

# Cycle counts and affinities in stochastic models of non-equilibrium systems

Patrick Pietzonka,<sup>1</sup> Jules Guioth,<sup>1,2</sup> and Robert L. Jack<sup>1,3</sup>

<sup>1</sup>*Department of Applied Mathematics and Theoretical Physics,  
University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom*

<sup>2</sup>*Univ. Lyon, ENS de Lyon, Univ. Claude Bernard,  
CNRS, Laboratoire de Physique, F-69342 Lyon, France*

<sup>3</sup>*Yusuf Hamied Department of Chemistry, University of Cambridge,  
Lensfield Road, Cambridge CB2 1EW, United Kingdom*

For non-equilibrium systems described by finite Markov processes, we consider the number of times that a system traverses a cyclic sequence of states (a cycle). The joint distribution of the number of forward and backward instances of any given cycle is described by universal formulae which depend on the cycle affinity, but are otherwise independent of system details. We discuss the similarities and differences of this result to fluctuation theorems, and generalize the result to families of cycles, relevant under coarse-graining. Finally, we describe the application of large deviation theory to this cycle counting problem.

## I. INTRODUCTION

Fluctuations in non-equilibrium systems continue to provide surprises and new insight in statistical physics. Among the most famous examples are fluctuation theorems [1–5], which come in different types. Some of them (detailed fluctuation theorems) are symmetries of probability distributions [1, 3, 4], while others allow the results of dynamical experiments to be related to (static) quantities such as free energies [2, 3]. More recently, thermodynamic uncertainty relations (TURs) have been derived [6–10], which are inequalities that relate the variances of physical observables to underlying properties of a system, particularly its entropy production.

These results reflect an elegant mathematical structure, that underpins the physical models to which they apply. They also have experimental relevance [11, 12]. Still, it is notable that fluctuation theorems for non-equilibrium steady states usually involve quantities such as affinities or the entropy production, which are difficult to characterise from experimental (or simulation) data. There is an ongoing effort to infer such quantities by TURs [13–15].

In this context, a recent result of Biddle and Gunawardena (BG) [16] offers a potentially new route towards inference of cycle affinities from data. Analyses of steady-state fluctuations often focus on currents, but BG’s results concern cycles – sequences of states that begin and end at the same point. They showed that for long times, the affinity of a cycle can be computed by counting the number of times that the cycle is traversed, in each direction. In their approach, completion of a cycle corresponds to the system visiting a particular set of states, in a particular order. This distinguishes their analysis from a different body of work, that involves decomposition of the full stochastic trajectory into a sequence of completed cycles [17–19]: in that case, a system’s progress around a single cycle may include other cycles within it, see also [20] for a recent example of this decomposition in practice. Yet another approach [4] is based on Schnaken-

berg network theory [21], where the probability currents between pairs of states are decomposed into a minimal set of cycle currents, see also [22, 23]. However, that construction does not provide information about any specific cyclic sequence of states.

Given their definition of a cycle as a specific sequence of states, the approach of [16] can be interpreted as a generalization to *continuous-time* Markov chains of a set of discrete-time problems, including counting the number of occurrences of a given word, in a random sequence of letters. Such problems are relevant for DNA sequence analysis [24–27]. There are also physical results for cycle counting in discrete-time, at least for unicyclic models [28].

In this paper, we follow [16], exploring in more detail the distribution of the number of times that cycles are traversed, in either direction. The results apply at all times. Hence they generalize the results of [16], which concern the mean of the distribution, for long trajectories. We also outline methods for analysis of large-deviation events [29–31], where the cycle count takes a non-typical value at very large times. As might be expected, the cycle affinity plays a central role in the statistics of cycle counts, especially for the cycle current, which is the difference in counts for forward and reverse cycles. By contrast, the statistics of the total count have a complex dependence on all model parameters, as expected for time-reversal symmetric (frenetic) quantities [32]. Some of the methods used here, particularly renewal theory and  $m$ th order Markov processes, are similar to those used for sequence analysis, although our results concern processes in continuous time.

The form of the paper is as follows: Sec. II defines the models and quantities of interest and Sec. III describes the main results for fluctuations of cycle counts, in finite time intervals. Sec. IV discusses these results, including some possible extensions, and the connection with fluctuation theorems. The relevant theory for large deviations of cycle counts is outlined in Sec. V. Finally, Sec. VI gives a short conclusion.

## II. MODEL AND DEFINITIONS

*Model* – We consider non-equilibrium systems that are modelled as Markov chains with discrete states, in continuous time. A very simple example system is shown in Fig. 1(a). The set of states is denoted by  $\Gamma$ , which in the example is  $\{A, B, C, D\}$ . The transition rate from state  $x$  to state  $y$  is denoted by  $w(x \rightarrow y) \geq 0$ . We emphasise that the results apply for any finite set  $\Gamma$  so they include complex systems like (finite) exclusion processes and are not at all restricted to simple models like the example of Fig. 1.

We restrict to irreducible systems and we also assume that if  $w(x \rightarrow y) > 0$  then also  $w(y \rightarrow x) > 0$  (which is sometimes called weak reversibility or microreversibility). This ensures that the steady state is unique and that every state  $x$  has a non-zero probability in the steady state, denoted by  $\pi(x)$ . Also the average entropy production rate is finite and non-negative.

*Trajectories* – A trajectory of the system can be specified for a time period  $[0, \tau]$ , an example is shown in Fig. 1(b). Let  $X$  denote a trajectory, it consists of a sequence of states

$$\mathbf{x}_X = (x_0, x_1, x_2, \dots, x_M) \quad (1)$$

and the associated transition times

$$\mathbf{t}_X = (t_1, t_2, \dots, t_M) \quad (2)$$

with  $0 < t_1 < t_2 < \dots < t_M < \tau$ . Here each  $x_i \in \Gamma$ , the system starts in state  $x_0$  and jumps to state  $x_i$  at time  $t_i$ . The number of transitions  $M$  is a random (trajectory-dependent) quantity.

*Words and cycles* – We consider sequences of states visited by the Markov chain. Similar objects have been studied for models in discrete time, especially in the context of DNA sequence analysis [24–27]. The continuous-time case is very similar, although it requires some additional book-keeping.

A sequence of states is called a word. We always restrict to words that can occur in trajectories of the system. (In the example of Fig. 1, AB or BAC would be suitable words, but ABD is excluded because  $w(B \rightarrow D) = 0$ .) A word that begins and ends with the same letter is called a *cycle*, and if  $\mathcal{C}$  is a cycle then  $\mathcal{C}^R$  denotes its time-reversal. For example if  $\mathcal{C} = ABCA$  then  $\mathcal{C}^R = ACBA$ . The cycle length  $m_{\mathcal{C}}$  is the number of transitions required to complete a cycle, so  $m_{\mathcal{C}} = 3$  in this example (one fewer than the corresponding word length).

Denote by  $\mathcal{C}_j$  the  $j$ th state in cycle  $\mathcal{C}$ , so  $1 \leq j \leq m_{\mathcal{C}} + 1$ . Then the cycle affinity for  $\mathcal{C}$  is

$$\mathcal{A}_{\mathcal{C}} = \sum_{j=1}^{m_{\mathcal{C}}} \ln \frac{w(\mathcal{C}_j \rightarrow \mathcal{C}_{j+1})}{w(\mathcal{C}_{j+1} \rightarrow \mathcal{C}_j)}, \quad (3)$$

which is also the entropy production for one cycle in the steady state. For allowed cycles then all rates are non-zero (by the weak reversibility property) so the affinity is finite.

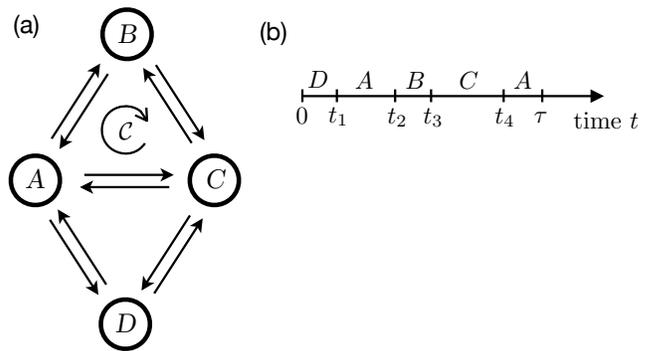


FIG. 1. (a) Diagram showing a simple system of four states, with transitions indicated by straight arrows. The cycle  $\mathcal{C} = ABCA$  is also indicated. (b) Example trajectory for a time period  $[0, \tau]$ , with jumps between states at times  $\mathbf{t}_X = (t_1, t_2, t_3, t_4)$ . The sequence of states is  $\mathbf{x}_X = (D, A, B, C, A)$  so  $n_{\mathcal{C}}(X) = 1$  and  $n_{\mathcal{C}^R}(X) = 0$ .

In a trajectory, it is convenient to define the *start time* of a word (or cycle) as the time of the jump between the first two states, and the *end time* as the time of the jump between the last two states. So for ABCA the word starts at the transition  $A \rightarrow B$  and ends at  $C \rightarrow A$ . The *completion time* is the difference between the start and end time.

As noted in Sec. I, this definition of a cycle [16] (which might also be called a *cyclic word*) – and the corresponding start/end times – differ from the definitions used in other works such as [19, 20, 33, 34]. Our definition leads to a simpler analysis, but some of the results and methods are similar.

*Counting cycles* – Given a trajectory  $X$  and a cycle  $\mathcal{C}$ , write  $\hat{n}_{\mathcal{C}}(X)$  for the number of occurrences of  $\mathcal{C}$  in  $X$ . This is the number of times that the word  $\mathcal{C}$  appears in the sequence  $\mathbf{x}_X$ . (The hat serves as a reminder that  $\hat{n}_{\mathcal{C}}$  is a random variable.) The cycle must appear exactly as in its definition, and different occurrences of the cycle may overlap. (For example the cycle ABABA appears twice in the sequence ABABABA.) Of course, generic words can be counted in the same way, not only cycles. It is convenient to write  $\hat{n}_{\mathcal{C}^R} = \hat{n}_{\mathcal{C}^R}$  for the number of occurrences of the reverse cycle.

*Non-revisiting cycles* – It will be convenient in the following to distinguish two kinds of cycle. Recalling that a cycle always begins and ends at the same point, we define a non-revisiting cycle as one in that does not return to its initial point, until the end. For example ABA is a non-revisiting cycle but ABCACBA is not. One sees that different occurrences of a non-revisiting cycle cannot overlap each other, and that a general cycle can be decomposed as the concatenation of non-revisiting cycles. The class of non-revisiting cycles is larger than that of simple cycles (for example ABCBCA is non-revisiting), but it is not as large as the class of non-overlapping cycles (or words) from [25].

### III. RESULTS: FINITE TIME

This Section contains some general results for the probability distribution of the number of cycle counts, for finite trajectories with time  $t \in [0, \tau]$ . We first summarise the results before giving the derivations. The analysis leading to these results is quite straightforward, but we argue that the results are interesting for two reasons: first, as a possible way to infer model parameters (specifically, affinities) from data [16]; and also as a starting point for more detailed analysis of cycle counts. Both these directions are discussed in later Sections.

#### A. Overview

Our results concern the random variables  $\hat{n}_{\mathcal{C}}, \hat{n}_{\mathcal{C}}^{\text{R}}$ , for cycles as defined above. In the palindromic case  $\mathcal{C} = \mathcal{C}^{\text{R}}$  then none of these results have any content, so we assume throughout that  $\mathcal{C} \neq \mathcal{C}^{\text{R}}$ . We are motivated by a result of [16], which is that for long trajectories

$$\lim_{\tau \rightarrow \infty} \frac{\hat{n}_{\mathcal{C}}(X)}{\hat{n}_{\mathcal{C}}^{\text{R}}(X)} = e^{\mathcal{A}_{\mathcal{C}}} . \quad (4)$$

Such formulae require some care because the left hand side is a random (trajectory-dependent) quantity but the right hand side is deterministic: the equation holds in the same sense as a law of large numbers. The physical idea behind (4) is that the cycle affinity can be inferred by counting cycles that are traversed in forward and backward directions.

