\mathcal{C}\leftarrow\mathcal{C}') \left[\ln \frac{W^{\rm var}(\mathcal{C}\leftarrow\mathcal{C}')}{W(\mathcal{C}\leftarrow\mathcal{C}')} - 1 \right] + W(\mathcal{C}\leftarrow\mathcal{C}'). \tag{20}$$

We note here an equivalent way of writing the objective function in (19) above. By averaging over the number of jumps in any small time interval after time t, starting from the current configuration $\mathcal{C}(t)$, one finds

$$\psi(s) = \lim_{t_{\rm obs}\to\infty} \left[\min_{\{W^{\rm var}\}} \frac{1}{t_{\rm obs}} \int_0^{t_{\rm obs}} {\rm d}t \left\langle s\, b(\mathcal{C}(t)) + \sum_{\mathcal{C}'} L(\mathcal{C}'\leftarrow\mathcal{C}(t)) \right\rangle_{\rm var} \right] \tag{21}$$

where we have also written out B explicitly as a time integral. The minima in (19,21) are obtained when the rates $W^{\rm var}$ are equal to the auxiliary rates defined by (9). A derivation of this result will be sketched in Sect. 4.2 below. We first give a brief discussion of its interpretation and potential usefulness.

The variational principle (19) arises in "optimal control theory" [32–35]: the idea is that $W^{\rm var}$ is a "controlled dynamics" that should be optimised in order to realise the rare event of interest. The content of (19) is that the controlled process should minimise $s\langle B\rangle_{\rm var}$, while deforming the original rates as little as possible. [Note that $L(\mathcal{C}'\leftarrow\mathcal{C})$ resembles a relative entropy between the sets of transition rates, with $L=0$ if $W^{\rm var}(\mathcal{C}'\leftarrow\mathcal{C}) = W(\mathcal{C}'\leftarrow\mathcal{C})$. In fact the final term in (21) is exactly the small-Δt limit of the relative entropy between the distributions of configurations reached from $\mathcal{C}$ in a small time interval Δt, for systems with rates W and $W^{\rm var}$. For the rates W, this distribution is $P_{\Delta t}(\mathcal{C}') = \Delta t W(\mathcal{C}'\leftarrow\mathcal{C})$ for $\mathcal{C}'\neq\mathcal{C}$ and $P_{\Delta t}(\mathcal{C}) = 1 - r(\mathcal{C})\Delta t$ otherwise; the relevant expressions for the rates $W^{\rm var}$ are analogous.] Since the maximum in (19) is obtained when $W^{\rm var} = W^{\rm aux}$, we may restrict the maximisation to rates $W^{\rm var}(\mathcal{C}'\leftarrow\mathcal{C}) = W(\mathcal{C}'\leftarrow\mathcal{C}){\rm e}^{[\Delta V^{\rm var}(\mathcal{C}) - \Delta V^{\rm var}(\mathcal{C}')]/2}$ of the same form as $W^{\rm aux}$. Then $\Delta V^{\rm var}$ has the interpretation of an effective potential that pushes the system towards the rare event of interest. In this context, (19) can be interpreted as an optimisation over the "controlling field" $\Delta V^{\rm var}$.

In the case of diffusive processes, (19) has a particularly simple form: consider a model defined by a Langevin equation (or stochastic differential equation)

$$\dot{x} = K(x) + \eta \tag{22}$$

where $K = K(x)$ is a force and η is a white noise. We then define a "controlled process" $\dot{x} = K - \partial_x V^{\rm var} + \eta$ where $V^{\rm var} = V^{\rm var}(x)$ is the controlling potential. The idea is to discretize in time using a small time interval Δt. For $x' \approx x$, one has

$$\frac{L(x' \leftarrow x)}{W^{\rm var}(x' \leftarrow x)} \approx (x' - x)(-\partial_x V^{\rm var}) + \exp((x' - x)\partial_x V^{\rm var}) - 1. \tag{23}$$

