

# Affinity and Fluctuations in a Mesoscopic Noria

M. Bauer · F. Cornu

Received: 15 September 2013 / Accepted: 4 March 2014 / Published online: 28 March 2014  
© Springer Science+Business Media New York 2014

**Abstract** We exhibit the invariance of cycle affinities in finite state Markov processes under various natural probabilistic constructions, for instance under conditioning and under a new combinatorial construction that we call “drag and drop”. We show that cycle affinities have a natural probabilistic meaning related to first passage non-equilibrium fluctuation relations that we establish.

**Keywords** Affinity · Fluctuation relations · First passage times

## 1 Introduction

Affinity, a term ultimately borrowed from alchemy where it described a “natural attraction” among certain elements, and later seen in early chemistry as the “force” that causes chemical reactions, has by now a precise thermodynamical definition, related to the rate of irreversible variation of entropy when a chemical reaction progresses [1]. By analogy, affinity can be defined in a much more general context for certain random processes describing the time evolution of a probability measure, which offers striking analogies with the time evolution of concentrations of reactants: the variation of the entropy of the probability measure as time goes by often splits in a natural way and allows to single out irreversible entropy variations, hence affinities [2].

The purpose of this study is twofold.

- The first is to study important invariances of affinities: affinities for single “reactions” are in general not invariant under natural probabilistic constructions, but affinities for cycles

---

M. Bauer (✉)

Institut de Physique Théorique de Saclay, CEA/DSM/IPhT, URA2306 du CNRS,  
CEA Saclay, 91191 Gif-sur-Yvette Cedex, France  
e-mail: michel.bauer@cea.fr

F. Cornu

Laboratoire de Physique Théorique, UMR 8627 du CNRS, Université Paris-Sud,  
Bât. 210, 91405 Orsay, France

of reactions are much more robust objects. We study this invariance for two important examples. The first, related to a combinatorial construction which we call “drag and drop” is tailored for this study and is explained in detail. The second, conditioning, relies on well-known universal principles. Both of them allow, starting from a finite state Markov process, to define natural processes satisfying certain constraints, while preserving cycle affinities.

- These observations allow us to obtain a probabilistic interpretation of cycle affinities by concentrating on a single cycle even if the initial pattern of reactions is more general. This interpretation is our second aim, and we show that it leads to relations similar (or better dual) to standard out-of-equilibrium fluctuation relations [3]: the efficiency of a cycle (i.e. the preferred direction in which it is traversed, and the speed at which this traversal is made) is quantified by studying the winding number, and the cycle affinity is shown to be the crucial observable that relates the probability distribution of the time it takes to observe a winding number or its opposite.

The origin of the term “dual” is that instead of looking at fluctuations of an observable at a given time, we look at the fluctuations of the time it takes to reach a certain value of an observable. The dual picture has some relevance: experimentally, looking at the fluctuations of the time it takes to observe something is in fact quite common. For instance certain experimental setups are better suited to measure the time it takes for a particle to reach a certain displacement than to measure the displacement at a fixed time.

This article is organized as follows:

- Section 2 contains mostly standard background material. It starts by motivating and introducing the basic notions related to affinities in the context of Markov processes: cumulative processes, exchange quantities and processes, exchange currents. On top of their physical importance, those objects also have deep roots in probability theory. Restricting our attention to micro-reversible Markov processes (for which an allowed transition from configuration  $\mathcal{C}$  to configuration  $\mathcal{C}'$  goes together with an allowed transition from configuration  $\mathcal{C}'$  to configuration  $\mathcal{C}$ ) we recall how affinity is related to a canonical (because it involves only that data of the Markov process itself) exchange process via entropy and entropy variations: we reproduce the standard computations that motivate this interpretation and add a seemingly unknown remark on the positivity of the affinity current itself (before average) in the stationary state [see (2.4)]. From the affinity, which is defined for each reaction, i.e. each unoriented edge of the graph associated to the Markov process, we review the construction of the cycle affinity and its invariances, and define a convenient technical tool which we call the affinity class.
- Section 3 shows that many natural probabilistic operations on Markov processes preserve cycle affinities and the affinity class. We introduce a combinatorial construction which we call “drag and drop”. It associates to every walk on a graph a walk on a given subgraph. Then we explain what kind of process is induced by “drag and drop” applied to samples of a Markov process. It turns out that the continuous time Markov property is lost, but a generalized renewal property survives, which is enough for all our purposes. We also review briefly the more familiar construction of conditioning. “Drag and drop” and conditioning amount, albeit in different ways, to discard certain reactions (i.e. transitions) occurring in the original Markov process. We then give a list of probabilistic constructions, including “drag and drop” and conditioning, that preserve the affinity class. We conclude with some remarks on the observability of conditioning and “drag and drop” in real systems.
- Section 4 concentrates on the important case when the transitions in the system define a single cycle. The relevance of this simple case is enhanced by our previous observations