In the following, we derive several results, related to (4). First, the derivation of [16] can be easily generalized to obtain a result for steady-state averages over trajectories of finite length  $\tau$ , with arbitrary initial condition. The result is

$$\frac{\langle \hat{n}_{\mathcal{C}} \rangle}{\langle \hat{n}_{\mathcal{C}}^{\text{R}} \rangle} = e^{\mathcal{A}_{\mathcal{C}}} \quad (5)$$

where  $\langle \cdot \rangle$  indicates an average over trajectories of the system (the dependence of the cycle counts on  $X$  is implicit). This result states that cycles with positive affinity happen more often in the forward direction, as expected. As  $\tau \rightarrow \infty$  then  $\hat{n}_{\mathcal{C}}(X) \rightarrow \langle \hat{n}_{\mathcal{C}} \rangle$  with probability one, this is a weak law of large numbers. The proposal of [4] was that (4) might be used to infer affinities from data; in this case (5) seems also useful since long trajectories are not required.

We note in passing that the mean number of cycles is not a simple linear function of the trajectory length, that is  $\langle \hat{n}_{\mathcal{C}} \rangle \neq \tau \omega(\mathcal{C})$  in general. (Here  $\omega(\mathcal{C})$  would be a cycle completion rate.) The reason is that there is typically a significant lag time between starting and ending a cycle. So the fact that (5) applies for all  $\tau$  is not trivial.

We now consider the joint distribution of  $n_{\mathcal{C}}, n_{\mathcal{C}}^{\text{R}}$ , which we denote by  $P_{\tau}(n_{\mathcal{C}}, n_{\mathcal{C}}^{\text{R}})$ . Our results for this distribution are restricted to non-revisiting cycles, but they hold for

any trajectory length  $\tau$  and for any initial condition (it is not necessary that the probabilities are evaluated in the steady state of the system). We show that

$$n_{\mathcal{C}} P_{\tau}(n_{\mathcal{C}}, n_{\mathcal{C}}^{\text{R}} - 1) = e^{\mathcal{A}_{\mathcal{C}}} n_{\mathcal{C}}^{\text{R}} P_{\tau}(n_{\mathcal{C}} - 1, n_{\mathcal{C}}^{\text{R}}). \quad (6)$$

The physical origin of (6) is that replacing any non-revisiting cycle  $\mathcal{C}$  by its time-reversed counterpart  $\mathcal{C}^{\text{R}}$  changes the trajectory probability by a factor  $e^{-\mathcal{A}_{\mathcal{C}}}$ . The prefactors  $n_{\mathcal{C}}$  and  $n_{\mathcal{C}}^{\text{R}}$  are of combinatorial origin.

Also, it is straightforward to show that for non-revisiting cycles

$$P_{\tau}(n_{\mathcal{C}}, n_{\mathcal{C}}^{\text{R}}) = P_{\tau}(n_{\mathcal{C}}^{\text{R}}, n_{\mathcal{C}}) \exp[(n_{\mathcal{C}} - n_{\mathcal{C}}^{\text{R}})\mathcal{A}_{\mathcal{C}}] , \quad (7)$$

which has some similarities to the fluctuation theorem of Andrieux and Gaspard [4], see Sec. IV B.

Both (6) and (7) are direct consequences of a binomial structure in the distribution  $P_{\tau}(n_{\mathcal{C}}, n_{\mathcal{C}}^{\text{R}})$ . In order to state this property conveniently, identify the total number of cycles in trajectory  $X$  and the corresponding net flux as

$$\begin{aligned} \hat{K}_{\mathcal{C}}(X) &= \hat{n}_{\mathcal{C}}(X) + \hat{n}_{\mathcal{C}}^{\text{R}}(X) \\ \hat{J}_{\mathcal{C}}(X) &= \hat{n}_{\mathcal{C}}(X) - \hat{n}_{\mathcal{C}}^{\text{R}}(X). \end{aligned} \quad (8)$$

Denoting the joint distribution of these quantities by  $\tilde{P}_{\tau}(K, J)$  and the marginal of  $K$  by  $\tilde{P}_{\tau}(K) = \sum_J \tilde{P}_{\tau}(K, J)$ , we show in Sec. III C that the conditional distribution of  $\hat{J}$ , namely  $\tilde{P}_{\tau}(K, J)/\tilde{P}_{\tau}(K)$ , is binomial, so that

$$\tilde{P}_{\tau}(K, J) = \tilde{P}_{\tau}(K) \binom{K}{\frac{1}{2}(K+J)} \frac{\exp(J\mathcal{A}/2)}{[2 \cosh(\mathcal{A}/2)]^K} . \quad (9)$$

Here and in the following, we sometimes omit the label  $\mathcal{C}$  for variables and affinities, where there is no ambiguity. The key point of (9) is that the dependence of  $\tilde{P}_{\tau}$  on  $J$  is explicit. This formula holds for all models and for any non-revisiting cycle  $\mathcal{C}$ . In discrete time, a similar result is given in [28], for the restricted case of unicyclic models.

These results extend the analysis of [16] from the most likely number of cycles to its full fluctuation spectrum. However, they are restricted to non-revisiting cycles: we emphasise that (6,7) are derived from the more general (9), so this restriction is necessary for all these results. For such cycles, one may then recover previous results for the mean, in particular (5) is obtained by summing both sides of (6) over  $n_{\mathcal{C}}, n_{\mathcal{C}}^{\text{R}}$ .

It is useful to recall that the path weight of trajectory  $X$  in our general model is

$$P[X] = P_0(x_0) \left[ \prod_{i=0}^{M-1} w(x_i \rightarrow x_{i+1}) e^{-r(x_i)(t_{i+1}-t_i)} \right] \times e^{-r(x_M)(\tau-t_M)} \quad (10)$$

where  $t_0 = 0$ ; also  $P_0(x_0)$  is the (arbitrary) distribution of the initial state and we introduced the exit rate from state  $x$ , as

$$r(x) = \sum_{y \neq x} w(x \rightarrow y) . \quad (11)$$

From this formula, the result (6) can be anticipated by observing that any instance of a non-revisiting cycle  $\mathcal{C}$  in  $X$  can be replaced by an instance of  $\mathcal{C}^R$ , so that  $P[X]$  changes by a factor  $e^{-A_C}$ . A precise argument along these lines is given in Sec. III C, see also Fig. 2. Alternatively, this result [and also (9)] may be derived using renewal theory, see Appendix B.

### B. Average cycle counts

We now derive (5). Consider the probability that an instance of cycle  $\mathcal{C}$  starts between times  $t$  and  $t + \delta t$  and ends before time  $\tau$ . For small  $\delta t$  denote this by  $P_{\mathcal{C}}(t, \tau)\delta t$ . Considering trajectories for the time period  $[0, \tau]$ , the average  $\langle \hat{n}_{\mathcal{C}} \rangle$  may then be decomposed as

$$\langle \hat{n}_{\mathcal{C}} \rangle = \int_0^{\tau} P_{\mathcal{C}}(t, \tau) dt. \quad (12)$$

Moreover, the probability that a transition  $\mathcal{C}_1 \rightarrow \mathcal{C}_2$  occurs between times  $t$  and  $t + \delta t$  is  $p(\mathcal{C}_1, t)w(\mathcal{C}_1 \rightarrow \mathcal{C}_2)\delta t$ , where  $p(\mathcal{C}_1, t)$  is the probability that the system is in state  $\mathcal{C}_1$  at time  $t$ . Since every instance of cycle  $\mathcal{C}$  starts with such a transition, it follows that

$$P_{\mathcal{C}}(t, \tau) = p(\mathcal{C}_1, t)w(\mathcal{C}_1 \rightarrow \mathcal{C}_2)p_{\text{seq}}(\mathcal{C})F_{\text{fin}}(\tau - t, \mathcal{C}), \quad (13)$$

where

$$p_{\text{seq}}(\mathcal{C}) = \prod_{j=2}^{m_{\mathcal{C}}} \frac{w(\mathcal{C}_j \rightarrow \mathcal{C}_{j+1})}{r(\mathcal{C}_j)} \quad (14)$$

is the probability that the system follows the correct sequence of states, and the factor  $F_{\text{fin}}(\tau - t, \mathcal{C})$  in (13) is the probability that the cycle is completed in a time less than  $\tau - t$ . [This cycle completion time is a sum of  $m_{\mathcal{C}} - 1$  exponentially distributed random variables with means  $r(\mathcal{C}_2)^{-1}, r(\mathcal{C}_3)^{-1}, \dots, r(\mathcal{C}_{m_{\mathcal{C}}})^{-1}$ .] Hence

$$\langle \hat{n}_{\mathcal{C}} \rangle = \int_0^{\tau} \left[ p(\mathcal{C}_1, t)r(\mathcal{C}_1) \prod_{j=1}^{m_{\mathcal{C}}} \frac{w(\mathcal{C}_j \rightarrow \mathcal{C}_{j+1})}{r(\mathcal{C}_j)} \times F_{\text{fin}}(\tau - t, \mathcal{C}) \right] dt. \quad (15)$$

Repeating the same argument for the reversed cycle  $\mathcal{C}^R$  one finds

$$\langle \hat{n}_{\mathcal{C}^R} \rangle = \int_0^{\tau} \left[ p(\mathcal{C}_1, t)r(\mathcal{C}_1) \prod_{j=1}^{m_{\mathcal{C}}} \frac{w(\mathcal{C}_{j+1} \rightarrow \mathcal{C}_j)}{r(\mathcal{C}_{j+1})} \times F_{\text{fin}}(\tau - t, \mathcal{C}^R) \right] dt. \quad (16)$$

Note that the time to complete cycle  $\mathcal{C}^R$  is the same sum of exponentially distributed random variables as for  $\mathcal{C}$ , so

$$F_{\text{fin}}(\tau - t, \mathcal{C}^R) = F_{\text{fin}}(\tau - t, \mathcal{C}), \quad (17)$$

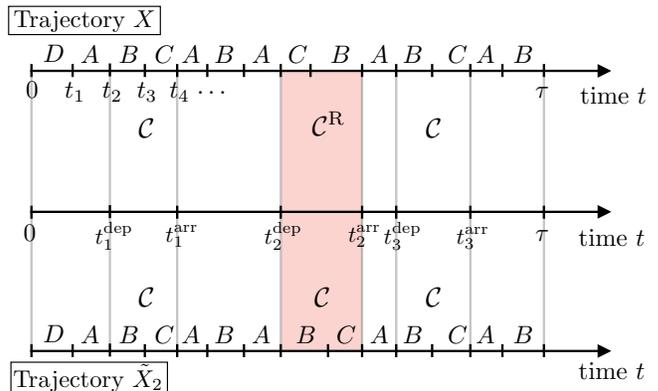


FIG. 2. (Top) A trajectory  $X$  of the four-state system in Fig. 1, and the corresponding sequence of completions of the cycle  $\mathcal{C}$  or  $\mathcal{C}^R$ . We keep track of the departure and arrival times for state  $A$  that enclose instances of  $\mathcal{C}$  or  $\mathcal{C}^R$ . The second such instance is highlighted in red. (Bottom) Applying time-reversal to the highlighted part of the trajectory leads to the trajectory  $\tilde{X}_2$ , transforming the cycle  $\mathcal{C}^R$  into  $\mathcal{C}$ .

see also [33, 34], an explicit formula for  $F_{\text{fin}}$  is given in (A4). Also, the fact that the cycle starts and ends at the same point means that  $\prod_{j=1}^{m_{\mathcal{C}}} r(\mathcal{C}_{j+1}) = \prod_{j=1}^{m_{\mathcal{C}}} r(\mathcal{C}_j)$ . Combining these facts with (3,15,16), one recovers (5).