Then averaging over x' with weight $W^{\rm var}$ reduces this to $\Delta t(\partial_x V^{\rm var})^2/2 + O(\Delta t^2)$. Hence

$$\psi(s) = \lim_{t_{\rm obs}\to\infty} \min_{V^{\rm var}} \frac{1}{t_{\rm obs}} \int_0^{t_{\rm obs}} {\rm d}t \left\langle s\, b(x(t)) + \frac{1}{2}[\partial_x V^{\rm var}(x(t))]^2 \right\rangle_{\rm var} \tag{24}$$

where one seeks to simultaneously minimise the average of sB and the magnitude of the controlling force $\partial_x V^{\rm var}$. The relationships between optimal control and large deviations for diffusive systems have been discussed in the physics literature [34,35], but while the results (19,21) for Markov chains are known in the mathematical literature [63], they have not, to our knowledge, been applied very far in physics.

In terms of future applications, it is clear that (19) gives bounds on ψ and allows variational estimates of $W^{\rm aux}$. In principle this enables variational analyses of large deviations in non-equilibrium settings, similar to those described for time-reversible systems in Sect. 3.1. However, there is an additional difficulty associated with (19), which arises from the estimation of the average with respect to the variational (controlled) dynamics. In the absence of detailed balance, these averages will typically need to be obtained by direct numerical simulation, in which case convergence to the limit of large $t_{\rm obs}$ may be non-trivial.

In the case of time-reversal symmetric ensembles, one can restrict to $W^{\rm var}$ that obey detailed balance, and (19) reduces to (18). To see this, replace the expectation value in (21) by an average with respect to the steady state of the controlled dynamics $\mu^{\rm var}(\mathcal{C}) \propto {\rm e}^{-E(\mathcal{C})-\Delta V(\mathcal{C})}$. The key point is that the logarithmic term in $L(\mathcal{C}' \leftarrow \mathcal{C})$ yields $\sum_{\mathcal{C},\mathcal{C}'} W^{\rm var}(\mathcal{C}' \leftarrow \mathcal{C})\mu^{\rm var}(\mathcal{C})[\Delta V(\mathcal{C}) - \Delta V(\mathcal{C}')]/2$; using the detailed balance relation $W^{\rm var}(\mathcal{C}' \leftarrow \mathcal{C})\mu^{\rm var}(\mathcal{C}) = W^{\rm var}(\mathcal{C} \leftarrow \mathcal{C}')\mu^{\rm var}(\mathcal{C}')$ and interchanging the summation variables shows that this term vanishes. Finally using $r(\mathcal{C}) = \sum_{\mathcal{C}'} W(\mathcal{C}' \leftarrow \mathcal{C})$, Eq. (21) reduces to

$$\psi = \min_{\{W^{\rm var}\}} \sum_{\mathcal{C}} \left[sb(\mathcal{C}) + r(\mathcal{C}) - \sum_{\mathcal{C}'} W^{\rm var}(\mathcal{C}' \leftarrow \mathcal{C}) \right] \mu^{\rm var}(\mathcal{C}) \tag{25}$$

which can be shown to be the same as (18).

We highlight two other potential routes for application of (19). First, it can provide simple bounds on ψ by appropriate simple choices of $W^{\rm var}$. For example if one biases by the total activity (number of spin flips), and the system has a configuration with sub-extensive escape rate [there exists a sequence of configurations $\mathcal{C}_N$ in systems of increasing size N such that $r(\mathcal{C}_N)/N \to 0$ as $N \to \infty$], then $\lim_{N\to\infty} \psi(s)/N \le 0$ and hence (given weak conditions on properties of the steady state) there must be a dynamical phase transition at $s = 0$. This is a non-equilibrium analogue of results proven for kinetically constrained models of the glass transition [12,13]. It is relevant for exclusion processes, where the same result may be derived either by exact solution [39] or within fluctuating hydrodynamics [24,25]. But the method based on (19) is both very simple and very general. Second, there should be possibilities of using (19) in numerical schemes, for example by generalising the method of Nemoto and Sasa [61]. This possibility remains to be explored.

