Rare switching events in non-stationary systems

Nils B. Becker[1] and Pieter Rein ten Wolde[1]

FOM Institute for Atomic and Molecular Physics (AMOLF), Science Park 104, 1098 XG Amsterdam,
The Netherlands

(Dated: January 2, 2014)

Physical systems with many degrees of freedom can often be understood in terms of transitions between a small number of metastable states. For time-homogeneous systems with short-term memory these transitions are fully characterized by a set of rate constants. We consider the question how to extend such a coarse-grained description to non-stationary systems and to systems with finite memory. We identify the physical regimes in which time-dependent rates are meaningful, and state microscopic expressions that can be used to measure both externally time-dependent and history-dependent rates in microscopic simulations.

I. INTRODUCTION

Physical systems with a large number of degrees of freedom and complicated dynamics can be understood in much simpler terms, if they exhibit a small number of metastable states. A natural coarse-grained description is then found in terms of residence times in each of the states, joined by nearly instantaneous transitions between them.

This idea has been widely used in the context of thermal equilibrium systems that are characterized by a clear separation of time scales between fast intra-state dynamics and slow global relaxation (e.g.[1]), giving rise to Markov State Models, see e.g.[2]. Here transitions are Poissonian, i.e. they happen with uniform propensity (probability per unit time); equivalently, the waiting time intervals are uncorrelated and exponentially distributed. Each transition is described by a single number, the rate constant.

In this article we are interested in systems that switch between two metastable states rarely and rapidly, but with a time-dependent propensity. Such time-dependence may be caused by an external force driving the system. Alternatively it may also arise due to the presence of internal degrees of freedom that do not fully equilibrate on the macroscopic time scale of interest, giving rise to non-Markovian macroscopic switching dynamics between the metastable states. We consider both equilibrium and non-equilibrium systems, in the sense that their unperturbed microscopic dynamics may or may not be time-reversible.

Clearly, in such time-inhomogeneous or non-Markovian systems the concept of a *rate constant* is inadequate. In this article we investigate in what physical regimes a time-dependent *rate function* or a history-dependent *rate kernel* may instead be meaningfully defined, and state the corresponding phenomenological rate equations. We then address the question how macroscopic rate functions or rate kernels can be defined in terms of microscopic correlation functions; we give microscopic expressions and show how they can be used to measure these generalized rates in computer simulations.

II. OVERVIEW

We consider systems whose coarse-grained, macroscopic dynamics can be described as switching between two macroscopic states A and B. The macroscopic states are assumed to be metastable in the sense that the mean waiting time between the switching events is much longer than the duration of the switching event itself. This makes it possible to partition the phase space into the metastable regions A and B, and a transition region C separating them, such that

$$\tau_C \ll \tau_{AB}, \tau_{BA}, \tag{1}$$

where τ_C is a typical duration for a traversal of C (a switching event), and τ_{AB}, τ_{BA} are the mean waiting times for switching in the forward and backward direction, respectively. This separation of time scales provides the justification for coarse graining the system as a two-state system, switching rarely but rapidly between the states A and B.

We are interested both in equilibrium systems that are microscopically reversible and in non-equilibrium systems with dissipation of energy. If the system is in stationary state, the stability of the states A and B and the forward and backward probability fluxes, q_{AB} and q_{BA}, are constant in time. Nonetheless, if the system has memory, the switching propensities will be history-dependent. Furthermore, out of stationary state, the stability of the states and the dynamics of switching between them will change in time. In both of these situations, rate constants k_{AB} and k_{BA} do not adequately describe the system, and a time-dependent generalization of the rate constant concept is called for.

Following Chandler in his derivation of rate constants in equilibrium systems[3], we imagine that from experimental observations we know a certain phenomenological expression for the macroscopic switching dynamics to be valid, on times much larger than a certain macroscopic time resolution Δt. This macroscopic expression is the rate equation. The goal is then to derive a microscopic expression for the dynamics of the system that is consistent with the rate equation, and allows us to measure the macroscopic rates in a microscopic simulation. We will thus not derive a macroscopic expression from

microscopic principles; rather, we will *assume* that the system obeys a given rate equation, and use this description as a starting point for the derivation of microscopic expressions for time-dependent rate functions[3].

The propensity of the system to switch from one macroscopic state to another at time t may depend on the current macroscopic state of the system only, in which case the system is Markovian. More generally, it may depend also upon the macroscopic history, i.e. the sequence of states visited in the past. For such a non-Markovian system, the switching propensity could depend on the time that has passed since the last switching event, but it is also conceivable that it depends on the system's dynamics prior to the last switching event. Indeed, all information about the history of the macroscopic switching dynamics is contained in the sequence of switching times. It is thus natural to capture the history dependence of the switching dynamics by writing the time evolution of the system in terms of the times of all previous switching events:

$$\frac{\partial}{\partial t}P_B(t;t',t'',\dots) = k_{AB}(t|t',t'',\dots)P_A(t;t',t'',\dots)$$
$$-k_{BA}(t|t',t'',\dots)P_B(t;t',t'',\dots),$$
$$(2)$$

In this Master equation, $P_A(t;t',t'',\dots)$ denotes the *joint* probability to be in A at time t and to have switched for the last time within $(t',t'+\mathrm{d}t')$ (i.e. from B to A), for the second-to-last time in $(t'',t''+\mathrm{d}t'')$ (from A to B), and so on. Moreover, $k_{AB}(t|t',t'',\dots)\mathrm{d}t$ is the conditional probability to leave A in the time interval $(t,t+\mathrm{d}t)$ given that the sequence of switching times was $t',t'',\dots$. An analogous equation holds with A and B interchanged. Importantly, this macroscopic phenomenological rate equation *defines* the general rate kernels $k_{AB}(t|t',t'',\dots)$ and $k_{BA}(t|t',t'',\dots)$. An equation for the switching dynamics at time t can be obtained from Eq. 2 by integrating over the switching times prior to t,

$$\frac{\mathrm{d}}{\mathrm{d}t}P_B(t) = \int_{t>t'>t''>\dots} \frac{\partial}{\partial t}P_B(t;t',t'',\dots)\mathrm{d}t'\mathrm{d}t''\dots. \quad (3)$$

Since the last, second-to-last, ... switches are each unique events, no over-counting occurs here. No approximation has been made up to this point; note however that 3 is not in general a closed equation for $P_{A,B}(t)$. The lower integration limit depends upon the experimental setup and will be discussed in more detail below.

While Eqs. 2 and 3 describe the switching dynamics of an arbitrary two-state system, they cannot be solved in general. They are also too detailed, requiring full information about the macroscopic history. At this point the experiment has to inform us about the most useful and meaningful phenomenological model, that is, the minimal model that captures the macroscopic switching dynamics of the system. More specifically, the experiment has to reveal whether the switching propensities do indeed depend on the current time t and the previous

switching times $t',t'',\dots$. The decision about the type of macroscopic expression that describes the experiment best, will then depend both on the physical properties of the system under consideration, and on the time resolution and the measurement uncertainty of the experiment. Below, we will discuss a number of stationary and non-stationary situations for which Eqs. 2 and 3 can be simplified, and for which microscopic expressions for the macroscopic rate constants can be found.

We now turn to the microscopic characterization of the system. We consider an ensemble of trajectories that start at $t=0$ at state space points x_0, following a specified initial phase-space distribution function $\rho(x_0)$[4]. If the dynamics obey detailed balance and microscopic reversibility, and if $\rho(x_0)$ is the canonical distribution, then this is an equilibrium ensemble. We may instead also consider the relaxation towards thermodynamic equilibrium, starting from a non-equilibrium $\rho(x_0)$. Alternatively, the ensemble could be that of a non-equilibrium system that does not obey detailed balance and microscopic reversibility. Again, the initial condition could either be in stationary state, or out of stationary state, in which case the non-equilibrium system may relax back to a stationary state. Finally, the system (time-reversible or not) may never reach a stationary state over the time course $[0,T]$ of the experiment. This can happen if relaxation is slower than T, or if the system is driven externally via some protocol $\phi(t)$, for instance, proteins that are unfolded under the influence of an external force.

To derive microscopic expressions that are consistent with the macroscopic rate equations, Eqs. 2 and 3, we first need to define functions $h_A(x_t)$ and $h_B(x_t)$ that indicate whether the system with configuration x_t at time t is in state A or B. It is useful to define these characteristic functions in terms of an order parameter $q(x_t)$ that serves to measure the progress of the transition between the states A and B:

$$h_A(x_t) = \theta[q_A - q(x_t)], \quad (4)$$
$$h_B(x_t) = \theta[q(x_t) - q_B]. \quad (5)$$

Here, θ is the Heaviside step function. With these definitions, the system is considered to be in A when $q(x_t) < q_A$, in B when $q(x_t) > q_B$, and in the transition region C otherwise. Moreover, we have the relation

$$h_A + h_B \equiv 1, \text{ if } q_A = q_B = q^*. \quad (6)$$

The same relation holds effectively if $q_A < q_B$ and the occupancy of the transition region C is low for all times.

The principal idea is now that the macroscopic rates $k_{AB,BA}(t)$ can be derived from the behavior of the microscopic correlation function

$$C(t) = \int \mathrm{d}x_0\mathrm{d}x_t\rho(x_0)p(x_t|x_0)h_B(x_t)$$
$$= \langle h_B(x_t)\rangle \quad (7)$$

Here, $p(x_t|x_0)$ is the probability that the system is in state x_t at time t given that it started in state x_0 at

time zero; $\langle \dots \rangle$ denotes an average over the ensemble of trajectories that start in x_0, following the phase-space distribution $\rho(x_0)$. The correlation function $C(t)$ gives the probability that the system is in state B at time t given that it started from the initial distribution $\rho(x_0)$, which may or may not be chosen to fully reside in A. The transition rates are derived from the flux into B and out of B, and are thus related to the time derivative of $C(t)$:

$$\dot{C}(t) = \langle \dot{h}_B(t) \rangle, \tag{8}$$

where we use the shorthand notation $h_B(t) = h_B(x_t)$. The task at hand is to rewrite this microscopic expression such that it can be identified with the macroscopic expressions of Eqs. 2 and 3. This is carried out below for a number of different classes of systems. In each case, we start with the macroscopic rate equation, and then derive a microscopic correlation function from which the macroscopic rates can be obtained.