Note that there is no restriction here to non-revisiting cycles, the physical reason is that (12) decomposes the mean number of cycles into a sum of independent averages. To see this, recall that the average number of cycles that start between time  $t$  and  $t + \delta t$  and end before time  $\tau$  is  $P_{\mathcal{C}}(t, \tau)\delta t$ . Integrating over  $t$  corresponds to summing these independent averages and gives the average number of completed cycles within the full trajectory. The possibility of overlapping cycles is important for fluctuations in their number, but not for the mean.

### C. Fluctuations for non-revisiting cycles

We now restrict to non-revisiting cycles, and we derive (6-9). We use a methodology similar to proofs of fluctuation theorems based on path weights [5], see Appendix B 1 for a derivation using concepts of renewal theory.

Suppose that the cycle  $\mathcal{C}$  of interest starts in state  $A$  (this does not lose any generality). For any trajectory  $X$  we can identify the sequence of completions of the cycle in either forward or backward direction, for example

$$\mathcal{S}[X] = (\mathcal{C}, \mathcal{C}^R, \mathcal{C}, \mathcal{C}, \mathcal{C}^R), \quad (18)$$

along with the sequence of start and end times of the cycles, denoted by  $(t_1^{\text{dep}}, t_2^{\text{dep}}, \dots)$  and  $(t_1^{\text{arr}}, t_2^{\text{arr}}, \dots)$  respectively, see Fig. 2. (The start/end times of the cycle correspond to the departure/arrival times from/to state  $A$ .) The probability to observe a specific sequence  $\mathcal{S}$  of forward and backward cycles within time  $\tau$  can be ex-

pressed as

$$\mathcal{P}_\tau(\mathcal{S}) = \sum_X P[X] \delta(\mathcal{S}, \mathcal{S}[X]), \quad (19)$$

where the sum runs over all trajectories of length  $\tau$ , suitably parameterised as a path integral. The function  $\delta(\mathcal{S}, \mathcal{S}') = 1$  if  $\mathcal{S} = \mathcal{S}'$ , and zero otherwise.

To derive (9), we first obtain formulae that relate the probabilities of specific trajectories; then we sum over (classes of) trajectories to obtain the distribution of  $(K, J)$ . Note that the sequence  $\mathcal{S}[X]$  consists of  $\hat{K}(X)$  elements. For every trajectory  $X$  we define a conjugate trajectory  $\tilde{X}_k$  as follows: If  $k \leq \hat{K}(X)$  then  $\tilde{X}_k$  is obtained from  $X$  by reversing in time the  $k$ th cycle in  $\mathcal{S}[X]$ , that is:

$$\tilde{X}_k(t) = \begin{cases} X(t_k^{\text{arr}} - t + t_k^{\text{dep}}), & \text{if } t_k^{\text{dep}} < t < t_k^{\text{arr}} \\ X(t), & \text{otherwise} \end{cases}. \quad (20)$$

See Fig. 2, which shows how  $\tilde{X}_k$  is obtained from  $X$  for  $k = 2$  by reversing the second instance of the cycle. Similar partial time-reversal operations have been considered before in the context of simple chemical reactions [33, 34] as well as in other works on cycle counting [19, 20]. It is convenient to extend this definition to include  $k > \hat{K}(X)$  by taking  $\tilde{X}_k(t) = X(t)$ .

Noting that the sojourn times in each state are unaffected by the partial time-reversal, we see from (10) and (3) that

$$\frac{P[X]}{P[\tilde{X}_k]} = \begin{cases} e^{\pm \mathcal{A}c}, & \text{if } k \leq \hat{K}(X) \\ 1 & \text{otherwise} \end{cases}, \quad (21)$$

where in the first case we take the plus sign if the  $k$ th entry in  $\mathcal{S}(X)$  is  $\mathcal{C}$ , and the minus sign if this entry is  $\mathcal{C}^{\text{R}}$ .

Now define  $\tilde{\mathcal{S}}_k[X] = \mathcal{S}[\tilde{X}_k]$ ; for example if  $\mathcal{S} = (\mathcal{C}, \mathcal{C}^{\text{R}}, \mathcal{C}, \mathcal{C}, \mathcal{C}^{\text{R}})$  and  $k = 2$  then  $\tilde{\mathcal{S}}_2 = (\mathcal{C}, \mathcal{C}, \mathcal{C}, \mathcal{C}, \mathcal{C}^{\text{R}})$ . The mapping between  $X$  and  $\tilde{X}_k$  is a bijection, which means that (19) can be expressed as

$$\begin{aligned} \mathcal{P}_\tau(\mathcal{S}) &= \sum_X P[\tilde{X}_k] \delta(\mathcal{S}, \mathcal{S}[\tilde{X}_k]) \\ &= \sum_X P[\tilde{X}_k] \delta(\mathcal{S}, \tilde{\mathcal{S}}_k[X]), \end{aligned} \quad (22)$$

where the first equality is obtained by relabelling the trajectories and the second uses the definition of  $\tilde{\mathcal{S}}$ . Using (21) yields

$$\mathcal{P}_\tau(\mathcal{S}) = e^{\pm \mathcal{A}c} \mathcal{P}_\tau(\tilde{\mathcal{S}}_k), \quad (23)$$

where we take the plus (or minus) sign if the  $k$ th entry in  $\mathcal{S}(X)$  is  $\mathcal{C}$  (or  $\mathcal{C}^{\text{R}}$ ), as in (21). (It is assumed that the number of entries in  $\mathcal{S}$  is at least as large as  $k$ .) The result (23) is a special example of a detailed fluctuation theorem [5].

Now, given a sequence  $\mathcal{S}$ , one may use (23) and successively replace all instances of  $\mathcal{C}^{\text{R}}$  by  $\mathcal{C}$  to obtain

$$\mathcal{P}_\tau(\mathcal{S}) = e^{-n_{\mathcal{C}}^{\text{R}} \mathcal{A}c} \mathcal{P}_\tau(\mathcal{C}, \mathcal{C}, \dots, \mathcal{C}) \quad (24)$$

where  $n_{\mathcal{C}}^{\text{R}}$  is the number of occurrences of  $\mathcal{C}^{\text{R}}$  in the sequence  $\mathcal{S}$ . Write  $K_{\mathcal{S}}$  for the number of entries in  $\mathcal{S}$ . Then the right hand side of (24) is the probability of  $K_{\mathcal{S}}$  forward cycles and no reverse ones, that is  $P_\tau(K_{\mathcal{S}}, 0)$  in the notation of (6). The probability  $P_\tau(K_{\mathcal{S}} - n, n)$  can be obtained by summing (24) over sequences  $\mathcal{S}$  with the requisite numbers of forward and reverse cycles: the number of elements in the sum is a binomial coefficient. One obtains

$$P_\tau(n_{\mathcal{C}}, n_{\mathcal{C}}^{\text{R}}) = \binom{n_{\mathcal{C}} + n_{\mathcal{C}}^{\text{R}}}{n_{\mathcal{C}}} e^{-n_{\mathcal{C}}^{\text{R}} \mathcal{A}c} P_\tau(n_{\mathcal{C}} + n_{\mathcal{C}}^{\text{R}}, 0). \quad (25)$$

Re-parameterisation in terms of  $K, J$  yields (9). As noted in Sec. III A, both (6) and (7) follow straightforwardly from this result.

Finally, we observe that while these results have been derived for non-revisiting cycles, this is not the most general case in which (23) and hence (9) apply. Eq. (23) does not apply for *all* cycles because if two instances of the same cycle can overlap each other, then it is not generally possible to reverse a single instance of the cycle, leaving all other instances unchanged. [For example, consider the (revisiting) cycle ABCABCA and a trajectory that contains the sequence ABCABCABCA.] A similar problem arises if an instance of  $\mathcal{C}$  can overlap with an instance of  $\mathcal{C}^{\text{R}}$ . The assumption of non-revisiting cycles is sufficient to ensure that such overlaps never occur and (23) holds, but this condition is not necessary. Classes of non-overlapping words are discussed (for example) in [25]; we avoid such issues here, to simplify the analysis.

## IV. DISCUSSION OF FINITE-TIME RESULTS

### A. Coarse-grained measurements: families of cycles

The results derived so far concern the statistics of completions of a given cycle  $\mathcal{C}$ , which is a specific sequence of states. Hence, any measurement of  $n_{\mathcal{C}}$  requires complete information about the trajectory of the system. Since we consider ‘‘mesoscopic’’ models that should be defined as coarse-grained representations of real physical systems, it is useful to consider how this requirement of complete information can be reconciled with a coarse-graining operation.

Note first that these results can be generalised to some situations where incomplete information is available. To see this, let  $\mathcal{F}$  represent a family (a set) of cycles, and let

$$\hat{n}_{\mathcal{F}}(X) = \sum_{\mathcal{C} \in \mathcal{F}} \hat{n}_{\mathcal{C}}(X) \quad (26)$$

be the total number of occurrences in trajectory  $X$  of all cycles  $\mathcal{C}$  from that family  $\mathcal{F}$ . Reversing all cycles in

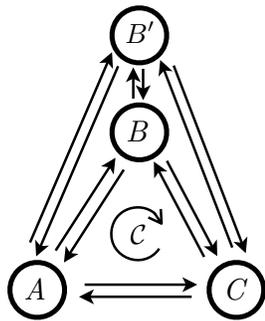


FIG. 3. A network where  $B$  and  $B'$  form a pair of sub-states. If the physical driving mechanism does not distinguish between these sub-states, variations of the cycle  $\mathcal{C}$  visiting either  $B$  or  $B'$  all have the same affinity and can be lumped together in a single family of cycles.

$\mathcal{F}$  yields the family  $\mathcal{F}^R$ , with the number of occurrences  $\hat{n}_{\mathcal{F}}^R$  defined analogously. (We assume that if  $\mathcal{C} \in \mathcal{F}$  then  $\mathcal{C}^R \notin \mathcal{F}$ .)

If all members of  $\mathcal{F}$  have the same affinity  $\mathcal{A}_{\mathcal{F}}$  then it is obvious that (5) still holds (with  $\hat{n}_{\mathcal{C}}, \hat{n}_{\mathcal{C}}^R, \mathcal{A}_{\mathcal{C}}$  replaced by  $\hat{n}_{\mathcal{F}}, \hat{n}_{\mathcal{F}}^R, \mathcal{A}_{\mathcal{F}}$ ). If all members of  $\mathcal{F}$  are also non-revisiting then (6-9) hold too. [This can be seen by constructing a modified sequence  $\mathcal{S}[X]$  in which the symbol  $\mathcal{C}$  represents a completion of any member of  $\mathcal{A}_{\mathcal{F}}$  and  $\mathcal{C}^R$  represents completion of the any member of the family  $\mathcal{F}^R$ . Then (23) holds and the analysis follows.]

A simple example of such a family is obtained by including repeated forwards and backwards steps within the cycle. For example, consider the family containing ABCA, ABCBCA, and all similar cycles obtained by repeatedly inserting instances of BC before the final A. All these cycles obviously have the same affinity and they are non-revisiting, so (5-9) still hold for the joint distribution of  $(\hat{n}_{\mathcal{F}}, \hat{n}_{\mathcal{F}}^R)$ . [To connect the results here with the framework of [18, 19], it is necessary to consider larger families, which include cycles that are constructed from a main (outer) cycle, and also include non-trivial subcycles; one should also extend the definition of a time-reversed cycle appropriately, so that only the main cycle is reversed, leaving the subcycles invariant. Such complex families are not our main concern in this work.]

Families of cycles with equal affinity also arise naturally in physical situations, especially where coarse-graining is considered. For example, suppose that a given state comes in two variants (perhaps  $B, B'$ ) which differ in a way that is irrelevant for the non-equilibrium driving force that controls the cycle affinity. Fig. 3 illustrates how this might appear in a simple model: there are two cycles that proceed via  $B, B'$  but have the same affinity (because the driving force is blind to the distinction between the states). Since these two cycles have the same affinity, they can be grouped into a family  $\mathcal{F}$  and (5-9) still hold for the combined counts. Moreover, the family could be extended by cycles that contain arbitrary numbers of forward and backward jumps between  $B$  and  $B'$ ,

which would become relevant when the transition rates between these to states are much faster than all other rates.