Finally one could also consider finite-$t_{\rm obs}$ analogues of (21). We define $\phi(\mathcal{C}, t_{\rm obs})$ as the minimum value of the objective function on the r.h.s. of (21) when starting

from a given configuration $\mathcal{C}$. It is then not difficult to argue that $\phi(\mathcal{C}, t_{\rm obs}) = \psi(s) + \Delta V_{\mathcal{C}}/(2t_{\rm obs})$ for large $t_{\rm obs}$, up to corrections that decay exponentially with $t_{\rm obs}$. If one allows the variational rates $W^{\rm var}$ to depend on time then one can also obtain a closed form for the evolution equation of the $\phi(\mathcal{C}, t_{\rm obs})$. Thus it may be possible to obtain the effective interactions from the finite-$t_{\rm obs}$ behaviour of the optimal cost $\phi(\mathcal{C}, t_{\rm obs})$ in the control theory approach.

4.2 Large deviations at "level-2.5"

To understand the origin of the variational result (19), it is useful to consider the large deviations of a very general set of observables [29]. For a given trajectory $\mathcal{C}(t)$, we define the *empirical current* which is a set of numbers $Q(\mathcal{C}' \leftarrow \mathcal{C})$, obtained by counting the jumps (transitions) between each pair of configurations, and dividing by $t_{\rm obs}$. Similarly, the *empirical measure* is a set of numbers $\mu(\mathcal{C})$ given by the fraction of time that the trajectory spent in each configuration. Note that for a given trajectory, $\sum_{\mathcal{C}'} Q(\mathcal{C}' \leftarrow \mathcal{C}) t_{\rm obs}$ is the total number of jumps that the systen makes out of configuration $\mathcal{C}$, and $\sum_{\mathcal{C}'} Q(\mathcal{C} \leftarrow \mathcal{C}') t_{\rm obs}$ is the total number of jumps into that configuration. These two numbers must be exactly equal unless $\mathcal{C}$ is the initial or final configuration, in which case they differ by at most unity. Dividing by $t_{\rm obs}$ we find a balance condition

$$\sum_{\mathcal{C}'} Q(\mathcal{C}' \leftarrow \mathcal{C}) = \sum_{\mathcal{C}} Q(\mathcal{C} \leftarrow \mathcal{C}'), \tag{26}$$

which holds at the level of individual trajectories, up to corrections of at most $\pm 1/t_{\rm obs}$ which are negligible in the large-$t_{\rm obs}$ limit.

4.2.1 Statement of the large deviation principle

The observables μ and Q are very high-dimensional objects if the state space is large, but as long as there are a finite number of them they obey a large deviation principle whose explicit rate function is known, for both biased and unbiased ensembles of trajectories. We first state the result [36]: for sufficiently large $t_{\rm obs}$, one has $p(\mu, Q) \sim {\rm e}^{-t_{\rm obs} I(\mu, Q)}$ with

$$I(\mu, Q) = \sum_{\mathcal{C}, \mathcal{C}'} \left\{ Q(\mathcal{C}' \leftarrow \mathcal{C}) \left[\log \frac{Q(\mathcal{C}' \leftarrow \mathcal{C})}{W(\mathcal{C}' \leftarrow \mathcal{C}) \mu(\mathcal{C})} - 1 \right] + W(\mathcal{C}' \leftarrow \mathcal{C}) \mu(\mathcal{C}) \right\}. \tag{27}$$

In an ensemble biased by a type-B observable according to (6), all trajectories with a given μ and Q are reweighted by the same factor ${\rm e}^{-sB} = \exp(-s \sum_{\mathcal{C}} b(\mathcal{C}) \mu(\mathcal{C}))$, so after including the normalization factor $1/{\rm e}^{-t_{\rm obs} \psi(s)}$ one has $p(\mu, Q) \sim {\rm e}^{-t_{\rm obs}[I(\mu, Q, s) - \psi(s)]}$ with

$$I(\mu, Q, s) = I(\mu, Q) + s \sum_{\mathcal{C}} b(\mathcal{C}) \mu(\mathcal{C}). \tag{28}$$

The result (27) is known as a "level-2.5" large deviation principle (LDP) since it is intermediate between an LDP for the empirical measure (known as level 2) and a full LDP for trajectories (known as level 3). A review of large deviations at level-2.5 is given in [36], including results for diffusive processes [64,65], while rigorous analysis of the case with countably infinite state spaces is given in [66–69], including a proof of (27).