of the good behavior of cycle affinities when edges are discarded either by conditioning or by “drag and drop”: what we say for a single cycle remains true for a cycle embedded in a more general pattern of reactions. In an out-of-equilibrium situation, a current, which deserves the name winding current, will flow through the cycle. This current is associated to an exchange process which is indeed the winding number. After an algebraic preliminary, we give a heuristic argument to show that if the affinity  $A$  is  $\geq 0$ , then the probability that the winding number will ever reach  $-1$  is  $e^{-A}$  while the probability to reach any positive winding number is 1. Then we refine the heuristic argument to prove this result, and a stronger one related to first passage times and which is a simple dual fluctuation relation involving the affinity again [see (4.16), which can be considered as the main equation in this work]. We conclude with a formula (4.18) for the mean of the first passage time at winding number 1, a quantitative measure of efficiency. As already said, instead of looking at the distribution of an observable at a given time, we look at the distribution of the time it takes to reach a certain value of an observable. The two approaches are related, and this allows us to check the formula (4.18) for the mean of the first passage time against the mean of the winding at large times. More generally, though we have not tried to prove it, there must exist relationships between the fluctuation relations derived here and those obtained in [3].

A few basic notions, included for completeness and to fix notations, are gathered in two appendices. *The reader is advised to look in the appendices whenever in trouble with a result used in the main text, or with terminology.*

- Appendix 1 summarizes the basic notions from graph theory that we use in the text.
- Appendix 2 (resp. 3) gather notions on (time-homogeneous) finite state Markov chains (resp. Markov processes). We stress that trajectories of Markov chains (resp. Markov processes) can be analyzed as jumps governed by a Markov chain separated by independent geometric (resp. exponential) waiting times. This viewpoint is most useful for our discussion and is often less familiar to physicists than the master equation approach. We briefly mention the graph theoretic interpretation of transitions.

There is an extensive literature dealing with the role of cycles appearing in samples of Markov chains or processes with a countable number of states, see e.g. [4, 5]. The fundamental role played by cycles, affinity and entropy production to understand recurrence properties and the structure of stationary measures has a number of applications to other fields including deterministic dynamical systems.

## 2 Affinities and Their Invariances

This section is mostly for motivations. Our aim is mainly to recall the fundamental role played by affinities in non-equilibrium systems.

Though the presentation itself is not totally standard, most results [in particular those concerned with decompositions of entropy variations, with the exception of (2.4)] are.

### 2.1 Cumulative Processes

Consider a Markov process  $(\mathbf{C}, \mathbb{W})$  (where  $\mathbf{C}$  is a finite set of configurations and  $\mathbb{W}$  is the matrix of transitions rates from one configuration to another, see Appendix 3 for further notations and definitions).

The trajectory  $\mathcal{C}_t$ ,  $t \in [0, +\infty[$ , a random function from  $[0, +\infty[$  to  $\mathbf{C}$  where  $\mathcal{C}_t$  denotes the configuration at time  $t$ , describes a particle jumping from one configuration to another at certain instants and an interesting class of observables just counts something at each jump. We can describe this by introducing an arbitrary matrix  $\mathbb{Q}$  on  $\mathbf{C}$  with vanishing diagonal elements. To  $\mathbb{Q}$ , we associate a process  $X^{\mathbb{Q}} = (X_t^{\mathbb{Q}})_{t \in [0, +\infty[}$  defined as follows:  $X_0^{\mathbb{Q}} = 0$ , and each time there is a jump from a configuration  $\mathcal{C}$  to a configuration  $\mathcal{C}'$ ,  $X^{\mathbb{Q}}$  jumps by  $(\mathcal{C}'|\mathbb{Q}|\mathcal{C})$ . Formally:

$$X_t^{\mathbb{Q}} \equiv \sum_{s \in ]0, t]} (\mathcal{C}_s | \mathbb{Q} | \mathcal{C}_{s-}),$$

so that indeed  $X^{\mathbb{Q}}$  changes only when the particle jumps, i.e. for those  $t$  such that  $\mathcal{C}_t \neq \mathcal{C}_{t-}$ . The sum is well-defined because on  $]0, t]$  there are only finitely many jumps with probability 1 (this is a theorem, but the intuitive reason is clear: on average, the time between two jumps is bounded below by the inverse of the absolute value of the smallest diagonal element of the Markov matrix). A generic name for this type of processes would be “cumulative processes”. They are a special case of the so-called “additive functionals” used in the mathematical literature. A natural candidate for  $\mathbb{Q}$  would be to take all the non diagonal matrix elements equal to 1. Then  $X^{\mathbb{Q}}$  would simply count the number of jumps in  $]0, t]$  – what some authors call activity – surely an important observable.

Note that the value of  $(\mathcal{C}'|\mathbb{Q}|\mathcal{C})$  is immaterial if  $(\mathcal{C}'|\mathbb{W}|\mathcal{C}) = 0$ . We could for instance take the convention that  $(\mathcal{C}'|\mathbb{Q}|\mathcal{C}) = 0$  whenever  $(\mathcal{C}'|\mathbb{W}|\mathcal{C}) = 0$ . With this convention,  $\mathbb{Q}$  would appear as a function on edges of the graph  $\mathbf{G}$  associated to  $(\mathbf{C}, \mathbb{W})$ . But later, we shall give arguments to concentrate on anti-symmetric  $\mathbb{Q}$ 's, and this might lead to conflicts with this convention. Under the hypothesis of micro-reversibility, which is our main interest, there is no such conflict.

Suppose the Markov process is known up to time  $t$ . What is  $X_{t+\Delta t}^{\mathbb{Q}} - X_t^{\mathbb{Q}}$  in average? By definition,  $X_{t+\Delta t}^{\mathbb{Q}} - X_t^{\mathbb{Q}} = \sum_{s \in ]t, t+\Delta t]} (\mathcal{C}_s | \mathbb{Q} | \mathcal{C}_{s-})$ . With probability  $1 - \Delta t \sum_{\mathcal{C}' \neq \mathcal{C}_t} (\mathcal{C}'|\mathbb{W}|\mathcal{C}_t) + o(\Delta t)$  there has been no jump in  $]t, t + \Delta t]$ , and with probability  $\Delta t (\mathcal{C}'|\mathbb{W}|\mathcal{C}_t) + o(\Delta t)$  there has been a jump to  $\mathcal{C}' \neq \mathcal{C}_t$ . The possibility of several jumps is negligible. So the average of  $X_{t+\Delta t}^{\mathbb{Q}} - X_t^{\mathbb{Q}}$  if the process is known up to time  $t$ , which is in the probabilistic language a conditional expectation<sup>1</sup> is

$$\Delta t \sum_{\mathcal{C}' \neq \mathcal{C}_t} (\mathcal{C}'|\mathbb{W}|\mathcal{C}_t) (\mathcal{C}'|\mathbb{Q}|\mathcal{C}_t) + o(\Delta t).$$

This leads to introduce

$$j_t^{\mathbb{Q}} \equiv \sum_{\mathcal{C}' \neq \mathcal{C}_t} (\mathcal{C}'|\mathbb{W}|\mathcal{C}_t) (\mathcal{C}'|\mathbb{Q}|\mathcal{C}_t) = (\mathcal{C}_t | \mathbb{Q}^\dagger \mathbb{W} | \mathcal{C}_t), \quad (2.1)$$

where  $\mathbb{Q}^\dagger$  denotes the transpose of  $\mathbb{Q}$ .