In this example, it is notable that the model may be coarse-grained exactly by combining the states  $B, B'$  into a single mesostate. As such, the example illustrates that the results presented here are consistent between different levels of coarse-graining. In fact, it is generally sufficient to observe the system on a coarse-grained level, as long as the coarse-graining does not mix cycles with different affinities. This mitigates the difficulty noted above, that the full trajectory of a system must be observed in order to apply our results.

## B. Relation to fluctuation theorems

We have emphasised the connection between the results (6-9) and fluctuation theorems [1–5]. As such, our derivations place the result (4) of Biddle and Gunawardena [16] in this context (under the restriction to non-revisiting cycles). The central result that enables this analysis is (23), which can be regarded as an instance of the “master fluctuation theorem” of Ref. [5], employing our partial time-reversal (20) as conjugate dynamics.

Nonetheless, the results (7,9) differ from usual fluctuation theorems, as they involve the total count of cycle completions in either direction, as well as the net flux around a cycle, see also results for the traffic or freeness [32, 35].

To connect to the more familiar case, note from (9) that  $\tilde{P}_{\tau}(K, J) = e^{J\mathcal{A}_{\mathcal{C}}} \tilde{P}_{\tau}(K, -J)$  and hence (summing both sides over  $K$ ):

$$\frac{\tilde{P}_{\tau}(J)}{\tilde{P}_{\tau}(-J)} = e^{J\mathcal{A}_{\mathcal{C}}}, \quad (27)$$

similar to (7). This result is reminiscent of the fluctuation theorem for currents by Andrieux and Gaspard [4], but there are several important differences.

In particular, (27) concerns counting observables for cycle completions: recall that  $n_{\mathcal{C}}$  and  $n_{\mathcal{C}}^R$  are the numbers of occurrences of specific sequences of states (for example  $\mathcal{C} = ABCA$  and  $\mathcal{C}^R = ACBA$ ) and  $J_{\mathcal{C}}$  is the difference between these numbers. On the other hand, the result of [4] concerns numbers of transitions between states (for example, one might consider a current defined as the difference between the number  $C \rightarrow A$  transitions and  $A \rightarrow C$  transitions). From these numbers of transitions, one defines cycle currents by an indirect method that involves a decomposition of steady-state current distributions in a basis that comes from Schnakenberg network theory [21].

We emphasise that the cycle currents in [4] are distinct objects from the counting observables for cycle completions that we consider here. For example, consider the model of Fig. 1: if we take ABCA and ACDA as the fundamental cycles in the sense of [4] (following the

Schnakenberg formalism) then the trajectory ABCDA would contribute +1 to each of the two *cycle currents* [4]. However, the trajectory does not complete either of these cycles in the exact sequence given, so both there are no *cycle completions* in the sense considered here (following [16]).

As a result of the indirect relationship between cycle currents and numbers of transitions, the fluctuation theorem of [4] appears as a symmetry of the joint distribution of *all* cycle currents. Moreover, the Schnakenberg theory applies to steady-state currents, which means that the result of [4] concerns the large-time limit of the current distribution. The result of [4] is a deep (and abstract) statement about the action of time-reversal on trajectories, and its implications for large deviations as  $\tau \rightarrow \infty$ . On the other hand, it does not generally imply a fluctuation theorem for the (marginal) distributions of currents associated with individual cycles [36–39].

By contrast, (27) is a much simpler result – it applies for all  $\tau$ , for individual cycles. The reason is that the cycle current is counted in a more direct way, by following the trajectory of the system throughout each instance of the cycle. Since the initial and final states of the cycle are always equal, replacing an instance of  $\mathcal{C}$  by  $\mathcal{C}^{\text{R}}$  in trajectory  $X$  has an effect on  $P(X)$  that is simple, and does not affect other parts of the trajectory.

For the very special case of a unicyclic network – and considering the family of cycles that include multiple forward and backward steps, as above – the fluctuation theorem of [4] follows from (23), in the long-time limit, see also [33, 34]. For multicyclic networks, the two results are distinct. Given that fluctuations of cycle-counting observables contain new information, it may be that these results – including that of Biddle and Gunawardena [16] – may prove useful for thermodynamic inference, following [40, 41]. For that purpose, it is likely that inference based on families of cycles is more practical than counting instances of a specific cycle; for example, counting cycles within a family will typically result in larger observed numbers, improving the statistics.

## V. LARGE DEVIATIONS AS $\tau \rightarrow \infty$

Given the connection to fluctuation theorems [4], and that the original result of [16] employed a large-time limit, it is useful to consider how cycle counting observables behave as  $\tau \rightarrow \infty$ . One may expect by ergodicity that the cycle completion rate  $\hat{n}_{\mathcal{C}}/\tau$  converges to its steady state average as  $\tau \rightarrow \infty$ , which would be consistent with (4,5). Large deviation theory provides a precise way to analyse this limit, and shows that this expectation is correct. The relevant large-deviation methods can be found in [29–31, 42, 43], we outline the theory here.

Define empirical time averages  $\bar{k} = \hat{K}_{\mathcal{C}}/\tau$  and  $\bar{j} = \hat{J}_{\mathcal{C}}/\tau$ : these are random (trajectory-dependent) quantities. Their joint probability density behaves for long

times as

$$P_{\tau}(\bar{k}, \bar{j}) \approx \exp[-\tau \mathcal{I}(\bar{k}, \bar{j})], \quad (28)$$

where  $\mathcal{I}$  is the rate function, which is non-negative. Such formulae are called large deviation principles – they show that the typical values of  $\bar{k}, \bar{j}$  occur with probability one (hence  $\mathcal{I} = 0$ ), while other values have probabilities that become exponentially small as  $\tau \rightarrow \infty$ . They have been analysed for a different type of cycle counts in [19].

The rate function may be characterised by the Gärtner-Ellis theorem as

$$\mathcal{I}(\bar{k}, \bar{j}) = \sup_{s, \lambda} [s\bar{k} + \lambda\bar{j} - \Psi(s, \lambda)], \quad (29)$$

where

$$\Psi(s, \lambda) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \left\langle \exp(s\hat{K} + \lambda\hat{J}) \right\rangle \quad (30)$$

is the scaled cumulant generating function (SCGF). Also, Varadhan’s lemma states that

$$\Psi(s, \lambda) = \sup_{\bar{k}, \bar{j}} [s\bar{k} + \lambda\bar{j} - \mathcal{I}(\bar{k}, \bar{j})]. \quad (31)$$

The marginal distribution for  $\bar{k}$  obeys

$$P_{\tau}(\bar{k}) \approx \exp[-\tau \mathcal{I}_1(\bar{k})] \quad (32)$$

with  $\mathcal{I}_1(\bar{k}) = \inf_{\bar{j}} \mathcal{I}(\bar{k}, \bar{j})$ , by the contraction principle.

A characterisation of  $\Psi$  will be given below, as the largest eigenvalue of a matrix. That analysis also ensures that the technical conditions required for (28) are satisfied, under our assumptions. Before that, we explore how (9) manifests itself in large deviations.

### A. Large deviations for non-revisiting cycles

For non-revisiting cycles, we note that for large  $K, J$  then (9) gives

$$\frac{1}{\tau} \log P_{\tau}(\bar{k}, \bar{j}) \approx \frac{1}{\tau} \log P_{\tau}(\bar{k}) - \mathcal{I}_2(\bar{j}, \bar{k}, \mathcal{A}) \quad (33)$$

where  $\mathcal{A}$  is the cycle affinity and (by Stirling’s approximation)

$$\begin{aligned} \mathcal{I}_2(\bar{j}, \bar{k}, \mathcal{A}) &= \bar{k} \log[\cosh(\mathcal{A}/2)] - \frac{\bar{j}\mathcal{A}}{2} \\ &+ \frac{\bar{k} + \bar{j}}{2} \log(\bar{k} + \bar{j}) + \frac{\bar{k} - \bar{j}}{2} \log(\bar{k} - \bar{j}). \end{aligned} \quad (34)$$

Then (28) implies that the rate function is

$$\mathcal{I}(\bar{k}, \bar{j}) = \mathcal{I}_1(\bar{k}) + \mathcal{I}_2(\bar{j}, \bar{k}, \mathcal{A}). \quad (35)$$

The function  $\mathcal{I}_2$  is closely related to the rate function for the time-averaged current  $\bar{j}$  of a biased random walk, which is related in turn to the binomial structure of (9).

Now define  $g(\lambda, \mathcal{A}) = \bar{k}^{-1} \sup_{\bar{j}} [\lambda \bar{j} - \mathcal{I}_2(\bar{j}, \bar{k}, \mathcal{A})]$  and observe that

$$g(\lambda, \mathcal{A}) = \log \frac{\cosh(\lambda + \mathcal{A}/2)}{\cosh(\mathcal{A}/2)}. \quad (36)$$

(It is important that this object does not depend on  $\bar{k}$ : while this is not obvious from its definition, it follows from the relationship of  $\mathcal{I}_2$  to a random walk.) Also define  $\Psi_1(s) = \Psi(s, 0)$  as the SCGF for  $\bar{k}$ . Then by (31) the (joint) SCGF  $\Psi$  has the simple form

$$\Psi(s, \lambda) = \Psi_1(s + g(\lambda, \mathcal{A})). \quad (37)$$

[The right hand side is the function  $\Psi_1$  evaluated at the point  $s + g(\lambda, \mathcal{A})$ .] The function  $g$  is symmetric with respect to  $\lambda = -\mathcal{A}/2$ . This symmetry gets inherited by the SCGF  $\Psi(s, \lambda)$ , where it reflects the fluctuation relation (27).

The expression (37) is simple in that the function  $g$  depends on system parameters only through the affinity  $\mathcal{A}$ , while the effects of all other properties of the system are encoded in a single function  $\Psi_1$ . Similarly in (9), the conditional distribution of  $J$  (given  $K$ ) is binomial and depends only on  $\mathcal{A}$ , but the distribution  $\tilde{P}_\tau(K)$  depends in a non-trivial way on all system parameters. In this sense, (37) is the consequence for large deviations of the detailed result (9) for finite times.

Fig. 4 illustrates (37) in the simple example of Fig. 1, for the cycle  $\mathcal{C} = ABCA$ . The numerical computation of the SCGFs was performed using the method described in Appendix C 2. Contours of the SCGF are the lines  $s = g(\lambda, \mathcal{A})$ : we show results for two sets of system parameters, which lead to the same value of  $\mathcal{A}$ ; hence the contour lines are the same in both cases, although the corresponding values of  $\Psi$  differ by an order of magnitude. Hence, the fact that these figures appear similar (despite the different model parameters) shows that the theoretical result (37) does indeed apply. This is a direct consequence (at the level of large deviations) of the binomial distribution of (9), which is the key result from which the other fluctuation properties are derived, in this work.

## B. Large deviations for words and cycles

This section outlines a general method for analysis of large deviations of cycle counts. This establishes that (28) does indeed hold, and provides a method for computation of SCGFs. Similar methods are used for analysis of word-counting in DNA sequence analysis [26] and in the statistics of repeated measurements [44], see also [19].