4.2.2 Connection to the auxiliary process

The typical empirical measure and current within the biased ensemble of (6) can be obtained by minimisation over μ and Q of the rate function $I(\mu, Q, s)$ in (27). The key point for our purposes is that this allows a variational determination of the auxiliary model of (9). We first cast the minimisation over μ, Q as a minimisation over a "variational auxiliary model": a model for which typical trajectories have current Q and measure μ. The transition rates of this model then have to be

$$W^{\mathrm{var}}(\mathcal{C}' \leftarrow \mathcal{C}) = Q(\mathcal{C}' \leftarrow \mathcal{C})/\mu(\mathcal{C}). \tag{29}$$

As one would expect, the balance constraint (26) on Q then ensures that the steady state of the process defined by the rates W^{var} is μ.

We rewrite (27) as

$$I(\mu, Q) = \sum_{\mathcal{C},\mathcal{C}'} \left\{ W^{\mathrm{var}}(\mathcal{C}' \leftarrow \mathcal{C}) \left[\log \frac{W^{\mathrm{var}}(\mathcal{C}' \leftarrow \mathcal{C})}{W(\mathcal{C}' \leftarrow \mathcal{C})} - 1 \right] + W(\mathcal{C}' \leftarrow \mathcal{C}) \right\} \mu(\mathcal{C}). \tag{30}$$

Including the bias term as in (28) gives the function $I(\mu, Q, s)$, which is minimised by the (μ, Q) that are most likely within the biased ensemble: we denote their values by (μ^*, Q^*). The variational rates W^{var} at this minimum of $I(\mu, Q, s)$ define a model for which typical trajectories have $(\mu, Q) = (\mu^*, Q^*)$ so they must be exactly the auxiliary rates W^{aux} associated with the biased ensemble.

Moreover, the minimal value of $I(\mu, Q, s)$ itself is just the dynamical free energy: $\psi(s) = I(\mu^*, Q^*, s)$. This follows from a contraction principle [9] because the observable $B = t_{\mathrm{obs}} \sum_{\mathcal{C}} b(\mathcal{C})\mu(\mathcal{C})$ is a simple function of the empirical measure: recall from (17) that $\langle \mathrm{e}^{-sB} \rangle_0 \sim \mathrm{e}^{-t_{\mathrm{obs}}\psi(s)}$. Hence, decomposing the average into contributions from all possible μ, Q, one has

$$\langle \mathrm{e}^{-sB} \rangle \sim \max_{\mu,Q} \left[\mathrm{e}^{-t_{\mathrm{obs}} I(\mu,Q)} \mathrm{e}^{-s t_{\mathrm{obs}} \sum_{\mathcal{C}} b(\mathcal{C})\mu(\mathcal{C})} \right] \sim \max_{\mu,Q} \mathrm{e}^{-t_{\mathrm{obs}} I(\mu,Q,s)}. \tag{31}$$

Summarising the ingredients so far, we have $\psi(s) = \min_{\mu,Q} I(\mu, Q, s)$ where the minimization can equivalently be done over the variational rates W^{var} rather than μ and Q. The final step in the argument is to realize that in the large t_{obs}-limit, the average over $\mathcal{C}(t)$ in the optimal control formulation (21) above becomes an average over the stationary measure $\mu(C)$, making the penalty term L equal to $I(\mu, Q)$ as rewritten in (30). Thus the variational principle (19) follows from the general level-2.5 result (27).