III. MARKOV SYSTEMS

In this section we consider systems that exhibit Markovian macroscopic switching dynamics. By this we mean that it is known from experiment that there exists a macroscopic time resolution Δt, on which the propensity to switch between the two macroscopic states A and B at time t is independent of the history of the macroscopic switching dynamics for earlier times $t' < t$. In this case the rate kernels are independent of the previous switching times, yet may depend on the current time if the system is not time-homogeneous: $k_{X\bar{X}}(t|t', t'', \dots) = k_{X\bar{X}}(t)$, where $X\bar{X} = AB$ or BA. We immediately conclude that the macroscopic rate equations Eqs. 2, 3 reduce to

$$\frac{\mathrm{d}}{\mathrm{d}t} P_B(t) = k_{AB}(t) P_A(t) - k_{BA}(t) P_B(t) \tag{9}$$

with rate functions $k_{AB}(t), k_{BA}(t)$. This equation together with its counterpart for P_A forms a closed set of equations for $P_{A,B}$. We remark that the form of Eq. 9 can always be obtained trivially from Eqs. 2, 3 by marginalizing. However, the resulting expression will then in general only be valid for a particular experiment. A Markov system is characterized by the property that Eq. 9 with a *fixed* set of rate functions $k_{X\bar{X}}(t)$ correctly describes multiple experiments, differing in their initial conditions and in their history before $t = 0$.

The fact that memory is lost on the macroscopic time scale Δt implies that the system locally equilibrates in the macroscopic states A and B on time scales τ_A and τ_B, respectively, that are shorter than Δt. We can then identify a transient time scale

$$\tau_{\mathrm{trans}} = \max\{\tau_A, \tau_B, \tau_C\} < \Delta t \ll \min\{\tau_{AB}, \tau_{BA}\} \tag{10}$$

which is the microscopic memory time of the system.

It is instructive to restrict the attention to trajectories that happen to start in A at $t = 0$. To derive microscopic

expressions for the macroscopic rate constants, we then consider the correlation function

$$\dot{C}(t) = \langle \dot{h}_B(t) \rangle_{A_0}, \tag{11}$$

where the subscript A_0 indicates restriction of $\rho(x_0)$ to this subensemble. We now rewrite this expression such that its terms can be identified with the macroscopic quantities appearing in Eq. 9. Taking $q_A = q_B = q^*$ and letting $h_X(t) = h_X(x_t)$, we can insert $h_A(t - \Delta t) + h_B(t - \Delta t) = 1$ and $h_A(t + \Delta t) + h_B(t + \Delta t) = 1$ into Eq. 11; the importance of Δt will become clear shortly. This yields

$$\begin{aligned}
\dot{C}(t) = &\langle h_A(t - \Delta t) \dot{h}_B(t) h_A(t + \Delta t) \rangle_{A_0} \\
&+ \langle h_A(t - \Delta t) \dot{h}_B(t) h_B(t + \Delta t) \rangle_{A_0} \\
&+ \langle h_B(t - \Delta t) \dot{h}_B(t) h_B(t + \Delta t) \rangle_{A_0} \\
&+ \langle h_B(t - \Delta t) \dot{h}_B(t) h_A(t + \Delta t) \rangle_{A_0}. \tag{12}
\end{aligned}$$

We now condition on the state prior to a transition, by multiplying and dividing the first two terms by $\langle h_A(t - \Delta t) \rangle_{A_0}$ and the last two terms by $\langle h_B(t - \Delta t) \rangle_{A_0}$. This gives

$$\begin{aligned}
\dot{C}(t) = &\langle h_A(t - \Delta t) \rangle_{A_0} \langle \dot{h}_B(t) h_A(t + \Delta t) \rangle_{A_0, A_{t - \Delta t}} \\
&+ \langle h_A(t - \Delta t) \rangle_{A_0} \langle \dot{h}_B(t) h_B(t + \Delta t) \rangle_{A_0, A_{t - \Delta t}} \\
&+ \langle h_B(t - \Delta t) \rangle_{A_0} \langle \dot{h}_B(t) h_B(t + \Delta t) \rangle_{A_0, B_{t - \Delta t}} \\
&+ \langle h_B(t - \Delta t) \rangle_{A_0} \langle \dot{h}_B(t) h_A(t + \Delta t) \rangle_{A_0, B_{t - \Delta t}} \tag{13}
\end{aligned}$$

The average $\langle \dots \rangle_{A_0, X_{t - \Delta t}}$ runs over trajectories that start in A at time $t = 0$ and are in $X = A, B$ at time $t - \Delta t$.

Since the system looses memory on the transient time scale τ_{trans}, the memory fo the initial state is lost for $(t - \Delta t) > \tau_{\mathrm{trans}}$, so that $\langle \dots \rangle_{X_{t - \Delta t}, A_0} = \langle \dots \rangle_{X_{t - \Delta t}}$. In this regime Eq. 13 becomes

$$\begin{aligned}
\dot{C}(t) = &\langle h_A(t - \Delta t) \rangle_{A_0} \langle \dot{h}_B(t) h_A(t + \Delta t) \rangle_{A_{t - \Delta t}} \\
&+ \langle h_A(t - \Delta t) \rangle_{A_0} \langle \dot{h}_B(t) h_B(t + \Delta t) \rangle_{A_{t - \Delta t}} \\
&+ \langle h_B(t - \Delta t) \rangle_{A_0} \langle \dot{h}_B(t) h_B(t + \Delta t) \rangle_{B_{t - \Delta t}} \\
&+ \langle h_B(t - \Delta t) \rangle_{A_0} \langle \dot{h}_B(t) h_A(t + \Delta t) \rangle_{B_{t - \Delta t}}. \tag{14}
\end{aligned}$$

We now rewrite the first and third term in this equation, which describe trajectories that re-cross into their original state $X = A, B$ after a time $2\Delta t$:

$$\begin{aligned}
\dot{C}(t) = &\frac{\langle h_A(t - \Delta t) \rangle_{A_0}}{\langle h_A(t - \Delta t) \rangle} \langle h_A(t - \Delta t) \dot{h}_B(t) h_A(t + \Delta t) \rangle \\
&+ \langle h_A(t - \Delta t) \rangle_{A_0} \langle \dot{h}_B(t) h_B(t + \Delta t) \rangle_{A_{t - \Delta t}} \\
&+ \frac{\langle h_B(t - \Delta t) \rangle_{A_0}}{\langle h_B(t - \Delta t) \rangle} \langle h_B(t - \Delta t) \dot{h}_B(t) h_B(t + \Delta t) \rangle \\
&+ \langle h_B(t - \Delta t) \rangle_{A_0} \langle \dot{h}_B(t) h_A(t + \Delta t) \rangle_{B_{t - \Delta t}} \\
\equiv &\, j_{AA}(t; \Delta t) + j_{AB}(t; \Delta t) \\
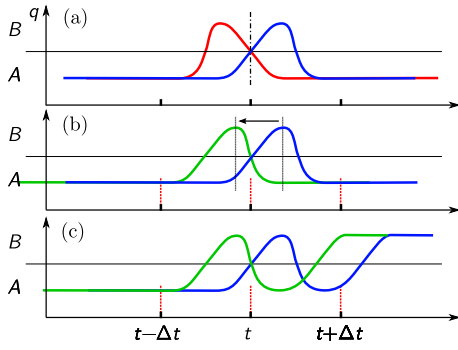&+ j_{BB}(t; \Delta t) + j_{BA}(t; \Delta t). \tag{15}
\end{aligned}$$

Figure 1. A visual interpretation of the term $j_{AA} = \langle h_A(t - \Delta t)\dot{h}_B(t)h_A(t+\Delta t)\rangle_{A_0}$ of Eq. 12. In microscopically reversible systems (a), for each trajectory (blue) there exists an identical but time-reflected trajectory (red). Consequently, $j_{AA} = 0$ for all values of Δt. In non-equilibrium systems that are memoryless and in stationary state (b), for each trajectory (blue) there exists a time-shifted trajectory (green) which cancels the contribution to j_{AA}, so that $j_{AA} = 0$ as well, irrespective of time-reversibility. In systems that are not time-reversible and remain correlated over Δt (c), time-shifted trajectories may fail to contribute to j_{AA}, as is the case for the trajectory shown in green. Here cancellation cannot occur and $j_{AA} \neq 0$. See Appendix A.

If the system obeys microscopic reversibility, and if the initial condition $\rho(x_0)$ coincides with the canonical distribution, then $\langle \ldots \rangle$ denotes an equilibrium average. In this case the recrossing terms

$$j_{AA}(t, \Delta t) = 0, \; j_{BB}(t, \Delta t) = 0 \qquad (16)$$

for all times t and for all values of Δt. This can be seen by noting that the equilibrium average $\langle h_X(x_{t-\Delta t})\dot{h}_B(x_t)h_X(x_{t+\Delta t})\rangle$ changes sign under time reversal and therefore vanishes in thermodynamic equilibrium. Out of equilibrium, the recrossing terms are non-zero in general. However, under the additional condition that over the given time scale Δt the system is approximately time-homogeneous and memoryless, Eq. 16 can be shown to hold nonetheless (Appendix A). Fig. 1 gives a graphical account of these observations.

Accepting Eq. 16 for the moment (this will be justified case-by-case below), Eq. 15 reduces to

$$\dot{C}(t) = \langle h_A(t - \Delta t)\rangle_{A_0}\langle \dot{h}_B(t)h_B(t+\Delta t)\rangle_{A_{t-\Delta t}}$$
$$+ \langle h_B(t - \Delta t)\rangle_{A_0}\langle \dot{h}_B(t)h_A(t+\Delta t)\rangle_{B_{t-\Delta t}} \quad (17)$$

We can now relate the microscopic Eq. 17 to the macroscopic Eq. 9 by identifying $\langle h_B(t-\Delta t)\rangle_{A_0} = P_B(t-\Delta t)$, $\langle h_A(x_{t-\Delta t})\rangle_{A_0} = P_A(t-\Delta t)$ and $\dot{C}(t) = \mathrm{d}P_B(t)/\mathrm{d}t$. This yields

$$\frac{\mathrm{d}}{\mathrm{d}t}P_B(t) = P_A(t - \Delta t)\langle \dot{h}_B(t)h_B(t+\Delta t)\rangle_{A_{t-\Delta t}}$$
$$+ P_B(t - \Delta t)\langle \dot{h}_B(t)h_A(t+\Delta t)\rangle_{B_{t-\Delta t}},$$
$$\simeq P_A(t)\langle \dot{h}_B(t)h_B(t+\Delta t)\rangle_{A_{t-\Delta t}}$$
$$+ P_B(t)\langle \dot{h}_B(t)h_A(t+\Delta t)\rangle_{B_{t-\Delta t}}, \qquad (18)$$

where the approximate equality results from the fact that Δt is the time resolution of the macroscopic description. Comparing this expression with the macroscopic rate equation Eq. 9, we deduce that

$$k_{AB}(t) = \langle \dot{h}_B(t)h_B(t + \Delta t)\rangle_{A_{t-\Delta t}},$$
$$k_{BA}(t) = -\langle \dot{h}_B(t)h_A(t + \Delta t)\rangle_{B_{t-\Delta t}}. \qquad (19)$$

In the next section, we recapitulate for completeness the classical scenario of a Markov system that relaxes towards stationary state starting from a non-stationary distribution[3]. In the subsequent section we consider Markov systems that are driven via an external protocol. In both cases, we will first consider the phenomenological rate equation, and then demonstrate that the macroscopic rate constants $k_{AB}(t), k_{BA}(t)$ are given by the microscopic expressions on the right-hand side of Eq. 19, provided Δt is chosen carefully, as discussed below.