Similar SCGFs to (30) appear when considering large deviations of the number of transitions between discrete states of Markov models – for example one might define  $\hat{n}_2$  as the number of transitions  $A \rightarrow B$  and  $\hat{n}_2^R$  as the number of transitions  $B \rightarrow A$ . Then consider (30) with

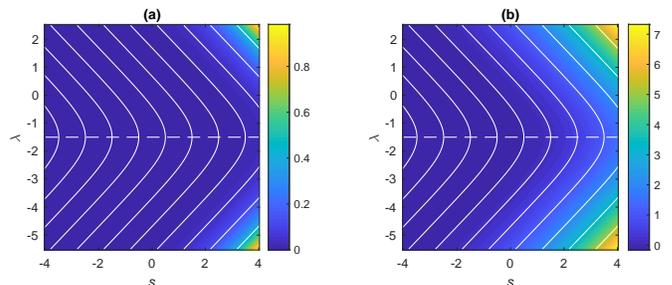


FIG. 4. SCGF  $\Psi(s, \lambda)$  (colour coded) for the current and traffic of the cycle  $\mathcal{C} = ABCA$  in Fig. 1. (a) and (b) differ in the choice of rates, but the affinity  $\mathcal{A}_c = 3$  is fixed. Lines of constant value of  $s + g$  are shown in solid white. They prescribe the overall shape of the SCGF, including the symmetry with respect to  $\lambda = -\mathcal{A}_c/2$  (dashed white line) corresponding to the fluctuation symmetry (27). Parameters: (a)  $w(A \rightarrow B) = w(B \rightarrow C) = w(C \rightarrow A) = 0.5$ ,  $w(C \rightarrow A) = w(D \rightarrow A) = 2$ ,  $w(B \rightarrow A) = w(A \rightarrow C) = 0.5e^{-2}$ ,  $w(C \rightarrow B) = 0.5e$ ,  $w(D \rightarrow C) = w(A \rightarrow D) = 2e^2$ ; (b)  $w(A \rightarrow B) = w(B \rightarrow C) = w(C \rightarrow A) = 1$ ,  $w(C \rightarrow D) = w(D \rightarrow A) = w(A \rightarrow D) = 0.1$ ,  $w(B \rightarrow A) = w(C \rightarrow B) = w(A \rightarrow C) = e^{-1}$ ,  $w(D \rightarrow C) = 0.1e^2$ .

$\hat{K}, \hat{J}$  replaced by  $\hat{K}_2 = \hat{n}_2 + \hat{n}_2^R$  and  $\hat{J}_2 = \hat{n}_2 - \hat{n}_2^R$  respectively. The resulting SCGF can be obtained by established methods [29–31, 42, 43] as the largest eigenvalue of a particular matrix that is called the *tilted generator*.

However, the established methodology is not applicable in the current setting because  $\hat{n}_c$  is not obtained by counting transitions between pairs of states (nor by considering state occupancies) – it requires that we count occurrences of specific *words*. The solution is to expand the state space of the original system to obtain an *extended system* in which each state is a word of length  $m$ . We illustrate this with the case  $m = 3$ . Suppose that the (original) system is in state C and the previous two states visited were A, B, in that order. Then the state of the extended system is the 3-letter word ABC. If the original process now makes a transition to A then the extended system makes a transition to BCA. (After the transition, the state is A and the previous two states were B, C.) This example is useful because this transition  $ABC \rightarrow BCA$  in the extended system corresponds exactly with a completed cycle in the original system. In other words, the problem of word-counting in the original model is reduced to a problem of counting transitions between states of the extended model. Since the extended model is still Markovian, established methods can then be used to compute the statistics of the relevant transitions, see below.

As a technical remark: this construction provides a mapping between trajectories of the original and extended systems, so that cycle counts of the original system can be inferred from the extended one. However, the initial  $m - 1$  states of a trajectory of the extended system are not fully-determined by a trajectory of the original system. This issue can cause some ambiguity in cycle

counts; but the problem can easily be rectified to obtain a one-to-one mapping of trajectories. Since the behavior of the first few states will not affect large-deviation analysis, we do not discuss this aspect.

To define more precisely the extended system, focus on a specific cycle and take  $m = m_{\mathcal{C}}$ . Each state of the extended system is an  $m$ -letter word (for example ABC or BCA), we denote these words by  $u, v, \dots$ . The transition rates of the extended system are denoted by  $W(u \rightarrow v)$ . The rate  $W(u \rightarrow v)$  is non-zero only if the first  $m - 1$  letters of word  $u$  are the same as the first  $m - 1$  letters of  $v$ . In this case  $W(u \rightarrow v) = w(u_f \rightarrow v_f)$  where  $u_f$  is the final letter of word  $u$ , and similarly  $v_f$  (recall that  $w$  indicates is a transition rate of the original system). One sees that construction of this extended system is a straightforward exercise, although it can be tedious because the number of states grows quickly with the word length and the number of states in the original system. For practical purposes, a milder extension of the state space is sufficient to establish specific results for cycle counts, see Appendix C 2.

Now write  $u_{\mathcal{C}}$  for the first  $m_{\mathcal{C}}$  letters of  $\mathcal{C}$  and  $v_{\mathcal{C}}$  for its last  $m_{\mathcal{C}}$  letters. (In the example  $\mathcal{C} = ABCA$  then  $u_{\mathcal{C}} = ABC$  and  $v_{\mathcal{C}} = BCA$ .) Then completion of cycle  $\mathcal{C}$  corresponds to a transition  $u_{\mathcal{C}} \rightarrow v_{\mathcal{C}}$  in the extended system, that is

$$\hat{n}_{\mathcal{C}}(X) = N_{u_{\mathcal{C}} \rightarrow v_{\mathcal{C}}}(X) \quad (38)$$

where  $N_{u \rightarrow v}(X)$  is the number of transitions  $u \rightarrow v$  in trajectory  $X$  of the extended system. In the same way,

$$\hat{n}_{\mathcal{C}}^{\mathbb{R}}(X) = N_{u_{\mathcal{C}}^{\mathbb{R}} \rightarrow v_{\mathcal{C}}^{\mathbb{R}}}(X) \quad (39)$$

where  $u_{\mathcal{C}}^{\mathbb{R}}$  indicates  $u_{\mathcal{C}^{\mathbb{R}}}$ , the first  $m_{\mathcal{C}}$  letters of  $\mathcal{C}^{\mathbb{R}}$ , and similarly  $v_{\mathcal{C}}^{\mathbb{R}}$ .

The extended system is itself Markovian, so standard methods can be used to analyse its large deviations. In particular, a method for counting transitions  $N(u \rightarrow v)$  between states is well-established [29–31, 42, 43], we give an outline, with details in Appendix C 1. The master equation of the extended system takes the standard form

$$\frac{\partial}{\partial t} P(u, t) = \sum_{v(\neq u)} [P(v, t)W(v \rightarrow u) - P(u, t)W(u \rightarrow v)]. \quad (40)$$

Now define a matrix  $\mathbb{W}$  with off-diagonal elements  $[\mathbb{W}^0]_{vu} = W(u \rightarrow v)$  and diagonal elements  $[\mathbb{W}^0]_{uu} = -\sum_v W(u \rightarrow v)$ . Then the master equation is  $\partial_t P = \mathbb{W}^0 P$ , where  $P$  is interpreted as a vector with elements  $P(u)$ . The SCGF can be obtained as the largest eigenvalue of the (“tilted”) matrix

$$\mathbb{W}(s, \lambda) = \mathbb{W}^0 + \mathbb{V}(s, \lambda), \quad (41)$$

where  $\mathbb{V}(s, \lambda)$  has only two elements that are non-zero:

$$\begin{aligned} [\mathbb{V}(s, \lambda)]_{v_{\mathcal{C}}, u_{\mathcal{C}}} &= (e^{s+\lambda} - 1)[\mathbb{W}^0]_{v_{\mathcal{C}}, u_{\mathcal{C}}}, \\ [\mathbb{V}(s, \lambda)]_{v_{\mathcal{C}}^{\mathbb{R}}, u_{\mathcal{C}}^{\mathbb{R}}} &= (e^{s-\lambda} - 1)[\mathbb{W}^0]_{v_{\mathcal{C}}^{\mathbb{R}}, u_{\mathcal{C}}^{\mathbb{R}}}. \end{aligned} \quad (42)$$

To establish that (28,30) hold, a few technical conditions are required on  $\mathbb{W}(s, \lambda)$ . Note that the extended process is Markov with a finite state space. In this case it is sufficient for it to have a unique steady state, which must hold if the original system is irreducible, as assumed above. Hence one has a large deviation result of the form (28).

Note that this construction is fully general, there was no assumption of non-revisiting cycles. If one does assume that  $\mathcal{C}$  is non-revisiting, the largest eigenvalue of  $\mathbb{W}(s, \lambda)$  must be of the form (37). An explicit derivation of this result is deferred to future work, which might also consider how large-deviation properties can be computed from the representation of the cycle-counting problem as a kind of renewal process via (B11), and what generalizations of the fluctuation theorems are possible for revisiting cycles.

## VI. CONCLUSION

We have analysed the joint distribution of cycle counts for forward and backward instances of a cycle  $\mathcal{C}$  in a discrete Markov process, as commonly used for analysis of non-equilibrium systems. The distribution is naturally characterised in terms of the cycle current  $J$  and the total count  $K$ . For non-revisiting cycles (which are those of primary physical relevance), the central result is (9), which shows that the conditional distribution of  $J$  given  $K$  is binomial and the only relevant parameter is the affinity. This shows that the conditional distribution of  $J$  is universal, with the affinity as its only parameter, while the distribution of  $K$  is free and depends on all system details.

For practical purposes, we point to (5) as a finite-time generalisation of (4), which might be useful as a way to infer affinities, as proposed in [16]. The counting of instances of cycle families rather than individual cycles, as discussed in Sec. IV A, might also help to improve this method.

We have also explained how large deviation theory can be applied to cycle counts. In particular, they do obey a large-deviation principle, whose properties can be computed from the extended system described here, by solving an eigenvalue problem.

These results suggest that further interesting structure may be present in distributions of cycle counts, either by analysis of the extended system, or by considering joint distributions of counts across more than one cycle. We look forward to future work in this direction.

## ACKNOWLEDGMENTS

We thank Jeremy Gunawardena and John Biddle for helpful discussions. J.G. acknowledges funding from the Royal Society under grant No. RP17002. This work was

funded in part by the European Research Council under the EU's Horizon 2020 Program, Grant No. 740269.

### Appendix A: Derivation of (13,17)

For completeness, we derive (13,17), starting from (10). First note that for any trajectory  $X$  starting at time  $t$  and ending at time  $\tau$ , the analogue of the path weight (10) can be written as

$$P[X] = p(x_0, t) \left[ \prod_{i=0}^{M-1} w(x_i \rightarrow x_{i+1}) e^{-r(x_i)\Delta_i} \right] \times e^{-r(X_M)(\tau-t_M)} \quad (\text{A1})$$

where  $\Delta_i = t_{i+1} - t_i$  is the sojourn time in state  $x_i$  and  $p(x_0, t)$  is the probability distribution of the initial state (at time  $t$ ). The states  $(x_0, x_1, \dots, x_M)$  and times  $(t_1, \dots, t_M)$  are indexed from time  $t$ , note also  $t_0 = t$ . This distribution  $P[X]$  is normalised in the sense that

$$1 = \sum_{M=0}^{\infty} \sum_{x_0 \dots x_M} \int_0^{\tau} dt_1 \int_{t_1}^{\tau} dt_2 \dots \int_{t_{M-1}}^{\tau} dt_M P[X]. \quad (\text{A2})$$

Using this distribution, and given a cycle  $\mathcal{C}$ , we compute the probability  $P_{\mathcal{C}}(t, \tau)\delta t$  of the following event: the trajectory has  $t_{m_{\mathcal{C}}} < \tau$  (from which it follows that  $M \geq m_{\mathcal{C}}$ ); also  $(x_0, x_1, \dots, x_{m_{\mathcal{C}}}) = (\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_{m_{\mathcal{C}}}, \mathcal{C}_1)$ , and  $t_1 \in [t, t + \delta t]$ . We use (A1) and sum over those  $x_k$  with  $k > m_{\mathcal{C}}$ , and integrate all the  $\Delta_i$ , to obtain [at leading order in  $\delta t$ ]:

$$P_{\mathcal{C}}(t, \tau)\delta t = p(\mathcal{C}_1, t)w(\mathcal{C}_1 \rightarrow \mathcal{C}_2)\delta t \prod_{j=2}^{m_{\mathcal{C}}} w(\mathcal{C}_j \rightarrow \mathcal{C}_{j+1}) \times \int d\Delta_2 \dots d\Delta_{m_{\mathcal{C}}} \Theta \left( \tau - t - \sum_{j=2}^{m_{\mathcal{C}}} \Delta_j \right) e^{-\sum_{j=2}^{m_{\mathcal{C}}} r(\mathcal{C}_j)\Delta_j}. \quad (\text{A3})$$