4.2.3 Derivation of (30)

A derivation of (30) for the case $s = 0$ is given in [29]. Here we outline their argument. The empirical current and measure (μ, Q) are typical for the process W^{var}, but they are not typical for the original process W. For a given trajectory, one may obtain an expression for the ratio $P[\mathcal{C}(t); 0]/P[\mathcal{C}(t); \mathrm{var}]$, using representations analogous to (15). Then one writes the probability of observing an empirical current and measure in the unbiased process as

$$\begin{aligned} \mathrm{e}^{-t_{\mathrm{obs}} I(\mu,Q)} &\sim \sum_{\mathcal{C}(t)|\mu,Q} P[\mathcal{C}(t); 0] = \sum_{\mathcal{C}(t)|\mu,Q} \frac{P[\mathcal{C}(t); 0]}{P[\mathcal{C}(t); \mathrm{var}]} P[\mathcal{C}(t); \mathrm{var}] \\ &\sim \left\langle \frac{P[\mathcal{C}(t); 0]}{P[\mathcal{C}(t); \mathrm{var}]} \right\rangle_{\mathrm{var}}. \end{aligned} \tag{32}$$

The summations in this equation should be interpreted as path integrals over all trajectories that are compatible with an empirical current and measure (μ, Q). The average on the r.h.s. is with respect to the W^{var} process: the restriction to a given (μ, Q) can be omitted here since this average is already dominated by such trajectories, which are typical for that process. Using the explicit form of the ratio to be averaged then yields (27). The analysis of [29] considered only the case $s = 0$, but the general result of (28) follows immediately as explained above, because the effect of the bias in (6) can be re-written as a bias that depends only on the empirical measure.

4.3 Large deviations at level-2

Finally, we recall a classical result of Donsker and Varadhan [26] for large deviations of the empirical measure. Without bias, these satisfy $p(\mu) \sim \mathrm{e}^{-t_{\mathrm{obs}} J(\mu)}$ with

$$J(\mu) = \max_{\rho} \left\{ -\sum_{\mathcal{C},\mathcal{C}'} \rho(\mathcal{C}') W(\mathcal{C}' \leftarrow \mathcal{C}) \frac{\mu(\mathcal{C})}{\rho(\mathcal{C})} + \sum_{\mathcal{C}} r(\mathcal{C}) \mu(\mathcal{C}) \right\} \tag{33}$$

where the maximisation is over a set of variational parameters $\rho(\mathcal{C}) > 0$. In the B-biased ensemble the relevant large deviation function just needs to add the effect of the bias as before, giving $p(\mu) \sim \mathrm{e}^{-t_{\mathrm{obs}}[J(\mu,s)-\psi(s)]}$ with

$$J(\mu, s) = J(\mu) + \sum_{\mathcal{C}} s\, b(\mathcal{C}) \mu(\mathcal{C}). \tag{34}$$

The corresponding expression for ensembles biased by type-A observables is given in [13, Appendix C]. Subsequent minimisation over μ yields the dynamical free energy ψ, and the $\rho(\mathcal{C})$ at the minimum are the $u(\mathcal{C})$ associated with the auxiliary dynamics of Eq. (9).

The result (33) can be obtained by minimisation of (27) over Q, subject to the balance constraints (26). Calling the minimum value $J(\mu)$, we want to show that it can be obtained alternatively from the maximisation problem (33). This can be done using Lagrangian duality: the Lagrangian for the original minimisation is

$$L(\mu, Q, \lambda) = I(\mu, Q) + \sum_{\mathcal{C},\mathcal{C}'} \lambda(\mathcal{C}) [Q(\mathcal{C}' \leftarrow \mathcal{C}) - Q(\mathcal{C} \leftarrow \mathcal{C}')]. \tag{35}$$

The dual Lagrangian is then defined as $\tilde{L}(\mu, \lambda) = \min_Q L(\mu, Q, \lambda)$. Since for any Q satisfying (26) one has $L(\mu, Q, \lambda) = I(\mu, Q)$, it follows that $\tilde{L}(\mu, \lambda) \leq J(\mu)$. Since the equality holds for the optimal Q, one has the dual representation $J(\mu) = \max_\lambda \tilde{L}(\mu, \lambda)$. Now setting the derivative of $L(\mu, Q, \lambda)$ to zero to find $\tilde{L}(\mu, \lambda)$ gives

$$\log \frac{Q(\mathcal{C}' \leftarrow \mathcal{C})}{W(\mathcal{C}' \leftarrow \mathcal{C}) \mu(\mathcal{C})} = \lambda(\mathcal{C}') - \lambda(\mathcal{C}). \tag{36}$$