A. Time-homogeneous Markov systems

We imagine that the Markov system has been driven via some protocol into an arbitrary initial condition $\rho(x_0)$ at $t = 0$, and consider its relaxation towards stationary state. We further imagine that experiments have revealed that the phenomenological rate equation, Eq. 9, can be reduced to

$$\frac{\mathrm{d}}{\mathrm{d}t}P_B(t) = k_{AB}P_A(t) - k_{BA}P_B(t), \qquad (20)$$

with rate constants k_{AB} and k_{BA}; thus k_{AB} is the (constant) propensity that the system switches to B, given that it is in A. The transitions between the states A and B on the macroscopic time scale Δt are described by a time-homogeneous Markov process. We note that even a system with time-homogeneous microscopic dynamics may fail to obey Eq. 20, if the system starts from a highly non-equilibrium initial condition; indeed, the experiment has to reveal whether the assumption of constant rates actually holds.

The solution of Eq. 20 for a system that starts in state A at time $t = 0$ is given by

$$P_B(t) = P_B^\infty(1 - e^{-t/\tau_{\mathrm{rxn}}}), \qquad (21)$$

where $\tau_{\mathrm{rxn}} = (k_{AB} + k_{AB})^{-1}$ is the macroscopic relaxation time of the switch and P_B^∞ is the probability of being in B in the stationary state, given by $P_B^\infty = k_{AB}/(k_{AB} + k_{BA})$.

Following the discussion above in this special case, we first need to establish that the recrossing fluxes vanish, Eq. 16. This holds whenever the system is both memoryless and time-homogeneous for time differences larger than Δt. These conditions are met in the present case, because Eq. 20, which is assumed to describe the experiment, implies that on the time scale Δt the system switches between A and B in a memoryless fashion with

constant rates. It then follows that the rates are given by Eq. 19:

$$k_{AB} = \langle \dot{h}_B(t) h_B(t + \Delta t) \rangle_{A_{t-\Delta t}},$$
$$k_{BA} = -\langle \dot{h}_B(t) h_A(t + \Delta t) \rangle_{B_{t-\Delta t}}. \tag{22}$$

While k_{AB} and k_{BA} on the left-hand side are independent of time, the expressions on the right-hand side appear to depend on t and Δt.

We first discuss the dependence on t. In writing down our macroscopic rate equation, we have *assumed* that the system switches between the states A and B with constant rates. This presupposes that the system relaxes inside the basins of A and B in between the switching events faster than the macroscopic time Δt, leading to Markovian switching dynamics and to loss of memory of the initial condition. Importantly, this assumption also implies that during relaxation towards stationary state, the probability of being in either A or B will change with time, but the state-space distribution *within* each macroscopic state A and B does not change with time; by deduction, the state-space distribution during relaxation, conditioned on being in either A or B, must be equal to that of the stationary distribution. In other words, while the ensemble brackets $\langle \ldots \rangle_{A_0}$ denote an average over trajectories that start from an arbitrary initial condition $\rho(x_0)$ in state A, this ensemble must, according to our assumption, be effectively equal to the stationary ensemble of trajectories, conditioned on starting in A.

This is essentially the content of Onsager's regression hypothesis[5,6], which states that the relaxation of an observable in a non-equilibrium experiment is proportional to the relaxation of a spontaneous fluctuation of that observable in the equilibrium system. It relies on the idea that the non-equilibrium initial distribution in a relaxation experiment follows a phase-space distribution that is similar to that of a spontaneous fluctuation in the equilibrium system[7]. The regression hypothesis is a well-known theorem for systems close to thermal equilibrium[8], but generalizations also exist for relaxation of perturbations towards non-equilibrium stationary state[9–12]. For rare switching events in equilibrium and stationary non-equilibrium systems, the presence of a dynamical bottleneck for switching means that the system can rapidly relax inside the basins of A and B, leading to loss of memory of the initial condition. This implies that the regressions theorem also holds, with exponential time-dependence of the relaxation function, at least on a time resolution Δt coarser than the transient time τ_{trans}.

We now turn to the dependence on Δt. Since the rates are constant, we can time-shift Eq. 22 to give:

$$k_{AB} = \langle \dot{h}_B(0) h_B(\Delta t) \rangle_{A_{-\Delta t}},$$
$$k_{BA} = -\langle \dot{h}_B(0) h_A(\Delta t) \rangle_{B_{-\Delta t}}. \tag{23}$$

These expressions, which are very similar to those of the reactive flux method of Bennett[13] and Chandler[3], hold only for a certain range of Δt values[7]. The macroscopic rate constant k_{AB} is defined as the (constant) propensity that the system switches from A to B given that at that moment in time it is in A. In contrast, the microscopic expression $\langle \dot{h}_B(0) h_B(\Delta t) \rangle_{A_{-\Delta t}}$ is the propensity that the system switches from A to B minus the propensity that it switches from B to A at a certain moment in time, given that at an earlier time $-\Delta t$ it was in A and at a later time Δt it is in B: $\langle \dot{h}_B(0) h_B(\Delta t) \rangle_{A_{-\Delta t}} = \langle \theta[\dot{h}_B(0)] \dot{h}_B(0) h_B(\Delta t) \rangle_{A_{-\Delta t}} - \langle \theta[-\dot{h}_B(0)](-\dot{h}_B(0)) h_B(\Delta t) \rangle_{A_{-\Delta t}}$. Indeed, the microscopic expression takes into account that the system may switch back and forth between the two states a number of times, and in fact, for $\Delta t \to \infty$, $\langle \dot{h}_B(0) h_B(\Delta t) \rangle_{A_{-\Delta t}}$ becomes equal to $\langle \dot{h}_B(\infty) \rangle \langle h_B(\infty) \rangle = 0$. Clearly, the macroscopic rate constant can only become equal to the microscopic expression if Δt is chosen to be smaller than the typical waiting time, $\Delta t \ll \tau_{AB}, \tau_{BA}$.

On the other hand, Δt cannot be made arbitrarily small. For $\Delta t \to 0$, Eqs. 23 reduce to the transition-state approximation

$$k_{AB}^{\text{TST}} = \langle \theta(\dot{q}) \dot{q} \delta(q - q^*) \rangle,$$
$$k_{BA}^{\text{TST}} = -\langle \theta(-\dot{q}) \dot{q} \delta(q - q^*) \rangle. \tag{24}$$

Transition-state theory assumes that every trajectory that crosses the dividing surface from A to B will end up in B before it returns to A on a time scale τ_{BA}. However, trajectories may also recross the dividing surface a (large) number of times before they settle in the new state B or the original state A. These correlated recrossings tend to decrease the correlation function $\langle \dot{h}_B(0) h_B(\Delta t) \rangle_{A_{-\Delta t}}$ as Δt is increased from zero. If the recrossings can be resolved experimentally, and if one wishes to characterize them, then Eq. 20 is not an appropriate model and the rate constants have to be defined differently; we will discuss this scenario in more detail in the next section. Here, we have been assuming that Eq. 20 is an appropriate model. In this model, the system equilibrates on time scales $\tau_A < \Delta t$ and $\tau_B < \Delta t$ inside the states A and B, before switching out of these states on a much longer time scale τ_{AB} and τ_{BA}, respectively. This implies that the correlation functions $\langle \dot{h}_B(0) h_B(\Delta t) \rangle_{A_{-\Delta t}}$ and $\langle \dot{h}_B(0) h_B(\Delta t) \rangle_{B_{-\Delta t}}$ reach a plateau for Δt in the range $\tau_{\text{trans}} < \Delta t \ll \tau_{AB}, \tau_{BA}$.

Figs. 2 and 3 illustrate these ideas for a particle performing a random walk in a piecewise-flat double-well potential according to Metropolis Monte Carlo dynamics (e.g.[14]). The macroscopic trajectory, shown in Fig. 2b, exhibits long committed episodes, but also short spikes caused by transient recrossings of the dividing surface $q^* = q_A = q_B$. While the correlation function $C(t) = \langle h_B(t) \rangle_{A_0}$ for this system shown in Fig. 3a rises exponentially for macroscopic times, its time derivative (Fig. 3b,c) does show a sudden drop for $t \lesssim \tau_{\text{trans}} \approx 10$; this is due to the rapid correlated re-crossings of the diving surface q^*. Importantly, for $t_{\text{trans}} < t < \tau_{\text{rxn}}$ the derivative of the correlation function exhibits a clear plateau. In this regime, the flux of trajectories from A to
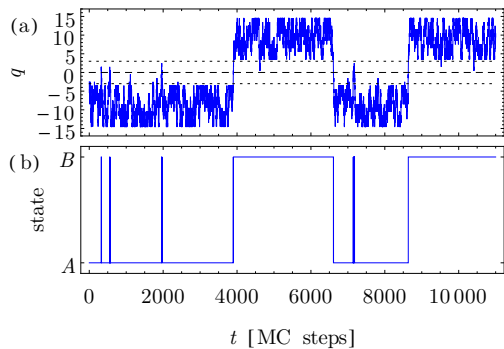
Figure 2. Microscopic (a) and macroscopic (b) trajectories in a simple model system of diffusion over a flat potential barrier. A particle starts at $q = -2.5$ and then performs a random walk visiting half-integer lattice points, with reflecting boundaries at $q = \pm 15$. The particle experiences a flat potential barrier given by $U(q)/k_B T = 3\theta(2 - |q|)$; diffusive steps are proposed in either direction and accepted according to a standard Metropolis criterion. The dividing surface defining the states A and B is located at $q_A = q_B = q^* = 0$.
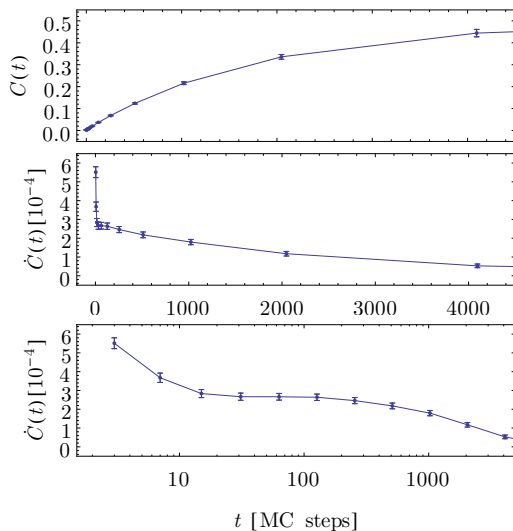

Figure 3. The correlation function $C(t)$, Eq. 7 and its time derivative for the model system of Fig. 2. Clearly on long times on the order of $\tau_{\text{rxn}} \simeq 1800$ MC steps the system relaxes exponentially. For very short times $t \lesssim 10$, the behavior is dominated by correlated transient re-crossings made by trajectories on the barrier top. A plateau emerges between these two times, whose value $\dot{C} \simeq 2.78 \times 10^{-4}$ equals the rate constant k_{AB}.