Here  $\Theta$  is the Heaviside (step) function; the  $\Delta_j$  are integrated over  $[0, \infty)$ ; we used that the integral for  $\Delta_0$  runs over  $[t, t + \delta t]$ , which yields the factor  $\delta t$ . Eq. (A3) coincides with (13) if we identify

$$F_{\text{fin}}(\tau - t, \mathcal{C}) = \int d\Delta_2 \dots d\Delta_{m_{\mathcal{C}}} \Theta \left( \tau - t - \sum_{j=2}^{m_{\mathcal{C}}} \Delta_j \right) \times \prod_{j=2}^{m_{\mathcal{C}}} \left[ r(\mathcal{C}_j) e^{-r(\mathcal{C}_j)\Delta_j} \right]. \quad (\text{A4})$$

To interpret this result, we identify  $\Delta_{\text{tot}} = \sum_{j=2}^{m_{\mathcal{C}}} \Delta_j$  as the sum of  $m_{\mathcal{C}} - 1$  exponential random variables with means  $r(\mathcal{C}_2)^{-1}, \dots, r(\mathcal{C}_{m_{\mathcal{C}}})^{-1}$ . As advertised in the main text, the result (A4) is simply the probability that this  $\Delta_{\text{tot}}$  is less than  $\tau - t$ . For any given  $m_{\mathcal{C}}$ , the integrals can

be performed, but we retain here the integral form, which shows the structure of the result. In particular, it is clear from (A4) that (17) holds, because  $\mathcal{C}^{\text{R}}$  contains the same states as  $\mathcal{C}$  (only the order is reversed), and the factor  $F_{\text{fin}}(\tau - t, \mathcal{C})$  from (A4) is invariant under permutation of the states  $\mathcal{C}_2, \dots, \mathcal{C}_{m_{\mathcal{C}}}$  within the cycle  $\mathcal{C}$ .

### Appendix B: Connection to renewal theory

#### 1. Alternative derivation of (9) by renewal theory

The results (6-9) for non-revisiting cycles can also be proven using a methodology similar to renewal processes. We include this analysis for completeness, and because the results provide additional information on the statistics of cycle completions, that may be useful for future work.

Suppose that the cycle  $\mathcal{C}$  of interest starts in state A. Any trajectory  $X$  can be decomposed into several pieces as in Fig. 5: an initial transient before the first visit to A, the time periods spent in A, the complete cycles between visits to A, and a final period between the last visit to A and the end of the trajectory at time  $\tau$ . Moreover, for any cycle  $\mathcal{C}$ , one can classify the complete cycles as instances of either  $\mathcal{C}$ , or  $\mathcal{C}^{\text{R}}$ , or some other cycle.

Hence, any trajectory  $X$  can be associated to a reduced trajectory  $Y$ , which is characterised by the sequences of arrival and departure times to/from A and the sequence of cycle types, for example

$$\mathcal{S}_Y = (\mathcal{C}, \mathcal{O}, \mathcal{C}, \mathcal{C}^{\text{R}}, \mathcal{O}, \dots) \quad (\text{B1})$$

where  $\mathcal{O}$  denotes any cycle other than  $\mathcal{C}, \mathcal{C}^{\text{R}}$ . (Separate occurrences of  $\mathcal{O}$  may indicate different cycles.) It is assumed that the cycle begins and ends with generic words that are indicated by  $\mathcal{W}$  in Fig. 5, these are not included in  $\mathcal{S}_Y$ . If the trajectory starts or ends in A then one or both of the  $\mathcal{W}$ s will have zero length. Compared with (18), this  $\mathcal{S}_Y$  is different in that it includes a separate element for every departure from A, not only those departures that lead to cycles  $\mathcal{C}$  or  $\mathcal{C}^{\text{R}}$ . Similarly, we use  $t_i^{\text{dep}}$  and  $t_i^{\text{arr}}$  in this Section to indicate the times of (all) departures/arrivals from/to A.

The mapping from  $X$  to the reduced trajectory  $Y$  is many-to-one because the times for transitions inside the cycles are not preserved, and nor are the sequences of states inside the generic cycles/words  $\mathcal{O}, \mathcal{W}$ . In the following, we consider the probabilities of the reduced trajectories  $Y$ , which are obtained by integrating over all possible trajectories  $X$  that reduce to  $Y$ . One sees that  $\hat{n}_{\mathcal{C}}(X)$  is the number of occurrences of  $\mathcal{C}$  in  $\mathcal{S}_Y$  (and similarly for  $\mathcal{C}^{\text{R}}$ ), so the full statistics of  $\hat{n}_{\mathcal{C}}, \hat{n}_{\mathcal{C}}^{\text{R}}$  can be computed from the statistics of the reduced trajectories  $Y$ .

The probabilities of the reduced trajectories have several useful properties, which are summarized here, with extra detail in Appendix B2. First, the times between each arrival in A and the next departure are all independent, they are exponentially distributed with mean

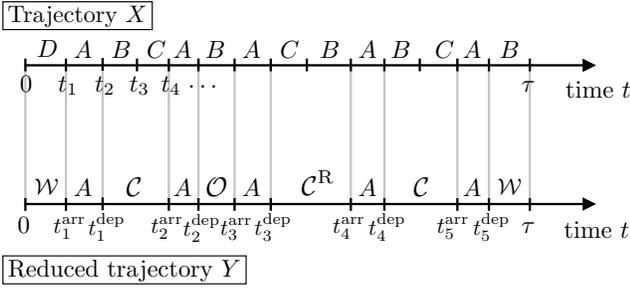


FIG. 5. A trajectory  $X$  of the four-state system in Fig. 1, and the corresponding reduced trajectory  $Y$ . The reduced trajectory keeps track of the arrival and departure times for state  $A$ , and on whether excursions from  $A$  are instances of  $\mathcal{C}$  or  $\mathcal{C}^R$  or some other cycle (indicated as  $\mathcal{O}$ ). The initial and final parts of the trajectory consist of generic words (indicated by  $\mathcal{W}$ ). This reduced trajectory has  $\mathcal{S}_Y = (\mathcal{C}, \mathcal{O}, \mathcal{C}^R, \mathcal{C})$

$r(A)^{-1}$ . Second, on departure from  $A$  at time  $t_{\text{dep}}$ , the subsequent behaviour is Markovian (independent of the previous history), as also occurs in renewal processes. The probability that any departure from  $A$  leads to a complete cycle  $\mathcal{C}$  is [similar to (13)]

$$P_{\mathcal{C}}^{\text{ren}}(\tau - t_{\text{dep}}) = \frac{w(A \rightarrow \mathcal{C}_2)}{r(A)} p_{\text{seq}}(\mathcal{C}) F_{\text{fin}}(\tau - t_{\text{dep}}, \mathcal{C}). \quad (\text{B2})$$

Also, given that such a cycle is completed, the time  $t_{\text{arr}}$  for the next arrival in  $A$  (which is the end time of the cycle) has cumulative distribution function

$$\mathcal{P}(t_{\text{arr}} < t_{\text{dep}} + \Delta t | \mathcal{C}) = \frac{F_{\text{fin}}(\Delta t, \mathcal{C})}{F_{\text{fin}}(\tau - t_{\text{dep}}, \mathcal{C})}. \quad (\text{B3})$$

Hence, given that the system departs from  $A$  at time  $t_{\text{dep}}$ , the probability density that it completes an instance of cycle  $\mathcal{C}$  and returns to  $A$  a time  $\Delta t$  later is

$$f_{\mathcal{C}}(\Delta t) = \frac{w(A \rightarrow \mathcal{C}_2)}{r(A)} p_{\text{seq}}(\mathcal{C}) \frac{\partial}{\partial(\Delta t)} F_{\text{fin}}(\Delta t, \mathcal{C}). \quad (\text{B4})$$

Using (17) and (3), the corresponding quantity for  $\mathcal{C}^R$  is

$$f_{\mathcal{C}^R}(\Delta t) = e^{-\mathcal{A}c} f_{\mathcal{C}}(\Delta t). \quad (\text{B5})$$

Following similar arguments, a formula is available for the probability of any reduced trajectory. This is given in Appendix B 2.

Notwithstanding that derivation, an important fact is already apparent from (B5): Given any reduced trajectory  $Y$ , one may obtain a new trajectory  $Y'$  by replacing any instance of  $\mathcal{C}$  in  $\mathcal{S}_Y$  by  $\mathcal{C}^R$  (keeping all other aspects of the trajectory fixed). The resulting trajectory probabilities are related as

$$\mathcal{P}(Y') = \mathcal{P}(Y) e^{-\mathcal{A}c} \quad (\text{B6})$$

[See also Appendix B 2, and note that this is analogous to (23).]

Recalling from (8) that  $\hat{K}$  is the total number of instances of  $\mathcal{C}$  and  $\mathcal{C}^R$ , one may define a set  $\Lambda$  containing  $2^{\hat{K}}$  (reduced) trajectories, formed by all possible replacements of  $\mathcal{C}$  by  $\mathcal{C}^R$ , and vice versa. Then the conditional probability of trajectory  $Y$  within this set is

$$\mathcal{P}(Y|\Lambda) = \frac{e^{-\mathcal{A}c \hat{n}_{\mathcal{C}}^R(Y)}}{(1 + e^{-\mathcal{A}c})^{K_{\Lambda}}} \quad (\text{B7})$$

where  $K_{\Lambda}$  is the value of  $\hat{K}$  for all trajectories in  $\Lambda$ . These trajectories have different values of  $\hat{J}$ ; the number of trajectories with any given value is a binomial coefficient. Hence [using (8)] the conditional distribution of  $\hat{J}$  is

$$P(J|\Lambda) = \binom{K_{\Lambda}}{\frac{1}{2}(K_{\Lambda} + J)} \frac{\exp(J\mathcal{A}/2)}{[2 \cosh(\mathcal{A}/2)]^{K_{\Lambda}}}. \quad (\text{B8})$$

Finally, the distribution  $\tilde{P}_{\tau}(K, J)$  of (9) can be obtained by conditional probability as  $\tilde{P}_{\tau}(K, J) = \sum_{\Lambda|K} P(J|\Lambda) P(\Lambda)$  where the sum (which might alternatively be expressed as an integral) is over all sets  $\Lambda$  with  $K_{\Lambda} = K$ , and  $P(\Lambda)$  is the probability that a random trajectory is in the set  $\Lambda$ . This yields (9).

## 2. Probabilities of reduced trajectories

We derive the probability of a reduced trajectory  $Y$ , whose definition is illustrated in Fig. 5. On each visit to state  $A$ , the system loses all memory of its previous history: this is a renewal. It follows that the probability of trajectory  $Y$  is given by a product of terms, one from each of its components. Denote the number of visits to  $A$  by  $\mathcal{N}$ , this is a random quantity but we do not write any hats, to lighten the notation. Hence  $Y$  is specified by  $\mathcal{N}$  arrival times and  $\mathcal{N}$  departure times, and the  $\mathcal{N} - 1$  elements of  $\mathcal{S}_Y$ . If the trajectory starts in  $A$  then we take  $t_1^{\text{arr}} = 0$  and if it ends in  $A$  then  $t_{\mathcal{N}}^{\text{dep}} = \tau$ . The initial condition of the system is given by a distribution  $p_{\text{ini}}$  over its states (it is not assumed that  $p_{\text{ini}}$  corresponds to the steady state).