Note that  $j_t^{\mathbb{Q}} = j^{\mathbb{Q}}(\mathcal{C}_t)$  if  $j^{\mathbb{Q}}(\mathcal{C}) \equiv \sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) (\mathcal{C}'|\mathbb{Q}|\mathcal{C})$ . The function  $j^{\mathbb{Q}}(\mathcal{C})$  (or the process  $j_t^{\mathbb{Q}}$ ) is called the current associated to  $\mathbb{Q}$ . By construction,  $\int_0^t j_s^{\mathbb{Q}} ds$  is a continuous process but, though the process  $X_t^{\mathbb{Q}}$  has jumps, they have the same average:

$$\langle X_t^{\mathbb{Q}} \rangle = \left\langle \int_0^t j_s^{\mathbb{Q}} ds \right\rangle = \int_0^t \langle j_s^{\mathbb{Q}} \rangle ds, \quad (2.2)$$

<sup>1</sup> On conditional expectations, see the general references given at the beginning of Appendix 2.

where  $\langle \cdot \rangle$  denotes the expectation or average with respect to the probability law  $P$  of the Markov process with characteristics  $(\mathbf{C}, \mathbb{W}, \nu)$  where  $\nu$  is the initial probability distribution. The proof is simple. If  $\langle X_{t+\Delta t}^{\mathbb{Q}} - X_t^{\mathbb{Q}} | \text{knowing } \mathcal{C}_s, s \in [0, t] \rangle$  denotes the average of  $X_{t+\Delta t}^{\mathbb{Q}} - X_t^{\mathbb{Q}}$  when the process is known up to time  $t$ , we have seen that

$$\langle X_{t+\Delta t}^{\mathbb{Q}} - X_t^{\mathbb{Q}} | \text{knowing } \mathcal{C}_s, s \in [0, t] \rangle = j_t^{\mathbb{Q}} \Delta t + o(\Delta t).$$

On the left-hand side we have already averaged over the fluctuations of the trajectory in the time interval  $[t, t + \Delta t]$  so if we average now over the possible trajectories  $\mathcal{C}_s, s \in [0, t]$ , we simply get  $\langle X_{t+\Delta t}^{\mathbb{Q}} - X_t^{\mathbb{Q}} \rangle$ . The right-hand side is insensitive to fluctuations in the trajectory after  $t$  so if we average now over the possible trajectories  $\mathcal{C}_s, s \in [0, t]$  we get  $\langle j_t^{\mathbb{Q}} \rangle \Delta t + o(\Delta t)$ . So  $\langle X_{t+\Delta t}^{\mathbb{Q}} - X_t^{\mathbb{Q}} \rangle = \langle j_t^{\mathbb{Q}} \rangle \Delta t + o(\Delta t)$ . The equality (2.2) follows by taking Riemann sums and letting  $\Delta t \rightarrow 0$ .

Note that  $\langle j_s^{\mathbb{Q}} \rangle = \sum_{\mathcal{C}, \mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C})(\mathcal{C}'|\mathbb{Q}|\mathcal{C})P(\mathcal{C}; s)$ , where  $P(\mathcal{C}; s) \equiv P(\mathcal{C}_s = \mathcal{C})$  is the probability to be in configuration  $\mathcal{C}$  at time  $s$ , and that  $\int_0^t P(\mathcal{C}; s)ds = \langle \int_0^t \mathbf{1}_{\mathcal{C}_s=\mathcal{C}} ds \rangle$  is just the expectation of the time spent at  $\mathcal{C}$  in the interval  $[0, t]$ . Hence,  $\langle X_t^{\mathbb{Q}} \rangle$  can be written in terms of occupation times.

Though this does not play any role in the sequel, let us stress that the relation between a cumulative process  $X^{\mathbb{Q}}$  and the associated current  $j^{\mathbb{Q}}$  has a deep probabilistic meaning.<sup>2</sup> We hope to return to this in a forthcoming work [6].

## 2.2 Exchange Processes

In the sequel, we shall concentrate on *anti-symmetric*  $\mathbb{Q}$ 's (which is not the case when one counts jumps). The physical reason is the following: we view the jumps in the Markov process as triggered by interactions with some external reservoirs, leading to the exchange of some conserved quantities (energy, particles,...). The reservoir responsible for a jump from  $\mathcal{C}$  to  $\mathcal{C}'$  would give to the system a quantity  $(\mathcal{C}'|\mathbb{Q}|\mathcal{C})$  which should be given back during the reverse jump from  $\mathcal{C}'$  to  $\mathcal{C}$ . Note that this is really a constraint only if  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$  and  $(\mathcal{C}|\mathbb{W}|\mathcal{C}')$  are both non-vanishing. When  $\mathbb{Q}$  is anti-symmetric, we call  $X_t^{\mathbb{Q}}$  the exchange process, and  $j_t^{\mathbb{Q}}$  the exchange current, associated to  $\mathbb{Q}$ . Note that we can rewrite  $j_t^{\mathbb{Q}}$  as a matrix product  $j_t^{\mathbb{Q}} = -(C_t|\mathbb{Q}\mathbb{W}|C_t)$ .