B is constant. Indeed, the rate constant k_{AB} is precisely given by the value of $\dot{C}(t)$ in this regime.

In summary, we reiterate that we have not derived the macroscopic rate equation from microscopic laws. By positing the macroscopic rate equation Eq. 20, we have made the implicit assumption that the system relaxes insides the basins and looses memory of its dynamics on

the experimental time scale Δt. The same assumption also entails that the non-stationary ensemble of trajectories is similar to the stationary ensemble of trajectories, conditioned on starting in A. In other words, if the macroscopic rate equation gives an accurate description of the switching dynamics, then this strongly suggests that the Onsager regression theorem holds as well.

B. Externally driven Markov systems

We now consider time-inhomogeneous Markov systems. We imagine that from $t = 0$ the Markovian system is under the influence of some time-dependent external force $\phi(t)$ that biases it towards one state or the other. The macroscopic rate equation thus has the general form of Eq. 9, and is valid for times longer than a macroscopic time Δt as inferred from experiment. The general solution can be written as

$$P_B(t) = P_B(0)e^{-\int_0^t \frac{\mathrm{d}t'}{\tau_{\text{rxn}}(t')}} + \int_0^t k_{AB}(t')e^{-\int_{t'}^t \frac{\mathrm{d}t''}{\tau_{\text{rxn}}(t'')}}\mathrm{d}t',$$ (25)

where we define $\tau_{\text{rxn}}^{-1}(t) = k_{AB}(t) + k_{BA}(t)$.

Examples of externally driven Markov systems include protein unfolding driven by a time-dependent force and (crystal) nucleation at a time-varying temperature. This class also contains systems where one (or more) of the degrees of freedom relaxes slowly on the time scale of a switching event, such as gene networks where at a given time a gene is turned on, which then induces the flipping of a genetic switch. The protein that drives the flipping of the switch could then be viewed as the external force $\phi(t)$ that acts on the switch. We consider external protocols with finite band-width such that there exist a scale of fastest force variation τ_ϕ and a scale of slowest variation T_ϕ, where T_ϕ is at most the duration of the experiment, $T_\phi \leq T$.

What microscopic scenarios are compatible with the macroscopic rate law, Eq. 9? The rate law implies that the system looses microscopic memory beyond Δt, so that $\tau_{\text{trans}} < \Delta t$ must hold. As before, we restrict our attention to systems with a time scale separation between intrastate relaxation and global relaxation, $\tau_{\text{trans}} \ll \tau_{AB,BA}$, where τ_{AB} and τ_{BA} are the mean waiting times for switching[15]. Under these constraints, there still is a variety of possible relations between the time scales of driving and the intrinsic microscopic time scales of the system.

The simplest situation is the quasi-static case: While the state-space distribution evolves with $\phi(t)$ in between the switching events, the system can locally (i.e. within the macroscopic states) adapt to the force, meaning that $\tau_{\text{trans}} < \tau_\phi$. Importantly, this quasi-static forcing scenario does not require the force to be slower than the global relaxation; $\tau_\phi, T_\phi \gtrless \tau_{AB}, \tau_{BA}$. We then expect that it should be straightforward to define time-dependent rate functions in terms of microscopic correlation functions.

To derive the microscopic expressions that are consistent with the macroscopic Eq. 20, we first note that in this quasi-static case we may choose Δt such that $\tau_{\text{trans}} < \Delta t < \tau_\phi, \tau_{\text{rxn}}$. On this time scale Δt the macroscopic switching dynamics is memoryless and approximately time-homogeneous, since the system adapts instantaneously to the force. Thus Eq. 16 applies and the expressions Eqs. 14-18 hold unchanged. We then indeed arrive at Eq. 19. The rates now depend on time t, but since the force is quasi-static, the dynamics in each Δt interval is governed only by the value of ϕ. Time dependence thus enters only via $\phi(t)$:

$$k_{X\bar{X}}(t) = k_{X\bar{X}}(\phi(t)), \text{ if } \tau_{\text{trans}} < \Delta t < \tau_\phi. \quad (26)$$

Furthermore, when Δt is in the range $\tau_{\text{trans}} < \Delta t < \tau_\phi, \tau_{\text{rxn}}$, we can expect that the correlation functions in Eq. 19 are constant as a function of Δt, in analogy to the time-homogeneous case.

Fig. 4 illustrates these relations in a version of the barrier-diffusion toy model, which is now augmented with an oscillatory force with a single frequency, $f(t) = ak_BT\sin(2\pi t/\tau_\phi)$ with $\tau_\phi = T_\phi = 400$, $a = 0.1$. We choose $\Delta t = 20$ so that $\tau_{\text{trans}} < \Delta t \ll \tau_\phi, \tau_{\text{rxn}}$. This separation of time scales suggests that we can treat this time-inhomogeneous Markov system as quasi-stationary, so that $j_{AA} \simeq 0$ (Eq. 16), and time enters the rates only via the force. Indeed, we observe that the recrossing fluxes approximately vanish, and that at the times when the force crosses zero, the rate constant equals that of the time-homogeneous Markov system of Fig. 3, supporting this idea.

An alternative simple scenario arises in the opposite case of rapid driving, where the external force variations are faster than the local relaxation within the macroscopic states, i.e. $\tau_\phi < T_\phi < \tau_{\text{trans}}$. We can then choose a macroscopic resolution $\Delta t > \tau_\phi, T_\phi, \tau_{\text{trans}}$, so that the macroscopic description effectively averages both over microscopic correlations of the system and over the variations in driving. In this case the macroscopic switching dynamics is described by the phenomenological rate equation of the time-homogeneous system, Eq. 20, although the magnitudes of the rate constants, which are independent of Δt, will be renormalized by the time-varying force. Moreover, because the system is effectively time-homogeneous and switches in a memoryless fashion between A and B, again Eq. 16 holds, so that the rate constants are given by the plateau values of the correlation functions in Eqs. 23 in the regime $\tau_\phi, \tau_{\text{trans}} < \Delta t \ll \tau_{AB}, \tau_{BA}$.

In the previous two scenarios we were able to choose Δt to be both well-separated from the driving time scales and within a well-defined plateau of the derivative of the microscopic correlation function; this allowed us to extract the rate as $k_{AB} = \dot{C}(\Delta t)$, irrespective of the precise value of Δt. We now consider a scenario in which this is not possible. Suppose that a certain macroscopic time resolution Δt is required, which still leads to Markovian switching ($\tau_{\text{trans}} < \Delta t$) but lies *within* the frequency
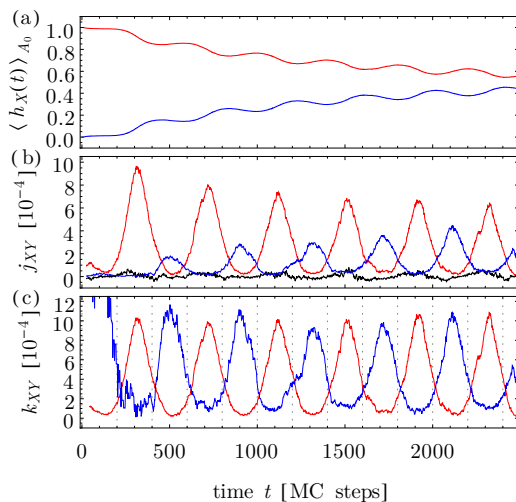

Figure 4. Time-dependent version of the system of Fig. 2, under the influence of an external force. The system is started in stationary state within A and relaxes with a time dependent modulation (a; $\langle h_A\rangle$, blue, $\langle h_B\rangle$, red). Fluxes defined in Eq. 12 (b; j_{AB}, blue, j_{BA}, red, j_{AA}, black) reflect the state occupancies. The rate functions, Eq. 19 (c; colors as in b) are independent of state occupancy. At the zeros of the force (dotted lines) the time dependent rates coincide with the rate constant of the stationary system, $k_{AB} = 2.78 \times 10^{-4}$ (see Fig. 3), in agreement with a quasi-stationary description on the time scale of the driving.

band of the driving, $\tau_\phi < \Delta t < T_\phi$. For instance, we may wish to describe a non-stationary experiment carried out with a given time resolution equal to Δt.

In the case that the microscopic relaxation of the system is the fastest process, $\tau_{\text{trans}} < \tau_\phi$ as in the quasi-stationary case studied above, then the correlation function does exhibit a plateau, but in a regime $\tau_{\text{trans}} < \widetilde{\Delta t} < \tau_\phi$ below the macroscopic resolution Δt. While this does allow us to uniquely define a force-dependent rate constant as $k_{AB}(\phi) = \dot{C}(\widetilde{\Delta t})$, this rate constant is inaccessible in experiments on the time scale Δt. The macroscopic rate constant as measured in the experiment now depends on the experimental time resolution Δt:

$$k_{AB}^{\text{eff}}(t) \simeq (\Delta t)^{-1} \int_{t-\Delta t/2}^{t+\Delta t/2} k_{AB}(\phi(t'))\mathrm{d}t'. \quad (27)$$

Alternatively, when the driving contains components that are faster than microscopic relaxation, $\tau_\phi < \tau_{\text{trans}}$, the system is Markovian on the time scale Δt but retains memory of the driving over times τ_{trans}. In this case ($\tau_\phi < \tau_{\text{trans}} < \Delta t < T_\phi$) the correlation function $C(t)$ does not exhibit a well-defined plateau on any time scale, and, as a result, a rate constant cannot be defined in a unique, protocol-independent manner. It is of course still possible to define a rate function in terms of a first passage time density, as a function of the external protocol; for analytical treatments in model systems with

simple driving protocols see for instance[16] and references therein.