The first contribution to the trajectory probability comes from the transient period before the first visit to  $A$ , it is a probability density for  $t_1^{\text{arr}}$ , which we write as  $f_{\text{beg}}(t_1^{\text{arr}})$ . This probability has two contributions, the first is  $p_{\text{ini}}(A)\delta(t_1^{\text{arr}})$  because the system may start in  $A$ . The second is the probability density that the system first reaches  $A$  at time  $t_1^{\text{arr}}$ . This distribution can be computed if necessary, for the purposes of this work it is sufficient that  $f_{\text{beg}}(t_1^{\text{arr}})$  exists, but the specific form is not required.

The next contribution comes from the visits to  $A$ . After each arrival, the system stays in  $A$  for a time  $t_{\text{dep}} - t_{\text{arr}}$  whose probability density is

$$f_A(t_{\text{dep}} - t_{\text{arr}}) = r(A) e^{-(t_{\text{dep}} - t_{\text{arr}})r(A)}. \quad (\text{B9})$$

The next contribution comes from completed cycles. On leaving  $A$  at time  $t_{\text{dep}}$ , the probability to complete a cycle  $\mathcal{C}$  and return a time  $\Delta t$  later is  $f_{\mathcal{C}}(\Delta t|t_{\text{dep}})$  as given in

(B4). A similar expression holds for cycle  $\mathcal{C}^R$ , see (B5). One must also consider the probability density to return to A by a different cycle (neither  $\mathcal{C}$  or  $\mathcal{C}^R$ ) after time  $\Delta t$ , which is denoted by  $f_{\mathcal{O}}(\Delta t|t_{\text{dep}})$ . The precise form of this function is not needed for the current purpose, only that it is well-defined (similar to  $f_{\text{beg}}$ ). Still, if one considers very long trajectories, a system that departs from A must eventually return to it, from which one deduces the normalization constraint

$$\int_0^\infty dt [f_{\mathcal{C}}(t) + f_{\mathcal{C}^R}(t) + f_{\mathcal{O}}(t)] = 1. \quad (\text{B10})$$

Finally one must consider the contribution to the trajectory probability from the final component, between the last departure from A and time  $\tau$ . This is denoted by  $p_{\text{end}}(\tau - t_{\mathcal{N}}^{\text{dep}})$ . The form of this contribution depends on whether the system ends the trajectory in state A (so  $\tau = t_{\mathcal{N}}^{\text{dep}}$ ) or not. In the latter case,  $p_{\text{end}}(t)$  is the probability that a system departing from state A does not return to it within time  $t$ . In the case  $\tau = t_{\mathcal{N}}^{\text{dep}}$  then  $p_{\text{end}}$  has a contribution  $r(\text{A})^{-1}\delta(t)$ , this factor combines with the  $f_{\text{A}}$  contributions to the trajectory probability to ensure that the distribution of times spent in A is correctly accounted for.

Combining all these ingredients, the probability density for the reduced trajectory  $Y$  is

$$\mathcal{P}(Y) = f_{\text{beg}}(t_1^{\text{arr}}) \left[ \prod_{i=1}^{\mathcal{N}} f_{\text{A}}(t_i^{\text{dep}} - t_i^{\text{arr}}) \right] f_{\text{end}}(\tau - t_{\mathcal{N}}^{\text{dep}}) \times \prod_{i=1}^{\mathcal{N}-1} f_{(\mathcal{S}_Y)_i}(t_{i+1}^{\text{arr}} - t_i^{\text{dep}}) \quad (\text{B11})$$

Here  $f_{(\mathcal{S}_Y)_i}$  is one of  $f_{\mathcal{C}}, f_{\mathcal{C}^R}, f_{\mathcal{O}}$ , according to which kind of cycle appears in the  $i$ th element of  $\mathcal{S}_Y$ .

From this final result one may directly check (B6), because the only change on replacing an instance of  $\mathcal{C}$  in  $Y$  by  $\mathcal{C}^R$  is to exchange a factor of  $f_{\mathcal{C}}(t_{i+1}^{\text{arr}} - t_i^{\text{dep}})$  for  $f_{\mathcal{C}^R}(t_{i+1}^{\text{arr}} - t_i^{\text{dep}})$ . Using this with (B5) yields (B6).

## Appendix C: Large deviation computation

### 1. SCGF for large deviations for cycle counts

We outline the derivation of the SCGF from (30) as the largest eigenvalue of the matrix  $\mathbb{W}(s, \lambda)$  in (41). We also explain why this SCGF cannot be derived by applying a ‘‘standard’’ tilting method to the original system.

Following (for example) [42, 45], we generalise the probability  $P(u, t)$  from (40) by defining  $P(u, n_{\mathcal{C}}, n_{\mathcal{C}^R}, t)$  as the probability for the extended system to be in state  $u$  at time  $t$ , having made  $n_{\mathcal{C}}$  completions of cycle  $\mathcal{C}$  and  $n_{\mathcal{C}^R}$  completions of  $\mathcal{C}^R$ . It is crucial that this  $P$  obeys its

own master equation:

$$\begin{aligned} \frac{\partial}{\partial t} P(u, n_{\mathcal{C}}, n_{\mathcal{C}^R}, t) &= \sum_{v(\neq u)} P(v, n_{\mathcal{C}}, n_{\mathcal{C}^R}, t) W(v \rightarrow u) \\ &\quad - \sum_{v(\neq u)} P(u, n_{\mathcal{C}}, n_{\mathcal{C}^R}, t) W(u \rightarrow v) \\ &+ [P(u_{\mathcal{C}}, n_{\mathcal{C}} - 1, n_{\mathcal{C}^R}, t) - P(u_{\mathcal{C}}, n_{\mathcal{C}}, n_{\mathcal{C}^R}, t)] W(u_{\mathcal{C}} \rightarrow v_{\mathcal{C}}) \delta_{u, v_{\mathcal{C}}} \\ &+ [P(u_{\mathcal{C}}, n_{\mathcal{C}}, n_{\mathcal{C}^R} - 1, t) - P(u_{\mathcal{C}}, n_{\mathcal{C}}, n_{\mathcal{C}^R}, t)] W(u_{\mathcal{C}}^R \rightarrow v_{\mathcal{C}}^R) \delta_{u, v_{\mathcal{C}}^R}. \end{aligned} \quad (\text{C1})$$

where the 3rd and 4th lines account for the fact that transitions  $u_{\mathcal{C}} \rightarrow v_{\mathcal{C}}$  and  $u_{\mathcal{C}}^R \rightarrow v_{\mathcal{C}}^R$  correspond to cycle completion events, in which the value of either  $n_{\mathcal{C}}$  or  $n_{\mathcal{C}^R}$  changes. [Recall Eq. (38).] Now define

$$\tilde{P}(u, s, \lambda, t) = \sum_{n_{\mathcal{C}}, n_{\mathcal{C}^R}} P(u, n_{\mathcal{C}}, n_{\mathcal{C}^R}, t) e^{(s+\lambda)n_{\mathcal{C}} + (s-\lambda)n_{\mathcal{C}^R}} \quad (\text{C2})$$

(the sums run from 0 to  $\infty$ ). Note that  $P$  is a normalised probability distribution over  $(u, n_{\mathcal{C}}, n_{\mathcal{C}^R})$  but  $\tilde{P}$  is not normalised. Then by (C1) one has

$$\frac{\partial}{\partial t} \tilde{P}(u, s, \lambda, t) = \sum_v [\mathbb{W}(s, \lambda)]_{u,v} \tilde{P}(v, s, \lambda, t) \quad (\text{C3})$$

where the matrix  $\mathbb{W}(s, \lambda)$  is defined in (41). This equation corresponds to  $\partial_t \tilde{P} = \mathbb{W}_{u,v}(s, \lambda) \tilde{P}$  from which one sees that the long-time behaviour of  $\tilde{P}$  is dominated by the largest eigenvalue of  $\mathbb{W}(s, \lambda)$ , that is  $\tilde{P}(u, s, \lambda, t) \simeq \tilde{P}_{\infty}(u, s, \lambda) e^{t\Psi(s, \lambda)}$  for large times, where  $\Psi(s, \lambda)$  is the largest eigenvalue. Summing over  $u$ , one may establish that the SCGF (30) coincides with this  $\Psi$ , as in [42, 45].

To see that this method requires the extended system, note that (C1) describes a Markovian dynamics for the evolution of  $(u, n_{\mathcal{C}}, n_{\mathcal{C}^R})$ , where  $u$  is the state of the extended system. By contrast, if one considers the original system (whose state is  $x$ ), the evolution of  $(x, n_{\mathcal{C}}, n_{\mathcal{C}^R})$  is not Markovian: the probability of an event where  $n_{\mathcal{C}}$  increases depends on the history of recently-visited states, and not only on the current state  $x$ . (Specifically,  $n_{\mathcal{C}}$  can only increase if the current state is  $\mathcal{C}_{m_{\mathcal{C}}}$  and the previous  $m_{\mathcal{C}} - 1$  states were  $\mathcal{C}_1, \dots, \mathcal{C}_{m_{\mathcal{C}}-1}$ .) As a result, the recipe given here – which connects SCGFs to eigenvalues – is only applicable at the level of the extended system.

In fact, the evolution of the state  $(x, n_{\mathcal{C}}, n_{\mathcal{C}^R})$  is an example of an  $m$ th order Markov process (we refer to [26] for applications of such models to word counting in discrete time, and to [46] for a discussion of large deviations in  $m$ th order Markov processes).

### 2. Practical calculation of large deviations for cycle counts

As discussed in Sec. V B, the SCGF  $\Psi$  for cycle counts can be characterised as the the largest eigenvalue of a

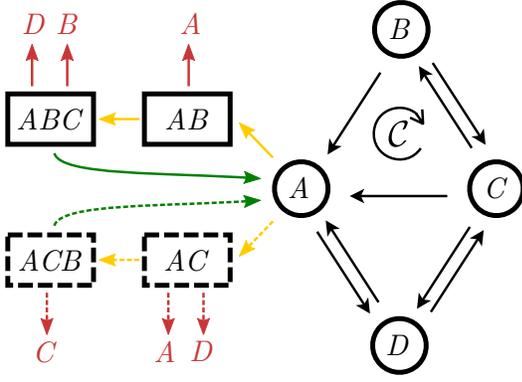


FIG. 6. Transitions in the extended state space for counting completions of the cycles  $\mathcal{C}$  and  $\mathcal{C}^R$  for the network of Fig. 1. Transitions within the original state space  $\Gamma$  (circled states) are shown in black. Additional states (boxed) are the partial completions of cycles  $\mathcal{C}$  (solid) and  $\mathcal{C}^R$  (dashed). Yellow arrows indicate attempted completions of cycles. These attempts may fail (via the transitions shown in red) or lead to a successful completion of a cycle (via the transitions shown in green).

matrix, which is a tilted generator for a Markov process on an extended state space. The size of this state space grows quickly with the model complexity, which makes explicit computations tedious. We explain here that a milder extension to the state space is already sufficient to obtain the SCGF (at least for non-revisiting cycles).

This (extended) state space contains all the elements of the original space  $\Gamma$ , along with states corresponding to progressive partial completions  $\tilde{\mathcal{C}}_2, \dots, \tilde{\mathcal{C}}_m$  of any cycle of interest (with length  $m$ ). As an example, we consider the cycles  $\mathcal{C} = ABCA$  and  $\mathcal{C}^R = ACBA$ , in which case the state space is extended by the states  $\tilde{\mathcal{C}}_2 = AB$ ,  $\tilde{\mathcal{C}}_3 = ABC$ , and  $\tilde{\mathcal{C}}_2^R = AC$ ,  $\tilde{\mathcal{C}}_3^R = ACB$ . A network representation of a Markov process on this extended space is shown in Fig. 6. Note that this extended state space grows only linearly with the length of the cycle of interest, as opposed to the exponential growth of the corresponding  $m$ -word space.