Another reason for the physical relevance of anti-symmetric  $\mathbb{Q}$ 's is that detailed balance is easily formulated: suppose that  $\mu$  is a probability on  $\mathbf{C}$ . The average of  $j^{\mathbb{Q}}$  under  $\mu$ ,  $\langle j^{\mathbb{Q}} \rangle_\mu \equiv \sum_{\mathcal{C}} j^{\mathbb{Q}}(\mathcal{C})\mu(\mathcal{C})$  vanishes for every anti-symmetric  $\mathbb{Q}$  if and only if  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})\mu(\mathcal{C}) = (\mathcal{C}|\mathbb{W}|\mathcal{C}')\mu(\mathcal{C}')$  for every  $\mathcal{C}, \mathcal{C}' \in \mathbf{C}$ . As usual, this implies that  $\mu$  is a stationary measure for  $(\mathbf{C}, \mathbb{W})$ . Hence, if  $P(\cdot; t)$  has a limit when  $t \rightarrow +\infty$ , this limit satisfies detailed balance if and only if the average of every exchange current  $\langle j_t^{\mathbb{Q}} \rangle$  goes to 0 when  $t \rightarrow +\infty$ , and then so does  $\frac{1}{t} \langle X_t^{\mathbb{Q}} \rangle = \frac{1}{t} \int_0^t \langle j_s^{\mathbb{Q}} \rangle ds$ . Note that we use the name “detailed balance” in a

<sup>2</sup> In proper mathematical language,  $X^{\mathbb{Q}}$  is a special semi-martingale, and its unique decomposition  $X_t^{\mathbb{Q}} = (X_t^{\mathbb{Q}} - \int_0^t j_s^{\mathbb{Q}} ds) + \int_0^t j_s^{\mathbb{Q}} ds$  as the sum of a martingale and a predictable finite variation process contains more than the mere coincidence  $\langle X_t^{\mathbb{Q}} \rangle = \langle \int_0^t j_s^{\mathbb{Q}} ds \rangle$ . Though powers of  $X^{\mathbb{Q}}$  are still special semi-martingales, their decomposition as a martingale plus a predictable finite variation process involves other currents. The systematics of this decomposition, which will be given elsewhere, is physically relevant. Indeed, in a number of experiments, the quantity of interest is some  $X^{\mathbb{Q}}$ , but only  $j^{\mathbb{Q}}$  and possibly a few other currents are measurable. The information carried by these currents on fluctuations of  $X^{\mathbb{Q}}$  (as embodied in moments  $\langle (X_t^{\mathbb{Q}})^k \rangle$  for instance) is limited and some combinatorial effort is required to extract it.

slightly generalized sense, because we have not specified an energy function on  $\mathbf{C}$ , defined a corresponding Boltzmann weight and so on.