IV. NON-MARKOVIAN SYSTEMS

For equilibrium systems, the two observations of rapid switching $\tau_C \ll \tau_{AB}, \tau_{BA}$ (Eq. 1) and rapid equilibration $\tau_A \ll \tau_{AB}; \tau_B \ll \tau_{BA}$ (Eq. 10) are often intimately connected. Rare events typically arise because of a single large free-energy barrier separating the two states A and B. As a consequence, the waiting time tends to be much longer than the switching event itself—making a two-state description meaningful—while it also allows the system to relax inside the basins of A and B in between the switching events—leading to memoryless switching on the macroscopic time scale Δt. In these equilibrium systems there is only one relevant macroscopic time scale τ_{rxn}, which is associated with the global relaxation of state occupancies. However, even for equilibrium systems, inertial effects may lead to correlated recrossings[17], possibly giving rise to non-Markovian switching dynamics on the macroscopic time scale Δt; if such correlated recrossings are important, then the macroscopic model of Eq. 20 needs to be refined.

One can readily find examples of non-equilibrium systems, however, that can be coarse grained as two-state systems, but in which switching between the two states occurs in a non-Markovian fashion even on time scales that are comparable to the typical waiting times τ_{AB}, τ_{BA}. This means that while the waiting time is much longer than the duration of the switching event, there exists another time scale in the system that is comparable to the waiting time for switching.

An example of such a non-Markovian, non-equilibrium system is the bacterial flagellar motor. This rotary motor can run in either clockwise or counterclockwise direction. The transitions between these two states happen much faster than the typical waiting time for switching, suggesting a two-state description. In experiments, the distribution of waiting times was observed to be non-exponential, and the power spectrum of the switching dynamics featured a peak around $1\mathrm{s}^{-1}$ [18], which means that there is a characteristic frequency at which the motor switches.

These data show that the flagellar motor cannot be described as a random telegraph process in which the switching events are independent, and the clockwise and counterclockwise intervals are uncorrelated and exponentially distributed; indeed, this system cannot be described by Eq. 20. The data also imply that motor switching is coupled to a non-equilibrium process[19–21]. It appears that the conformational dynamics of the rotor proteins is coupled to the slow relaxation dynamics of the flagellum, which in turn is influenced by the rotation of the motor, driven by a proton motive force[21].

To describe systems like the bacterial flagellar motor in which memory is important, we now consider non-

Markovian systems that when left unperturbed, reach a stationary state (equilibrium or not). As before, we imagine that the system flips between two macroscopic states A and B that are metastable in the sense that the waiting time between the switching events is much longer than the duration of the switching event itself. These system have memory on the macroscopic time scale Δt which means that the propensity of switching between the two macroscopic states at a given moment in time depends upon the history of the switching dynamics.

To derive microscopic expressions for the rate functions in the presence of memory, we start again with the phenomenological description of the system. If the propensity to switch depends upon the macroscopic history even prior to the last switching event, then Eq. 2 cannot be simplified much further, and remains of little use. However, it is conceivable that the switching propensity depends upon the macroscopic history only since the last switching event, to a good approximation. The bacterial flagellar motor is an example of such a system: when the experimental clockwise and counterclockwise intervals of the flagellar motor were randomly shuffled, the power spectrum was unchanged; moreover, the observed power spectrum could be reproduced from the measured waiting-time distributions only[18]. This strongly suggests that the different intervals are in fact temporally uncorrelated, so that the switching propensity only depends upon the time that has passed since the last switching event. We refer to this as the *clock-resetting* scenario—after each switching event, the system looses memory; in other words, switching is a renewal process. In what follows we will restrict our attention to this case. We do not exclude at this point the presence of external driving; in that case the system has only a single-switch memory but is time-inhomogeneous.

In the clock-resetting scenario, the phenomenological rate equation Eq. 2 can be simplified:

$$\frac{\mathrm{d}}{\mathrm{d}t}P_B(t) = \int_{t>t'} k_{AB}(t|t')P_A(t;t') - k_{BA}(t|t')P_B(t;t')\mathrm{d}t'. \tag{28}$$

Here, $P_A(t;t')\mathrm{d}t'$ is the joint probability that the system is in state A at time t and has switched into this state for the last time within the earlier interval $(t', t' + \mathrm{d}t')$. It can be decomposed as the propensity $q_{BA}(t')$ to enter A at t', times the survival probability $S_A(t|t')$ to stay in A until at least time t: $P_A(t;t') = q_{BA}(t')S_A(t|t')$.

The rate kernel $k_{AB}(t|t')$ is defined as the propensity that the system switches from A to B at time t given that it has switched into A at time $t' < t$ and is still in A at t. It is given by the propensity $q_{AB}(t|t') \equiv -\partial_t S_A(t|t')$ that a trajectory that entered A at time t', switches from A to B at a later time t, divided by the probability that it is still in A at time t: $k_{AB}(t|t') = q_{AB}(t|t')/S_A(t|t') = -\partial_t \ln S_A(t|t')$. If the system is time-homogeneous, i.e. without external time dependence, then we have $k_{AB}(t|t') = k_{AB}(t - t')$. If the system is in stationary state, then furthermore $P_B(t;t') = P_B(t - t')$. We note that this phenomenologi-

cal model can be solved analytically[21], see also Appendix C.

Before we derive microscopic expressions for the rate kernels that are consistent with Eq. 28, we first discuss the time scales that are relevant in such a system. A two-state description implies that the duration of the switching event is much shorter that the waiting time for switching: $\tau_C \ll \tau_{AB}, \tau_{BA}$. Moreover, we expect that inside the states A and B most degrees of freedom relax quickly, within τ_A and τ_B respectively, where $\tau_A \ll \tau_{AB}$ and $\tau_B \ll \tau_{BA}$; we can thus identify a fast transient time scale $\tau_{\text{trans}} \simeq \max\{\tau_C, \tau_A, \tau_B\}$. If these were the only time scales in problem, then the macroscopic switching dynamics could be described as that of a two-state Poisson process, assuming that we choose a macroscopic time scale $\Delta t > \tau_{\text{trans}}$. However, the observation that the macroscopic switching dynamics exhibits memory on the scale Δt, implies that there is another time scale in the problem, τ_{slow}, which is at least on the order $\tau_{\text{slow}} \gtrsim \Delta t$.

The bacterial flagellar motor provides a concrete illustration of these ideas: while the switching dynamics of the motor in the presence of the flagellum is non-Poissonian, the dynamics in the absence of the flagellum is Poissonian, with constant rates, leading to exponential waiting-time distributions[22]. The reason for this change in behavior is that the relaxation dynamics of the flagellum introduces a slow time scale, which becomes comparable to, and can even set, the typical waiting time for switching. After a switching event, the flagellum first unwinds, but then, driven by the rotation of the motor, winds up in the new direction; this leads to an increase in the force on the rotor proteins, which then tends to switch their conformation, and thereby the rotation direction of the motor[21].

We extend our diffusive-barrier crossing model to capture the non-equilibrium switching of the bacterial flagellar motor, see Fig. 5. The particle diffuses in a piecewise constant potential with a single flat barrier of height $4k_{\text{B}}T$ at $|q| < 2$; the state boundaries are at $q_A = q_B = q^* = 0$. Whenever the particle enters one of the two potential wells (say, at t'), the clock is reset and a restoring force ramps up over time, $f_{\text{clk}} = -\text{sgn}(q)f_0[1 - e^{-(t-t')/\tau_\phi}]$ with time constant $\tau_\phi = 100$ and saturating magnitude $f_0 = 0.8k_{\text{B}}T$. The addition of a restoring force leads to a faster global relaxation, which for the chosen value of f_0 is on the order of τ_ϕ. Fig. 5 shows that the relaxation of this system is not exponential: There is an initial lag time, and relaxation occurs in an oscillatory fashion, because the restoring force introduces a characteristic time scale for switching, and because particle clocks are taken to be synchronized at $t = 0$. We remark that an exponential relaxation observed in a single experiment would *not* by itself imply memoryless switching: For instance, the same system with unsynchronized clocks at $t = 0$ relaxes exponentially (not shown).

To derive microscopic expressions for the rate kernels that are consistent with the macroscopic rate equation of
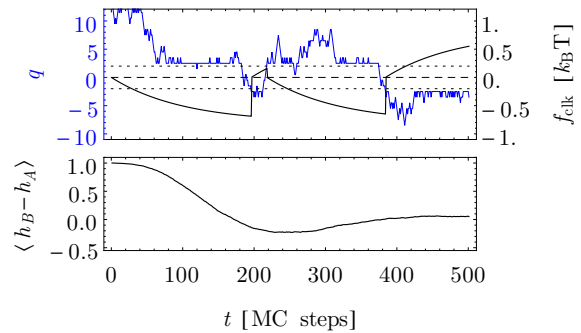

Figure 5. Particle diffusing with residence-time dependent restoring force. A trajectory (top, blue, left axis) reflects the restoring force (black, right axis). The relaxation of state occupancies when started uniformly in the B state basin (bottom) is oscillatory, and more rapid than without force (cf. Fig. 3).

Eq. 28, we need to define an indicator function that serves to measure the time since the last switching event. We let $H_X(t,t') \equiv \prod_{t \geq t'' > t'} h_X(t'')$. By construction, $H_X(t,t')$ takes the value 1 if the given trajectory was in $X = A, B$ without interruption from t' to t and 0 otherwise, and $H_X(t,t) = h_X(t)$. One also verifies (Appendix D) that

$$\partial_t H_X(t,t') = \dot{h}_X(t)H_X(t,t'), \text{ and}$$
$$\partial_{t'} H_X(t,t') = H_X(t,t')\dot{h}_X(t'). \quad (29)$$

The temporal derivatives of $H_X(t,t')$ with respect to t' and t (Eq. 29) thus make it possible to count the last entry time and the first exit time out of X, respectively. Although the definition of H may appear convoluted, it is straightforward to measure in a simulation, by recording switching times along trajectories. Interestingly an indicator function similar to H has been employed by Jung *et al*[23] to characterize fluctuations of trajectories in glassy systems.