Substituting back into $L(\mu, Q, \lambda)$, the log term cancels with the Lagrange multiplier contribution and one is left with

$$\tilde{L}(\mu, \lambda) = \sum_{\mathcal{C},\mathcal{C}'} \left\{ -W(\mathcal{C}' \leftarrow \mathcal{C}) \mu(\mathcal{C}) \mathrm{e}^{\lambda(\mathcal{C}) - \lambda(\mathcal{C}')} + W(\mathcal{C}' \leftarrow \mathcal{C}) \mu(\mathcal{C}) \right\}. \tag{37}$$

Identifying $\rho(\mathcal{C}) = \mathrm{e}^{-\lambda(\mathcal{C})}$ and carrying out the sum over $\mathcal{C}'$ in the second term then gives (33) as desired.

To our knowledge, Eq. (34) has had limited application for estimation of $\psi(s)$ and the $u(\mathcal{C})$. One obstacle is that this requires a *maximisation* over ρ, followed by a minimisation over μ. For this reason, straightforward bounds on ψ are not directly available, unlike the case of (27) where one minimises over both μ and Q.

5 Outlook

We have summarised a range of analytical and numerical results related to the effective potentials encoded by (11). Section 3 reviews some previous results where these effective potentials have been estimated, mostly in time-reversal symmetric ensembles. Section 4 shows how the effective potentials can be interpreted in terms of the controlling forces that achieve rare events most efficiently, in the sense of the "objective function" L in (20). We have discussed how the variational results described in Sects. 4.1 and 4.2 might be useful for generalising these kinds of method to systems without detailed balance, and for developing new numerical methods, possibly following Ref. [61]. The application of these results to biased ensembles for open quantum systems [42] might also provide useful insights.

Another general challenge coming from biased ensembles is the description of biased states that are inhomogeneous in space and time. The "addivity principle" leads to some exact results in homogeneous systems, but an accurate description of spatially inhomogeneous (phase-separated) states remains outstanding in some cases. Biased ensembles also support "travelling-wave" states which are inhomogeneous in both space and time [39,70]: it might be useful to investigate variational techniques based on (27) in order to address these problems.

From a fundamental point of view, the relation between effective interactions and the thermodynamic limit is also important. Biased ensembles in general will be characterised by some stationary measure $\mu(\mathcal{C})$. Restricting for convenience to systems with time-reversal symmetry one then expects that this has the form $\mu(\mathcal{C}) \propto \mu_0(\mathcal{C})\mathrm{e}^{-\Delta V_\mathcal{C}}$, where μ_0 is the stationary distribution of the unbiased process, and $\Delta V_\mathcal{C}$ an effective potential. However, in the thermodynamic limit, a question arises as to whether the measure μ is "Gibbsian" [62,71]: that is, whether ΔV can be written as a well-defined sum of interaction terms of increasing range. If such a description is not possible, even the definition of effective interactions becomes problematic in the thermodynamic limit. If one considers large deviations of the total energy (type-B) in the Ising model, there is evidence that the resulting effective interactions may not be Gibbsian [30]. It is also not clear whether the limits of large system size and large-t_{obs} should commute in such cases, and what consequences this might have. It would be interesting to analyse these questions further in future work.

We thank Raphael Chétrite, Hugo Touchette, Carsten Hartmann, Vivien Lecomte, Fred van Wijland, Juan Garrahan, and David Chandler for many useful discussions on the issues discussed here. RLJ thanks the EPSRC for support through grant EP/I003797/1.