There is also a mathematical reason, related to (co)homology, to focus on anti-symmetric  $\mathbb{Q}$ 's. Assuming micro-reversibility, we view anti-symmetric  $\mathbb{Q}$ 's as functions on edges of the graph  $\mathbf{G}$ , changing sign when the edge orientation is reversed. By definition, these are just the *1-cocycles* on  $\mathbf{G}$ . Special 1-cocycles are *1-coboundaries*, when  $(\mathcal{C}'|\mathbb{Q}|\mathcal{C}) = \mathcal{O}(\mathcal{C}') - \mathcal{O}(\mathcal{C})$  for some observable  $\mathcal{O}$ , i.e. for some function on  $\mathbf{C}$ . Notice that in that case  $X_t^{\mathbb{Q}} = \mathcal{O}(\mathcal{C}_t) - \mathcal{O}(\mathcal{C}_0)$  depends only on the configurations at time  $t$  and time 0, but that nevertheless the compensator  $\int_0^t \langle J_s^{\mathbb{Q}} \rangle ds$  depends on the whole past. A sequence  $(\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_n)$  of vertices of  $\mathbf{G}$  is called a walk (of length  $n$ ) on  $\mathbf{G}$  if  $(\mathcal{C}_m, \mathcal{C}_{m+1})$  is an edge of  $\mathbf{C}$  for  $0 \leq m < n$ . By construction, a trajectory of the Markov process  $(\mathbf{C}, \mathbb{W})$  up to (and including) the  $n^{\text{th}}$  jump is a walk of length  $n$  on  $\mathbf{G}$ . A walk with  $n \geq 3$  and  $\mathcal{C}_n = \mathcal{C}_0$  is called a *1-cycle* of  $\mathbf{G}$ . One can “integrate” a 1-cocycle along a walk:  $\hat{\int}_{(\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_n)} \mathbb{Q} \equiv \sum_{m=0}^{n-1} (\mathcal{C}_{m+1}|\mathbb{Q}|\mathcal{C}_m)$  (the notation  $\hat{\int}$  is not standard, it is introduced here just to stress the formal analogy with more familiar line integrals). The process  $X_t^{\mathbb{Q}}$  is just the “integral” of  $\mathbb{Q}$  along the walk described by the trajectory up to time  $t$ .

The “integral” of a 1-coboundary along a 1-cycle is always 0, and one can prove that conversely, if the integral of a 1-cocycle along any 1-cycle of  $\mathbf{G}$  vanishes, then the 1-cocycle is a 1-coboundary. The “integral” of a 1-coboundary along a path depends only on the origin and end of the path. If the 1-cocycle  $\mathbb{Q}$  is in fact a hidden 1-coboundary, one can recover the corresponding observable by choosing one “base” configuration in each connected component of  $\mathbf{G}$ , and setting  $\mathcal{O}(\mathcal{C}) \equiv \hat{\int}_{\text{Any path from a base configuration to } \mathcal{C}} \mathbb{Q}$ . Of course  $\mathcal{O}$  is only defined up to arbitrary additive constants (one for each connected component).

One can view 1-cycles as analogs of thermodynamic cycles, and 1-coboundaries as variations of state functions, while 1-cocycles are in some sense analogs of heat exchanges.

### 2.3 Affinities

In this section, where we assume micro-reversibility (see Appendix 1 and 3 for reference) all along, we come to particular exchange processes: those which can be defined solely in terms of the basic data of the Markov process  $(\mathbf{C}, \mathbb{W})$ . Namely, we have at our disposal  $\mathbb{W}$ , and possibly a probability distribution  $\mu$  on  $\mathbf{C}$ , or the time evolved probability distribution  $P(\mathcal{C}; t)$ .

The first exchange that comes to mind is probably  $\mathbb{W} - \mathbb{W}^\dagger$ , but it does not play an important role in the sequel.

The next one in terms of complexity is perhaps  $\mathbb{S}$  defined by

$$(\mathcal{C}'|\mathbb{S}|\mathcal{C}) \equiv \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')} \text{ if } \{\mathcal{C}, \mathcal{C}'\} \text{ is an edge of } \mathbf{G} \text{ and 0 if it isn't.}$$

The quantity  $\mathbb{S}$  is appealing for a number of reasons.

- It is dimensionless.
- The corresponding process  $X_t^{\mathbb{S}}$  is exactly the Lebowitz-Spohn action functional [7] for the trajectory up to time  $t$ . From a purely Markov process viewpoint,  $e^{X_t^{\mathbb{S}}}$  is, up to boundary terms, nothing but the ratio of the weight of the trajectory  $\mathcal{C}_s$ ,  $s \in [0, t]$  and its time reversal  $\mathcal{C}_{t-s}$ ,  $s \in [0, t]$  (a Radon-Nykodim derivative to be precise).
- The statement “The 1-cocycle  $\mathbb{S}$  for  $(\mathbf{C}, \mathbb{W})$  is a 1-coboundary” (for which it is enough to check that the “integral”  $\hat{\int}_{\text{cycle}} \mathbb{S}$  vanishes for every cycle of  $\mathbf{G}$ ) is an elegant mathematical way to say “Detailed balance is satisfied for  $(\mathbf{C}, \mathbb{W})$ ”. Indeed, if  $\mathbb{S}$  is a 1-coboundary we can write  $(\mathcal{C}'|\mathbb{S}|\mathcal{C}) = \ln w(\mathcal{C}') - \ln w(\mathcal{C})$  for a strictly positive function  $w$  on  $\mathbf{C}$ . This