Applying the relations of Eq. 29 and the identity $\dot{h}_A + \dot{h}_B = 0$, we can obtain the following relation:

$$\dot{h}_B(t) = \dot{h}_B(t) \sum_{X=A,B} H_X(t,t)$$

$$= \dot{h}_B(t)\Big[\sum_{X=A,B} H_X(t,t_0) + \int_{t_0}^t \partial_{t'} H_X(t,t')\text{d}t' \Big]$$

$$= \partial_t H_B(t,t_0) - \partial_t H_A(t,t_0) \quad (30\text{a})$$

$$+ \int_{t_0}^t \partial_t\partial_{t'} H_B(t,t') - \partial_t\partial_{t'} H_A(t,t')\text{d}t'. \quad (30\text{b})$$

Here the probability flux in and out of B is partitioned into contributions from trajectories that remained in one state since t_0 and then switch at t (Eq. 30a), and trajectories that had their last switching event at $t' > t_0$ before switching at t (Eq. 30b). The latter terms allow us to derive correlation functions which express the rate kernels depending on the last switching event.

To derive a microscopic expression that is consistent with the the macroscopic equation 28, we take the ensemble average $\langle \ldots \rangle$ of Eq. 30 (see also Eq. 8). The ensemble average is defined by the initial distribution $\rho(x_0)$ as before (although here we do not condition on starting in state A), and possibly by a prescription for $t < 0$, as discussed below. We now make the identifications

$$\langle h_B(t) \rangle = P_B(t), \tag{31a}$$

$$\partial_{t'} \langle H_X(t,t') \rangle \mathrm{d}t' = \langle H_X(t,t') \dot{h}_X(t') \rangle \mathrm{d}t'$$
$$= P_X(t;t') \mathrm{d}t', \tag{31b}$$

$$\frac{\partial_t \partial_{t'} \langle H_X(t,t') \rangle}{\partial_{t'} \langle H_X(t,t') \rangle} = \frac{\langle \dot{h}_X(t) H_X(t,t') \dot{h}_X(t') \rangle}{\langle H_X(t,t') \dot{h}_X(t') \rangle}$$
$$= -k_{X\bar{X}}(t|t'), \tag{31c}$$

where $X = A, B$ and $\bar{X} = B, A$. We can then rewrite the microscopic equation 30 in the form of Eq. 28.

$$\frac{\mathrm{d}}{\mathrm{d}t} P_B(t) = \partial_t \langle H_B(t,t_0) \rangle - \partial_t \langle H_A(t,t_0) \rangle \tag{32a}$$

$$+ \int_{t_0}^{t} k_{AB}(t|t') P_A(t;t') - k_{BA}(t|t') P_B(t;t') \mathrm{d}t'. \tag{32b}$$

The additional boundary terms Eq. 32a appear here due to a finite lower integration limit t_0. They can be treated in two ways. The first option is taking $t_0 \to -\infty$. Then the terms Eq. 32a vanish, because the system is not expected to stay in a given state forever. To extract rate constants from a microscopic measurement, we then have to be able to determine the integrand in Eq. 32b for negative t'. This will be possible for instance if we are dealing with a time-homogeneous system which was prepared in a stationary state for negative times, since then $P_X(t;t') = q_{\bar{X}X,\mathrm{ss}} S_X(t-t')$ for all $t' < 0$, with a constant, stationary-state value of the influx $q_{\bar{X}X,\mathrm{ss}}$; see also Appendix C.

Instead, for non-stationary systems one may wish to explicitly account for the switching events only after $t > 0$ without specifying the history before $t < 0$. Then one would let $t_0 \to 0$ and retain the boundary terms, Eq. 32a. They give a transient contribution to the macroscopic evolution of the system, which summarizes how the system was prepared before $t = 0$. We can incorporate this contribution by defining

$$\frac{\partial_t \langle H_X(t,0) \rangle}{\langle H_X(t,0) \rangle} = -k_{X\bar{X}}^{<0}(t), \tag{33a}$$

$$\langle H_X(t,0) \rangle = P_X^{<0}(t). \tag{33b}$$

In this case, Eq. 32 becomes

$$\frac{\mathrm{d}}{\mathrm{d}t} P_B(t) = k_{AB}^{<0}(t) P_A^{<0}(t) - k_{BA}^{<0}(t) P_B^{<0}(t) \tag{34a}$$

$$+ \int_0^t k_{AB}(t|t') P_A(t;t') - k_{BA}(t|t') P_B(t;t') \mathrm{d}t', \tag{34b}$$

which has an effective Markovian part (Eq. 34a) and a non-Markovian (Eq. 34b) part.

Scrutinizing the arguments above more carefully, we should expect that the system exhibits transient recrossings of the dividing surface q^* that are not persistent on the macroscopic time scale Δt. Moreover, these transient recrossings are not expected to lead to loss of memory: only macroscopic switching events, in which the systems fully transitions from one basin of attraction to another, are likely to reset the clock. This means that the identifications made in Eqs. 31, like the TST expressions Eqs. 24, are erroneous in that they overcount transient switching events. An additional point concerns the experimental time resolution. It is conceivable that the experimental resolution is sufficiently high that transient crossing events can be detected that do not reset the clock. If one wishes to describe these events, then the phenomenological model of Eq. 28, which only describes the crossing events that lead to loss of memory, needs to be modified. Here, we consider the case that Eq. 28 is an appropriate macroscopic model. The task is then to come up with a microscopic description that integrates over the transient recrossings of the dividing surface that do not reset the clock.

To integrate over the transient recrossings of the diving surface that separates the macroscopic states A and B, we must modify the indicator function $H_{X\bar{X}}$. This can be achieved by introducing a 'grace interval' Δt over which such short excursions are tolerated. Therefor we define a relaxed version of h_X as

$$h_X(t; \Delta t) = \theta\Big[\int_{t-\Delta t}^{t} h_X(t') \mathrm{d}t' - \Delta t/2\Big]; \tag{35}$$

this new indicator function takes the value 1 precisely if in the preceding short interval, the system was predominantly in state X; the arbitrary convention of using preceding and not e.g. centered intervals becomes unimportant on the macroscopic scale. Observe that, as for h_X, we have $h_A(t; \Delta t) + h_B(t; \Delta t) \equiv 1$ (we still take the states to be adjacent, $q_A = q_B = q^*$); the previous definition $h_X(t)$ is recovered in the limit $\Delta t \to 0$. Using this coarse-grained state indicator function, the indicator for uninterrupted residence in a state is then defined as before,

$$H_X(t, t'; \Delta t) = \prod_{t \geq t'' > t'} h_X(t''; \Delta t). \tag{36}$$

This expression remains 1 for instance if all excursions from X between $t' - \Delta t$ and t are shorter than $\Delta t/2$. By definition, $H_X(t, t'; \Delta t)$ is a monotonically decreasing function of t. Furthermore, $H_X(t, t; \Delta t) = h_X(t; \Delta t)$, and $H_X(t, t'; \Delta t) \to H_X(t, t')$ as $\Delta t \to 0$.

Fig. 6 shows the average uninterrupted state occupancies for the model system introduced above. While the state occupancy $H_A(t, t; \Delta t) = h_A(t; \Delta t)$ displays the oscillatory global relaxation in this system, $H_X(t, t'; \Delta t)$ decays with increasing residence time $t - t'$. This decay is slower when Δt is finite due to the fact that brief excursions from the state are tolerated.
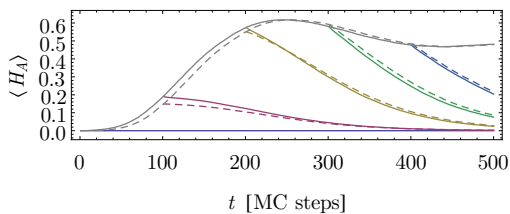
Figure 6. Uninterrupted state occupancies $\langle H_A(t,t';\Delta t)\rangle$ for the system from Fig. 5, for $\Delta t = 0, 20$ (solid and dashed lines, respectively). The system starts in state B and accumulates in A with occupancy $\langle H_A(t,t;\Delta t)\rangle - \langle h_A(t;\Delta t)\rangle$, shown in gray. Crossings back to B then decrease $\langle H_A(t,t';\Delta t)\rangle$, shown in colors for fixed values $t' = 0, 100, 200, 400$ from left to right.

With the new definitions for h and H, the equalities in Eq. 30 still hold. Therefore, also Eq. 32 holds unchanged if we make the new identifications

$$\langle h_B(t,\Delta t)\rangle = P_B(t), \tag{37a}$$

$$\partial_{t'}\langle H_X(t,t';\Delta t)\rangle \mathrm{d}t' = P_X(t;t';\Delta t)\mathrm{d}t', \tag{37b}$$

$$\frac{\partial_t \partial_{t'}\langle H_X(t,t';\Delta t)\rangle}{\partial_{t'}\langle H_X(t,t';\Delta t)\rangle} = -k_{X\bar{X}}(t|t';\Delta t), \tag{37c}$$

replacing Eqs. 31.

If the system exhibits transient recrossings on a time scale $\tau_{\text{trans}} \ll \tau_{\text{slow}}$, then $k_{X\bar{X}}(t|t';\Delta t)$ should become independent of Δt if we choose Δt to be in the range $\tau_{\text{trans}} < \Delta t < \tau_{\text{slow}}$. Moreover, we can expect that with this choice of Δt, the rate constant $k_{X\bar{X}}(t|t';\Delta t)$ agrees with the experimentally measured rate constant $k_{X\bar{X}}(t|t')$ when the experimental time resolution is indeed of the order Δt, such that the transient recrossings are not detected, while the presence of a slow time scale is detected.

These observations are illustrated in Fig. 7, for the clock-resetting model system introduced earlier. We recall that the quantity $\langle \partial_{t'} H_A(t,t';\Delta t)\rangle = P_A(t;t';\Delta t)$ measures the probability that the system enters A at time t' and stays in A at least until t; it is thus given by the influx propensity $q_{BA}(t';\Delta t)$ at t' times the survival probability in A, $S_A(t|t';\Delta t)$: $P_A(t;t';\Delta t) = q_{BA}(t';\Delta t)S_A(t|t';\Delta t)$. Fig. 7a shows that $P_A(t,t';\Delta t)$ decreases with $t - t'$ because $S_A(t|t';\Delta t)$ decreases with $t - t'$. $P_A(t;t';\Delta t)$ is non-monotonic as a function of t', reflecting the fact that the flux $q_{BA}(t';\Delta t)$ from B into A shows oscillations as the system equilibrates—these arise because the probability that the system is in B relaxes in an oscillatory fashion, see Figs. 5, 6.