References

1. S. Auer, D. Frenkel, Nature **409**, 6823 (2001)
2. R.P. Sear, J. Phys.: Cond. Matt. **19**, 033101 (2007)
3. W. Ren, E. vanden-Eijnden, P. Maragakis, W. E, J. Chem. Phys. **123**, 134109 (2005)
4. P. Hänggi, P. Talkner, M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990)
5. D.J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003)
6. W.E, W. Ren, E. vanden-Eijden, J. Phys. Chem. B **109**, 6688 (2005)

7. P.G. Bolhuis, D. Chandler, C. Dellago, P.L. Geissler, Ann. Rev. Phys. Chem. **53**, 291 (2002)
8. R.J. Allen, D. Frenkel, P.R. ten Wolde, J. Chem. Phys. **124**, 024102 (2006)
9. H. Touchette, Phys. Rep. **478**, 1 (2009)
10. G. Gallavotti, E.G.D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995)
11. J.L. Lebowitz, H. Spohn, J. Stat. Phys. **95**, 333 (1999)
12. J.P. Garrahan, R.L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, F. van Wijland, Phys. Rev. Lett. **98**, 195702 (2007)
13. J.P. Garrahan, R.L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, F. van Wijland, J. Phys. A **42**, 075007 (2009)
14. L.O. Hedges, R.L. Jack, J.P. Garrahan, D. Chandler, Science **323**, 1309 (2009)
15. T. Speck, D. Chandler, J. Chem. Phys. **136**, 184509 (2012)
16. J.K. Weber, R.L. Jack, V.S. Pande, J. Am. Chem. Soc. **135**, 5501 (2013)
17. A.S.J.S. Mey, P.L. Geissler, J.P. Garrahan, Phys. Rev. E **89**, 032109 (2014)
18. J.K. Weber, R.L. Jack, C.R. Schwandtes, V.S. Pande, Biophys. J **107**, 974 (2014)
19. J. Tailleur, J. Kurchan, Nat. Physics **3**, 203 (2007)
20. K.-D.N.T. Lam, J. Kurchan, D. Levine, J. Stat. Phys. **137**, 1079 (2009)
21. B. Derrida, J.L. Lebowitz, Phys. Rev. Lett. **80**, 209 (1998)
22. L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, C. Landim, Phys. Rev. Lett. **87**, 040601 (2001)
23. T. Bodineau, B. Derrida, Phys. Rev. Lett. **92**, 180601 (2004)
24. L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, C. Landim, Rev. Mod. Phys. (in press) (2015)
25. L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, C. Landim, Rev. Mod. Phys. (preprint) [arXiv:1404.6466]
26. M.D. Donsker, S.R.S. Varadhan, Commun. Pure Appl. Math. **28**, 1 (1975)
27. R.M.L. Evans, Phys. Rev. Lett. **92**, 150601 (2004)
28. R.M.L. Evans, J. Phys. A **38**, 293 (2005)
29. C. Maes, K. Netocny, EPL **82**, 30003 (2008)
30. R.L. Jack, P. Sollich, Prog. Theor. Phys. Supp. **184**, 304 (2010)
31. R. Chétrite, H. Touchette, Ann. Henri Poincaré (in press) (2014), doi: 10.1007/s00023-014-0375-8
32. W.H. Fleming, *Stochastic control and large deviations*, in *Future Tendencies in Computer Science, Control and Applied Mathematics* (Springer, Berlin, 1992), p. 291, doi: 10.1007/3-540-56320-2_66
33. W.H. Fleming, H.M. Soner, *Controlled Markov Processes and Viscosity Solutions* (Springer, Berlin, 2005)
34. C. Hartmann, C. Schütte, J. Stat. Mech., P11004 (2012)
35. V.Y. Chernyak, M. Chertkov, J. Bierkens, H.J. Kappen, J. Phys. A **47**, 022001 (2013)
36. A.C. Barato, R. Chetrite [arXiv:1408.5033]
37. V. Lecomte, C. Appert-Roland, F. van Wijland, J. Stat. Phys. **127**, 51 (2007)
38. R. Chétrite, H. Touchette, Phys. Rev. Lett. **111**, 120601 (2013)
39. T. Bodineau, B. Derrida, Phys. Rev. E **72**, 066110 (2005)
40. D.W. Stroock, *An Introduction to Markov Processes* (Springer, Berlin/Heidelberg, 2005)
41. M. Cameron, E. Vanden-Eijnden, J. Stat. Phys. **156**, 427 (2014).
42. J.P. Garrahan, I. Lesanovsky, Phys. Rev. Lett, **104**, 160601 (2010)
43. F. Ritort, P. Sollich, Adv. Phys. **52**, 219 (2003)
44. J.P. Garrahan, P. Sollich, C. Toninelli, *Kinetically constrained models*, Ch. 10 in *Dynamical heterogeneities in glasses, colloids, and granular media*, edited by L. Berthier, G. Biroli, J.-P. Bouchaud, L. Cipelletti, W. van Saarloos (Oxford University Press, Oxford, 2011)
45. D. Chandler, J.P. Garrahan, Ann. Rev. Phys. Chem. **61**, 191 (2010)
46. R.L. Jack, J.P. Garrahan, Phys. Rev. E **81**, 011111 (2010)
47. Y.S. Elmatad, R.L. Jack, J.P. Garrahan, D. Chandler, PNAS **107**, 12793 (2010)
48. R.L. Jack, P. Sollich, J. Phys. A **47**, 015003 (2014)