means that  $\frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')} = \frac{w(\mathcal{C}')}{w(\mathcal{C})}$  for each edge of  $\mathbf{G}$ . Multiplying  $w$  by a function which is constant on each connected component of  $\mathbf{G}$  leaves this relation invariant. In particular, one may assume that  $w$  is a probability (i.e.  $\sum_{\mathcal{C}} w(\mathcal{C}) = 1$ ), leading to detailed balance.

- Finally, we emphasize that anti-symmetry is related to an interpretation of exchanges with reservoirs, and simple physical models [2, 8, 9] support the interpretation that  $(\mathcal{C}'|\mathbb{S}|\mathcal{C})$  is just the variation of entropy of the reservoirs when the system transits from  $\mathcal{C}$  to  $\mathcal{C}'$ .

Our next aim is to recall a standard entropy argument, which is essentially already in [2], supporting this view.

### 2.3.1 Entropy Variations and Chemical Affinity

The first observation is that  $j^{\mathbb{S}}(\mathcal{C}) = \sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}'|\mathbb{W}|\mathcal{C})(\mathcal{C}'|\mathbb{S}|\mathcal{C})$ , so that  $\langle j_t^{\mathbb{S}} \rangle = \sum_{\mathcal{C}, \mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C})(\mathcal{C}'|\mathbb{S}|\mathcal{C}) P(\mathcal{C}; t)$  which by anti-symmetry can be rewritten as

$$\langle j_t^{\mathbb{S}} \rangle = \frac{1}{2} \sum_{\mathcal{C}, \mathcal{C}'} \dot{\xi}_{\mathcal{C}; \mathcal{C}'} \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')},$$

where we have set  $\dot{\xi}_{\mathcal{C}; \mathcal{C}'} \equiv -\dot{P}_{\mathcal{C}; \mathcal{C}'}(t) = -(\mathcal{C}|\mathbb{W}|\mathcal{C}') P(\mathcal{C}'; t) + (\mathcal{C}'|\mathbb{W}|\mathcal{C}) P(\mathcal{C}; t)$ . The rationale for this notation is that:

- The equation for the time evolution of  $P(\mathcal{C}; t)$  can be rewritten as

$$\frac{dP(\mathcal{C}; t)}{dt} = \sum_{\mathcal{C}'} \dot{P}_{\mathcal{C}; \mathcal{C}'}(t),$$

i.e.  $\dot{P}_{\mathcal{C}; \mathcal{C}'}(t)$  is the contribution of edge  $\{\mathcal{C}, \mathcal{C}'\}$  to the variation of  $P(\mathcal{C}; t)$  with time.

- Consequently, viewing  $\mathcal{C}, \mathcal{C}'$  as chemical species and the edge  $\{\mathcal{C}, \mathcal{C}'\}$  as a chemical reaction,

$$\int_0^t \dot{\xi}_{\mathcal{C}; \mathcal{C}'}(s) ds = - \int_0^t \dot{P}_{\mathcal{C}; \mathcal{C}'}(s) ds$$

is by definition the “extent of reaction” (modulo a substitution of probabilities for concentrations), quantifying how much the reaction  $\mathcal{C} \leftrightarrow \mathcal{C}'$  has progressed between time 0 and  $t$ . The minus sign is because the extent of the reaction grows when the reactant concentration drops down.

The next step is to introduce the (dimensionless) Shannon-Gibbs entropy at time  $t$ . Recall that if  $\mu$  is any probability distribution on  $\mathbf{C}$ , then  $S^{SG}[\mu] \equiv -\sum_{\mathcal{C}} \mu(\mathcal{C}) \ln \mu(\mathcal{C}) = -\langle \ln \mu \rangle_{\mu}$  where  $\langle \cdot \rangle_{\mu}$  is the expectation for a function on  $\mathbf{C}$  under the probability distribution  $\mu$ , not to be confused with  $\langle \cdot \rangle$  which denotes in this paper the expectation with respect to the Markov process probability measure on paths.<sup>3</sup> The entropy of the probability distribution  $P(\mathcal{C}; t)$  is thus

$$S^{SG}[P(t)] = - \sum_{\mathcal{C}} P(\mathcal{C}; t) \ln P(\mathcal{C}; t). \quad (2.3)$$

Note that Boltzmann’s constant is set equal to 1, hence the name “dimensionless”.