The rate kernel $k_{AB}(t|t';\Delta t)$ is defined as the flux of trajectories that enter A at time t' and leave A for the first time at time t divided by the flux of trajectories that enter A at time t' and remain in A till time t, see Eqs. 31c, 37c. This conditioning removes the dependence on the start time t' if the system is time-homogeneous. Fig. 7b shows that $k_{AB}(t|t';\Delta t)$ indeed depends only on the time difference: $k_{AB}(t|t';\Delta t) = k_{AB}(t - t';\Delta t)$, so

that we can conclude that our model system is indeed time homogeneous. By the same token, we may now average data for the same $t-t'$ but different t', leading to the result shown in Fig. 7c. It is seen that $k_{AB}(t-t';\Delta t)$ starts out with a high crossing rate, passes through a minimum and then rises to saturate at a constant value. The initial drop is due to the rapid correlated recrossings of the diving surface $q_A = q_B = q^*$; importantly, the detection of these rapid recrossings is strongly suppressed when the grace interval Δt is chosen to be larger than the microscopic relaxation time τ_{trans}, $\Delta t \geq \tau_{\text{trans}} \approx 10$, as seen in panel c. The subsequent minimum in $k_{AB}(t-t';\Delta t)$ corresponds to local relaxation of the system inside the basin of the new macroscopic state. The rate kernel remains small as long as the force that tends to flip the system back to the other state on a time scale τ_ϕ is still low. Finally, as the switching force increases with time, the rate k_{AB} increases, to reach a plateau value when the switching force becomes constant. Fig. 7c also shows that $k_{AB}(t-t';\Delta t)$ becomes independent of Δt when $\Delta t > \tau_{\text{trans}} \approx 10$ (and $t-t' > \Delta t$). This is crucial: it allows us to define the macroscopic rate function $k_{AB}(t-t')$, and it justifies *a posteriori* the two-state description with a rate kernel that depends on the time since the last switching event.

V. DISCUSSION

It is instructive to compare the microscopic expression for the rate kernel in a non-Markovian system, Eq. 37, with that for the rate constant in a Markovian system, Eq. 23. Both expressions count only those switching events where the system goes from one macroscopic state to the other over a time $2\Delta t$. For a Markov system in stationary state, the recrossing flux $j_{AA} \propto \langle h_A(t-\Delta t)\dot{h}_B(t)h_A(t+\Delta t)\rangle = 0$, as discussed above (see also Appendix A). That is, if we chose to count recrossings in the microscopic expression for the rate function instead of using Eq. 23, this would not change the net result except for additional noise. In fact, the Bennett-Chandler expression $\langle \dot{h}_B(t)\rangle_{A_0} = \langle \dot{q}(0)\delta(q(0)-q^*)\theta(q(t)-q^*)\rangle/\langle h_A\rangle$ also counts trajectories that come from B, cross and recross the dividing surface (a number of times), and then go to B; however, their net contribution to the average is indeed zero[3].

For a non-Markovian system, $\langle h_A(t-\Delta t)\dot{h}_B(t)h_A(t+\Delta t)\rangle$ may be non-zero even in stationary state, due to the presence of memory over Δt. In contrast to the Markov case, here these transient recrossings should be explicitly excluded, for the reasons already mentioned above: If we take the macroscopic rate function to be defined as the propensity to switch, given the time that has passed since the last macroscopic switching event, then we should indeed only count the truly macroscopic switching events in Eq. 37.

Whether a switching event counts as a macroscopic switching event ultimately depends on the chosen macro-
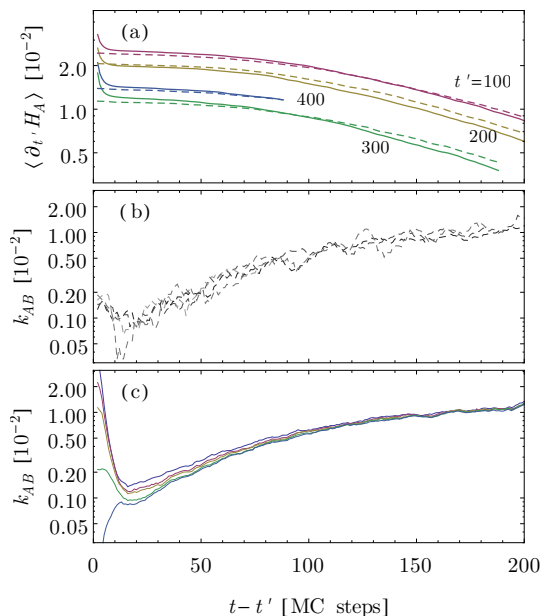
Figure 7. Time-dependent survival probabilities and rate kernels for the model system from Fig. 5. (a) The joint influx and survival probability $\langle \partial_{t'} H_A(t, t'; \Delta t)\rangle = P_A(t; t'; \Delta t) = q_{BA}(t') S_A(t|t')$ as a function of $t - t'$, shown for different values of t' (t'=0,100,200,400, as indicated) and for $\Delta t = 0, 10$ (solid and dashed lines, respectively). $P_A(t; t'; \Delta t)$ is the probability that the system enters A at time t' and still is in A at t. It decreases monotonically with $t - t'$, yet more slowly when the grace interval $\Delta t = 10$ is finite; but is non-monotonic in t', as explained in the text. (b) The rate kernel $k_{AB}(t|t'; \Delta t)$ as a function of $t - t'$, for different values of t', and for $\Delta t = 5$. It is seen that the rate kernels for the different times t' collapse. (c) The rate kernels $k(t - t'; \Delta t)$ averaged over t', for chosen time lags $\Delta t = 0, 2, 4, 10, 20$, from top to bottom. It is seen that for $t - t' > \Delta t$ and $\Delta t > \tau_{\text{trans}} \approx 10$, the rate kernels become independent of Δt. The macroscopic rate kernel can be identified with this limiting function.

scopic time resolution Δt. We can define a macroscopic switching event in an unambiguous manner if the time scale for recrossing τ_{trans}, is well separated from the slow memory time scale τ_{slow} of the system, so that $k(t|t'; \Delta t)$ is independent of Δt in the regime $\tau_{\text{trans}} < \Delta t < \tau_{\text{slow}}$. It may happen that the time scales τ_{trans} and τ_{slow} are not well separated; then, to match the experimentally measured rate constant, it becomes critical to tune Δt exactly to the experimental time resolution. However, the rate kernel defined in this way will then depend on the details of the experiment at least over some range of residence times $t - t'$, which limits its use as a description of intrinsic properties of the system. Thus in a non-Markovian system described by Eq. 28, a time scale separation between rapidly decaying transient memory and a well-defined slow memory time scale is again essential.

In this context, it is noteworthy that we have used con-

ventional indicator functions $h_A(t)$ and $h_B(t)$, instead of the indicator functions $h_{\mathcal{A}}(t)$ and $h_{\mathcal{B}}(t)$ as introduced by Van Erp and coworkers[24]. The latter are defined such that $h_{\mathcal{A}}(t)$ is 1 if the system was more recently in A than in B, and vice versa. The advantage of those indicator functions is that by a judicious choice of the dividing surfaces q_A and q_B, only the macroscopic switching events are counted, i.e. the crossing events where the system goes from one basin of attraction to the other[24]. If a clear separation exists between the transient time scale and the slow time scale, and macroscopic switching events can thus be defined uniquely (as discussed above), then the indicator functions $h_{\mathcal{A}, \mathcal{B}}$ are useful also in the framework presented here. For example, these indicator functions would then obviate the need to introduce a grace interval, as we have done for the non-Markovian system. The advantage of the conventional indicator functions is, however, that the corresponding correlation functions will inform us whether there exists such a clear separation of time scales: if so, the correlation functions exhibit a plateau, from which the macroscopic rate constant can be obtained.

Finally, now that we have derived microscopic expressions for macroscopic rate constants for both time-inhomogeneous Markov systems and non-Markov systems, the question arises whether these expressions can also be evaluated efficiently in a computer simulation. In the accompanying paper[25] we present a new numerical technique that makes this possible.

Appendix A: Vanishing recrossing fluxes

We consider the terms $j_{XX}(t) = \langle h_X(t - \Delta t)\dot{h}_B(t) h_X(t + \Delta t)\rangle$ appearing in Eq. 15, where $X = A, B$. Noting that (a) for $q_A = q_B$ we have $\dot{h}_B(t) + \dot{h}_A(t) = 0$, we can restrict our attention to the quantity $\tilde{j}(t) = \langle h_X(t - \Delta t)\dot{h}_X(t) h_X(t + \Delta t)\rangle$.

Assuming memory loss over time intervals Δt (b), and regarding $t^{\pm} = t \pm \Delta t$ as fixed, we can factor the joint probabilities as

$$\tilde{j}(t) = \partial_t \langle h_X(t^-) h_X(t) h_X(t^+)\rangle$$
$$= \partial_t(\langle h_X(t^-)\rangle \langle h_X(t)\rangle_{X_{t^-}} \langle h_X(t^+)\rangle_{X_t})$$
$$= \langle h_X(t^-)\rangle \partial_t \langle h_X(t)\rangle_{X_{t^-}} \langle h_X(t^+)\rangle_{X_t}. \quad (A1)$$

We further assume (c) that the system is time-homogeneous over the time Δt, in other words, any ex-

ternal driving, or background relaxation of slow degrees of freedom, happens much slower. Then we may rewrite the conditional probabilities as functions of the time difference,

$$\langle h_X(t) \rangle_{X_{t'}} = P_{X|X}(t - t'). \qquad (A2)$$

Then by the product rule

$$
\begin{aligned}
\tilde{\jmath}(t) &\propto \partial_t P_{X|X}(t - t^-) P_{X|X}(t^+ - t) \\
&= \dot{P}_{X|X}(\Delta t) P_{X|X}(\Delta t) - P_{X|X}(\Delta t) \dot{P}_{X|X}(\Delta t) \\
&= 0. \qquad (A3)
\end{aligned}
$$

By consequence both recrossing fluxes j_{XX} in Eq. 15 vanish.

Importantly, this conclusion rests on the three assumptions of (a) a transition region with negligible occupancy, (b) a time lag Δt long enough for any fast internal memory of the system to decay, and (c) at the same time short enough so that the system state is quasi-stationary with respect to any external driving or other slow relaxation dynamics.