49. V. Popkov, G.M. Schütz, J. Stat. Phys. **142**, 627 (2011)
50. R.L. Jack, I.R. Thompson, P. Sollich, Phys. Rev. Lett. **114**, 060601 (2015)
51. S. Torquato, F.H. Stillinger, Phys. Rev. E **68**, 041113 (2003)
52. A. Gabrielli, B. Jancovici, M. Joyce, J.L. Lebowitz, L. Pietronero, F. Sylos Labini, Phys. Rev. D **67**, 043506 (2003).
53. C.E. Zachary, Y. Jiao, S. Torquato, Phys. Rev. Lett. **106**, 178001 (2011)
54. Y. Jiao, T. Lau, H. Hatzikirou, M. Meyer-Hermann, J.C. Corbo, S. Torquato, Phys. Rev. E **89**, 022721 (2014)
55. L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, C. Landim, Phys. Rev. Lett. **94**, 030601 (2005)
56. P.I. Hurtado, C.P. Espigares, J.J. del Pozo, P.L. Garrido, J. Stat. Phys. **154**, 214 (2014)
57. M. Merolle, J.P. Garrahan, D. Chandler, PNAS **102**, 10837 (2005)
58. G.E. Crooks, D. Chandler, Phys. Rev. E **64**, 026109 (2001)
59. C. Giardina, J. Kurchan, L. Peliti, Phys. Rev. Lett. **96**, 120603 (2006)
60. V. Lecomte, J. Tailleur, J. Stat. Mech., P03004 (2007)
61. T. Nemoto, S. Sasa, Phys. Rev. Lett. **122**, 090602 (2014)
62. C. Maes, F. Redig, A. van Moffaert, J. Stat. Phys. **96**, 69 (1999)
63. P. Dupuis, R.S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations* (Wiley, New York, 1997)
64. C. Maes, K. Netocny, B. Wynants, Physica A **287**, 2675 (2008)
65. V.Y. Chernyak, M. Chertkov, S.V. Malinin, R. Teodorescu, J. Stat. Phys. **137**, 109 (2009)
66. L. Bertini, D. Gabrielli, A. Faggionnato, Ann. Henri Poincaré (in press) (2015)
67. L. Bertini, D. Gabrielli, A. Faggionnato, Ann. Henri Poincaré (preprint) [`arXiv:1210.2004`] (2012)
68. L. Bertini, A. Faggionato, Gabrielli, Markov Process. Relat. Fields **20**, 545 (2014)
69. L. Bertini, A. Faggionato, Gabrielli (preprint) [`arXiv:1212.6908`] (2012)
70. L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, C. Landim, J. Stat. Phys. **123**, 237 (2006)
71. A.C.D. van Enter, R. Fernández, A.D. Sokal, J. Stat. Phys. **72**, 879 (1993)