We order the three sub-spaces of the extended state space as  $(\Gamma, \tilde{\mathcal{C}}_2, \dots, \tilde{\mathcal{C}}_m, \tilde{\mathcal{C}}_2^R, \dots, \tilde{\mathcal{C}}_m^R)$  and accordingly construct a rate matrix of the block form

$$\mathbb{W}^0 = \begin{pmatrix} \mathbb{W}_\Gamma & \mathbb{W}_{\Gamma\mathcal{C}} & \mathbb{W}_{\Gamma\mathcal{C}^R} \\ \mathbb{W}_{\mathcal{C}\Gamma} & \mathbb{W}_{\mathcal{C}} & \mathbf{0} \\ \mathbb{W}_{\mathcal{C}^R\Gamma} & \mathbf{0} & \mathbb{W}_{\mathcal{C}^R} \end{pmatrix}. \quad (\text{C4})$$

We use the notation  $[\mathbb{W}^0]_{\nu\mu} \equiv W(\mu \rightarrow \nu)$  to label off-diagonal elements with column  $\mu$  and row  $\nu$  of the full matrix  $\mathbb{W}^0$  in the extended state space.

The block  $\mathbb{W}_\Gamma$  describes transitions within  $\Gamma$ , that do not mark the start of an attempted cycle, i.e.,  $W(x \rightarrow y) = w(x \rightarrow y)$  for all  $x, y \in \Gamma$ , except for  $W(\mathcal{C}_1 \rightarrow \mathcal{C}_2) = 0$  and  $W(\mathcal{C}_1^R \rightarrow \mathcal{C}_2^R) = 0$ . These transitions are marked in black in Fig. 6. [It is understood that  $w(\mathcal{C} \rightarrow \mathcal{C}) = 0$

in the formulae of this Section, because transitions only take place between distinct states.]

The blocks  $\mathbb{W}_{\mathcal{C}\Gamma}$  and  $\mathbb{W}_{\mathcal{C}^R\Gamma}$  have one non-zero entry each, marking the start of an attempted cycle. They have the rates  $W(\tilde{\mathcal{C}}_1 \rightarrow \tilde{\mathcal{C}}_2) = w(\mathcal{C}_1 \rightarrow \mathcal{C}_2)$  and  $W(\tilde{\mathcal{C}}_1^R \rightarrow \tilde{\mathcal{C}}_2^R) = w(\mathcal{C}_1^R \rightarrow \mathcal{C}_2^R)$ . Fig. 6 shows the relevant transitions for example system as yellow arrows leaving  $A$ .

The blocks  $\mathbb{W}_{\mathcal{C}}$  and  $\mathbb{W}_{\mathcal{C}^R}$  convey the successful continuation of the attempted cycle. Their non-zero rates are  $W(\tilde{\mathcal{C}}_i \rightarrow \tilde{\mathcal{C}}_{i+1}) = w(\mathcal{C}_i \rightarrow \mathcal{C}_{i+1})$  and  $W(\tilde{\mathcal{C}}_i^R \rightarrow \tilde{\mathcal{C}}_{i+1}^R) = w(\mathcal{C}_i^R \rightarrow \mathcal{C}_{i+1}^R)$  for  $2 \leq i < m$ , corresponding to the other yellow arrows in Fig. 6.

The blocks  $\mathbb{W}_{\Gamma\mathcal{C}}$  and  $\mathbb{W}_{\Gamma\mathcal{C}^R}$  convey transitions that mark the end of an (attempted) cycle. These are mostly unsuccessful terminations of an attempted cycle (shown in red in Fig. 6), except for a single transition for each cycle that closes it correctly (shown in green). The transition rates are  $W(\tilde{\mathcal{C}}_i \rightarrow y) = w(\mathcal{C}_i \rightarrow y)$  for  $2 \leq i \leq m$  and  $y \in \Gamma$ , except for  $W(\tilde{\mathcal{C}}_i \rightarrow \mathcal{C}_{i+1}) = 0$  when  $i < m$ ; and likewise for  $\mathcal{C}^R$ .

Finally, the diagonal elements of the transition matrix are set to  $[\mathbb{W}^0]_{\nu\nu} = -\sum_{\mu \neq \nu} [\mathbb{W}^0]_{\mu\nu}$ . The SCGF is obtained as the largest eigenvalue of the tilted matrix  $\mathbb{W}(s, \lambda) = \mathbb{W}^0 + \mathbb{V}(s, \lambda)$ , analogously to Eq. (41), where we count successful transitions from  $\tilde{\mathcal{C}}_m$  to  $\mathcal{C}_{m+1}$  or from  $\tilde{\mathcal{C}}_m^R$  to  $\mathcal{C}_{m+1}^R$  by setting

$$\begin{aligned} [\mathbb{V}(s, \lambda)]_{\mathcal{C}_{m+1}, \tilde{\mathcal{C}}_m} &= (e^{s+\lambda} - 1) [\mathbb{W}^0]_{\mathcal{C}_{m+1}, \tilde{\mathcal{C}}_m}, \\ [\mathbb{V}(s, \lambda)]_{\mathcal{C}_{m+1}^R, \tilde{\mathcal{C}}_m^R} &= (e^{s-\lambda} - 1) [\mathbb{W}^0]_{\mathcal{C}_{m+1}^R, \tilde{\mathcal{C}}_m^R}, \end{aligned} \quad (\text{C5})$$

and all other entries of  $[\mathbb{V}(s, \lambda)]_{\nu\mu}$  to zero.

For the particular example of Fig. 6, writing  $w_{xy} \equiv w(x \rightarrow y)$  and  $r_x \equiv r(x)$ , we obtain the tilted matrix (omitting zero elements):

$$\mathbb{W}(s, \lambda) = \begin{bmatrix} -r_A & w_{BA} & w_{CA} & w_{DA} & w_{BA} & w_{CA}e^{s+\lambda} & w_{CA} & w_{BA}e^{s-\lambda} \\ & -r_B & w_{CB} & & & w_{CB} & & \\ & & w_{BC} & -r_C & w_{DC} & & & w_{BC} \\ w_{AD} & & w_{CD} & -r_D & & w_{CD} & w_{CD} & \\ \hline w_{AB} & & & & -r_B & & & \\ & & & & w_{BC} & -r_C & & \\ \hline w_{AC} & & & & & & -r_C & \\ & & & & & & w_{CB} & -r_B \end{bmatrix}. \quad (\text{C6})$$

If  $s, \lambda = 0$ , this is a stochastic matrix – its columns sum to zero. Rows and columns correspond to the extended state space

$$(A, B, C, D, AB, ABC, AC, ACB). \quad (\text{C7})$$

The SCGFs in Fig. 4 were obtained by finding the largest eigenvalue of this matrix (for various  $s, \lambda$ ).

- 
- [1] G. Gallavotti and E. G. D. Cohen, *J. Stat. Phys.* **80**, 931 (1995).
- [2] C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
- [3] G. E. Crooks, *Phys. Rev. E* **61**, 2361 (2000).
- [4] D. Andrieux and P. Gaspard, *J. Stat. Phys.* **127**, 107 (2007).
- [5] U. Seifert, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [6] A. C. Barato and U. Seifert, *Phys. Rev. Lett.* **114**, 158101 (2015).
- [7] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, *Phys. Rev. Lett.* **116**, 120601 (2016).
- [8] P. Pietzonka, A. C. Barato, and U. Seifert, *Phys. Rev. E* **93**, 052145 (2016).
- [9] J. M. Horowitz and T. R. Gingrich, *Nat. Phys.* **16**, 15 (2019).
- [10] A. Dechant and S.-i. Sasa, *Proc. Nat. Acad. Sci. USA* **117**, 6430 (2020).
- [11] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, *Science* **296**, 1832 (2002).
- [12] S. Ciliberto, *Phys. Rev. X* **7**, 021051 (2017).
- [13] A. C. Barato and U. Seifert, *J. Phys. Chem. B* **119**, 6555 (2015).
- [14] J. Li, J. M. Horowitz, T. R. Gingrich, and N. Fakhri, *Nat. Commun.* **10**, 1666 (2019).
- [15] S. K. Manikandan, D. Gupta, and S. Krishnamurthy, *Phys. Rev. Lett.* **124**, 120603 (2020).
- [16] J. W. Biddle and J. Gunawardena, *Phys. Rev. E* **101**, 062125 (2020).
- [17] D.-Q. Jiang, M. Qian, and M.-P. Qian, *Mathematical Theory of Nonequilibrium Steady States* (Springer, Berlin/Heidelberg, 2004).
- [18] S. L. Kalpazidou, *Cycle Representations of Markov Processes* (Springer, 1995).
- [19] C. Jia, D.-Q. Jiang, and M.-P. Qian, *Ann. Appl. Prob.* **26**, 2454 (2016).
- [20] M. Polettini, G. Falasco, and M. Esposito, arxiv:2106.00425.
- [21] J. Schnakenberg, *Rev. Mod. Phys.* **48**, 571 (1976).
- [22] B. Altaner, S. Grosskinsky, S. Herminghaus, L. Katthän, M. Timme, and J. Vollmer, *Phys. Rev. E* **85**, 041133 (2012).
- [23] L. Bertini, A. Faggionato, and D. Gabrielli, *Stochastic Process. Appl.* **125**, 2786 (2015).
- [24] S. Schbath, *ESAIM: Probability and Statistics* **1**, 1 (1997).
- [25] S. Robin and J. J. Daudin, *J. Appl. Prob.* **36**, 179 (1999).
- [26] M. Lothaire, “Statistics on words with applications to biological sequences,” in *Applied Combinatorics on Words*, Encyclopedia of Mathematics and its Applications (Cambridge University Press, Cambridge, 2005) pp. 268–352.
- [27] E. Roquain and S. Schbath, *Adv. Appl. Prob.* **39**, 128 (2007).
- [28] É. Roldán and P. Vivo, *Phys. Rev. E* **100**, 042108 (2019).
- [29] H. Touchette, *Phys. Rep.* **478**, 1 (2009).
- [30] R. Chétrite and H. Touchette, *Ann. Henri Poincaré* **16**, 2005 (2015).
- [31] R. L. Jack, *Eur. Phys. J. B* **93**, 74 (2020).
- [32] C. Maes, *Physics Reports* **850**, 1 (2020).
- [33] H. Ge, *J. Phys. Chem. B* **112**, 61 (2008).
- [34] H. Ge, *J. Phys. A* **45**, 215002 (2012).
- [35] C. Maes and K. Netočný, *EPL* **82**, 30003 (2008).
- [36] J. Mehl, B. Lander, C. Bechinger, V. Blickle, and U. Seifert, *Phys. Rev. Lett.* **108**, 220601 (2012).
- [37] M. Polettini and M. Esposito, *Phys. Rev. Lett.* **119**, 240601 (2017).
- [38] M. Uhl, P. Pietzonka, and U. Seifert, *J. Stat. Mech. Theory Exp.* , 023203 (2018).
- [39] M. Kahlen and J. Ehrich, *J. Stat. Mech. Theory Exp.* **2018**, 063204 (2018).
- [40] K. Hayashi, H. Ueno, R. Iino, and H. Noji, *Phys. Rev. Lett.* **104**, 218103 (2010).
- [41] A. Alemany, M. Ribezzi-Crivellari, and F. Ritort, *New J. Phys.* **17**, 075009 (2015).
- [42] V. Lecomte, C. Appert-Rolland, and F. van Wijland, *J. Stat. Phys.* **127**, 51 (2007).
- [43] L. Bertini, A. Faggionato, and D. Gabrielli, *Ann. Inst. Henri Poincaré Probab. Stat.* **51**, 867 (2015).
- [44] M. van Horsen and M. Guță, *J. Math. Phys.* **56**, 022109 (2015).
- [45] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, *J. Phys. A* **42**, 075007 (2009).
- [46] R. Chétrite, A. Faggionato, and D. Gabrielli, (unpublished).