Appendix B: Influence of the macroscopic time scale

Fig. 8 is analogous to Fig. 4b in the main text, but for choices of time lag Δt that are not well separated from either microscopic relaxation or external driving; the recrossing flux j_{AA} is seen to deviate from 0.
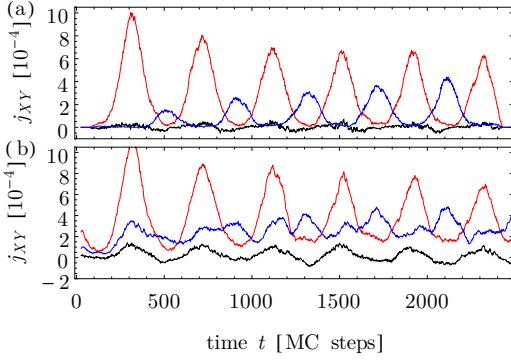

Figure 8. Variants of Fig. 4b, with Δt set to 100 (a) or 2 (b). In either case, the recrossing flux does not vanish; in (a) the force changes noticeably during Δt, while in (b) the system retains memory on the scale Δt, which is seen clearly in the higher harmonic components in j_{BA}. Rate functions are not given by Eq. 19.

Appendix C: Solution for the renewal rate equations

The rate equations 28 in the non-Markovian clock-resetting model may be rewritten in Matrix form as

$$\mathbf{q}(t) = \int_0^t -\partial_t \mathbf{S}(t|t') \mathbf{q}(t') \mathrm{d}t' - \partial_t \tilde{\mathbf{S}}(t) \mathbf{p}(0), \qquad (C1)$$

where

$$
\mathbf{S}(t|t') = \begin{bmatrix} 0 & S_A(t|t') \\ S_B(t|t') & 0 \end{bmatrix}, \quad \tilde{\mathbf{S}}(t) = \begin{bmatrix} \tilde{S}_A(t) & 0 \\ 0 & \tilde{S}_B(t) \end{bmatrix},
$$

$$
\mathbf{q}(t) = \begin{bmatrix} q_{AB}(t) \\ q_{BA}(t) \end{bmatrix}, \text{ and } \mathbf{p}(t) = \begin{bmatrix} P_A(t) \\ P_B(t) \end{bmatrix}.
$$

The net probability flux appearing on the left hand side of Eq. 28 is then recovered as $\partial_t P_B(t) = -\partial_t P_A(t) = q_{AB}(t) - q_{BA}(t)$. We are concerned with the time-homogeneous case only, so that we may write $\mathbf{S}(t|t') = \mathbf{S}(t - t')$.

The macroscopic pre-history for $t < 0$ in Eq. C1 is summarized in the $\tilde{\mathbf{S}}\mathbf{p}$ term. More precisely,

$$-\partial_t \tilde{\mathbf{S}}(t) \mathbf{p}(0) = \int_{-\infty}^0 -\partial_t \mathbf{S}(t - t') \mathbf{q}(t') \mathrm{d}t' \qquad (C2)$$

describes the first-switch propensies of trajectories that started in states A, B with probabilities $\mathbf{p}(0)$ at $t = 0$ (having arrived at earlier times). In the case that the system was held in a stationary state for all times $t' < 0$, it is possible to show that

$$\tilde{S}_X(t) = \frac{\int_{-\infty}^0 S_X(t - t') \mathrm{d}t'}{\int_{-\infty}^0 S_X(-t'') \mathrm{d}t''}. \qquad (C3)$$

To write a closed-form solution to Eq. C1, note that this is a convolution integral. After Laplace transformation ($f(s) = \int_0^\infty e^{-st} f(t)$) we may rewrite the equation as

$$
\begin{aligned}
\mathbf{q}(s) &= [-s\mathbf{S}(s) + \mathbf{S}(t = 0)]\mathbf{q}(s) \qquad (C4) \\
&\quad + [-s\tilde{\mathbf{S}}(s) + \tilde{\mathbf{S}}(t = 0)]\mathbf{p}(t = 0) \\
&= [-s\mathbf{S}(s) + \bar{\mathbf{1}}]\mathbf{q}(s) + [-s\tilde{\mathbf{S}}(s) + \mathbf{1}]\mathbf{p}(t = 0),
\end{aligned}
$$

where

$$\tilde{\mathbf{S}}(t = 0) = \mathbf{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \mathbf{S}(t = 0) = \bar{\mathbf{1}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

We may rearrange to give an explicit solution to the clock-resetting rate equation in Laplace space,

$$\mathbf{q}(s) = \left[\mathbf{1} - \bar{\mathbf{1}} + s\mathbf{S}(s) \right]^{-1} \left[\mathbf{1} - s\tilde{\mathbf{S}}(s) \right] \mathbf{p}(0). \qquad (C5)$$

Here the transient survival matrix $\tilde{\mathbf{S}}$ is either given as part of the specification of the problem or may be obtained from $\mathbf{S}$ via Eq. C3 in the case of stationary pre-history.

The stationary survival matrix **S** is given as the solution of the equation

$$\partial_t \mathbf{S}(t|t') = -\mathbf{k}(t|t')\mathbf{S}(t|t'); \; \mathbf{S}(t'|t') = \bar{\mathbf{1}}, \qquad \text{(C6)}$$

$$\text{where } \mathbf{k}(t|t') = \begin{bmatrix} k_{AB}(t|t') & 0 \\ 0 & k_{BA}(t|t') \end{bmatrix}, \qquad \text{(C7)}$$

which is just given by $\exp\left[-\int_{t'}^{t} \mathbf{k}(t''|t')\mathrm{d}t''\right]\bar{\mathbf{1}}$.

Appendix D: Definition of the uninterrupted indicator function H

When writing the function $H_A(t,t')$ as a continuous product, some purely mathematical subtleties arise. These are immaterial for the physics of two-state switching, and disappear when discretizing time as would effectively be done in either experiments or simulation.

Without restriction, we take the state indicator functions $h_X(t)$ to be right-continuous; i.e. we impose that $h_X(t) = h_X(t_+) = \lim_{\epsilon \searrow 0} h_X(t+\epsilon)$. This means in particular that the time derivatives are pre-point delta functions at the switching times: $\dot{h}_X(t) = \pm\delta(t - t_{\text{switch}-})$. It is then convenient to define

$$H_X(t,t') \equiv \prod_{t > t'' \geq t'} h_X(t''); \qquad \text{(D1)}$$

this definition makes $H_X(t,t')$ left-continuous in t and right-continuous in t': $H_X(t_-, t'_+) = H_X(t,t')$. Then the relations Eq. 29 given in the main text,

$$\partial_t H_X(t,t') = \dot{h}_X(t)H_X(t,t') \leq 0 \text{ and}$$
$$\partial_{t'} H_X(t,t') = H_X(t,t')\dot{h}_X(t') \geq 0, \qquad \text{(D2)}$$

are justified. Other conventions for the sample paths are possible but may require the insertion of limits in Eq. 29.

REFERENCES

[1] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after Kramers. *Reviews of Modern Physics*, 62(2):251, April 1990.

[2] Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D. Chodera, Christof Schütte, and Frank Noé. Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics*, 134:174105, 2011.

[3] David Chandler. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *The Journal of Chemical Physics*, 68(6):2959, 1978.

[4] If the microscopic dynamics is not Markovian, the initial condition will have to be supplemented with a prescription of how the system was prepared at $t < 0$.

[5] Lars Onsager. Reciprocal relations in irreversible processes. I. *Physical Review*, 37(4):405–426, February 1931.

[6] Lars Onsager. Reciprocal relations in irreversible processes. II. *Physical Review*, 38(12):2265–2279, December 1931.

[7] David Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, USA, 1 edition, September 1987.

[8] Herbert B. Callen and Theodore A. Welton. Irreversibility and generalized noise. *Physical Review*, 83(1):34–40, July 1951.

[9] G. S. Agarwal. Fluctuation-dissipation theorems for systems in non-thermal equilibrium and applications. *Zeitschrift für Physik*, 252:25–38, February 1972.

[10] Itamar Procaccia, David Ronis, and Irwin Oppenheim. Statistical mechanics of stationary states. IV. far-from-equilibrium stationary states and the regression of fluctuations. *Physical Review A*, 20(6):2533–2546, December 1979.

[11] J. Prost, J-F. Joanny, and J. M R Parrondo. Generalized fluctuation-dissipation theorem for steady-state systems. *Phys Rev Lett*, 103(9):090601, Aug 2009.

[12] U. Seifert and T. Speck. Fluctuation-dissipation theorem in nonequilibrium steady states. *EPL*, 89(1):6, 2010.

[13] C. H. Bennett. Molecular dynamics and transition state theory: The simulation of infrequent events. In *Algorithms for Chemical Computations*, volume 46 of *ACS Symposium Series*, pages 63–97. American Chemical Society, June 1977. 0.

[14] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications*. Academic Press, 2 edition, November 2001.

[15] Since the transition rates are now time-dependent, the definition of the waiting times for transitions, τ_{AB} and τ_{BA} is less clear-cut. A way to obtain these time scales is to take the average 'instantaneous waiting times' over the duration of the protocol. E.g. $\tau_{AB} = 1/T \int_0^T k_{AB}(t)^{-1}\mathrm{d}t$.

[16] Benjamin Lindner. Moments of the first passage time under external driving. *Journal of Statistical Physics*, 117(3-4):703–737, 2004.

[17] Titus S van Erp. Dynamical rare event simulation techniques for equilibrium and non-equilibrium systems. *1101.0927*, January 2011.

[18] Ekaterina A. Korobkova, Thierry Emonet, Heungwon Park, and Philippe Cluzel. Hidden stochastic nature of a single bacterial motor. *Phys Rev Lett*, 96(5):058105, Feb 2006.

[19] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, 2nd edition, 1992.

[20] Yuhai Tu and G. Grinstein. How white noise generates power-law switching in bacterial flagellar motors. *Phys Rev Lett*, 94(20):208101, May 2005.

[21] Siebe B. van Albada, Sorin Tănase-Nicola, and Pieter Rein ten Wolde. The switching dynamics of the bacterial flagellar motor. *Mol Syst Biol*, 5:316, 2009.

[22] Fan Bai, Richard W. Branch, Dan V Nicolau, Jr, Teuta Pilizota, Bradley C. Steel, Philip K. Maini, and Richard M. Berry. Conformational spread as a mechanism for cooperativity in the bacterial flagellar switch. *Science*, 327(5966):685–689, Feb 2010.

[23] YounJoon Jung, Juan P Garrahan, and David Chandler. Dynamical exchanges in facilitated models of supercooled liquids. *The Journal of Chemical Physics*, 123(8):084509, September 2005.

[24] Titus S. van Erp, Daniele Moroni, and Peter G. Bolhuis. A novel path sampling method for the calculation of rate constants. *The Journal of Chemical Physics*, 118(17):7762, 2003. TIS.

[25] Nils B Becker, Rosalind Allen, and Pieter Rein ten Wolde. Non-stationary forwad flux sampling. submitted to JCP.