<sup>3</sup> There is a relationship though. If  $\mathcal{O}$  is an observable, i.e. a function on  $\mathbf{C}$ , one can associate to it a random variable depending on the path,  $\mathcal{O}(\mathcal{C}_t)$ . The path dependence is only via the configuration  $\mathcal{C}_t$  at time  $t$ , and in that case the expectation with respect to the probability measure on paths can be computed by taking an expectation with respect to its one-time marginal  $P(\mathcal{C}, t)$  which is a probability on  $\mathbf{C}$ :  $\langle \mathcal{O}(\mathcal{C}_t) \rangle = \langle \mathcal{O} \rangle_{P(t)}$ .

Simple manipulations show that

$$\frac{dS^{SG}}{dt} = \frac{1}{2} \sum_{\mathcal{C}, \mathcal{C}'} \dot{\xi}_{\mathcal{C}; \mathcal{C}'}(t) \ln \frac{P(\mathcal{C}; t)}{P(\mathcal{C}'; t)}$$

which can be split in two:

$$\frac{dS^{SG}}{dt} = \frac{1}{2} \sum_{\mathcal{C}, \mathcal{C}'} \dot{\xi}_{\mathcal{C}; \mathcal{C}'}(t) \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C}; t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}'; t)} - \langle j_t^S \rangle$$

Each summand in the first term is  $\geq 0$  (it is of the form  $-(x' - x) \log(x/x')$ ), so it contributes to a systematic increase of entropy, and can be interpreted as a sign of irreversibility of the chemical reaction  $\mathcal{C} \leftrightarrow \mathcal{C}'$ . Then each summand in the second term can be interpreted as the entropy change due to exchanges with reservoirs, thus recovering the usual splitting of entropy variations as the sum of two contributions: one which is always positive and signs irreversibility, and one leading to entropy variations in the reservoirs:  $dS^{SG} = (dS^{SG})_{\text{irr}} + (dS^{SG})_{\text{exch}}$ . Note that the interpretation of  $e^{X_t^S}$  given at the beginning of Sect. 2.3 as related to time reversal supports strongly that  $(dS^{SG})_{\text{exch}}$  is a reversible contribution. The standard name in the literature for the irreversible contribution  $(dS^{SG})_{\text{irr}}$  is simply “entropy production” and its crucial role to understand probabilistic properties (for instance recurrence) is widely recognized (see e.g. [2, 7, 4, 5, 10]).

The above discussion is only one of the lines of reasoning leading to this decomposition (see e.g. [10, 11]). In certain circumstances, when exchanges with reservoirs are explicitly built in the transition rates (see e.g. [8, 9]), there is a direct computation at the level of the reservoirs leading to the equality (not just an abstract interpretation)  $(dS^{SG})_{\text{exch}} = -dS^{\text{Res}}$ , where  $dS^{\text{Res}}$  is the entropy variation of the reservoirs due to exchanges with the system. Let us stress however that even if these arguments, analogies and identifications are convincing and well-motivated, their universal validity for out-of-equilibrium systems is not proven.

Returning to the main stream of the argument, we thus have

$$(dS^{SG})_{\text{irr}} = \frac{1}{2} \sum_{\mathcal{C}, \mathcal{C}'} \dot{\xi}_{\mathcal{C}; \mathcal{C}'}(t) \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C}; t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}'; t)} dt \quad (dS^{SG})_{\text{exch}} = -\langle j_t^S \rangle dt.$$

Finally, we use the chemical definition of affinity: if  $\xi$  is the “extent of a reaction” and  $S$  is the entropy, the affinity  $A$  is defined<sup>4</sup> via  $(dS)_{\text{irr}} = Ad\xi$ . For the reaction  $\mathcal{C} \leftrightarrow \mathcal{C}'$ , we have seen that  $d\xi = \dot{\xi}_{\mathcal{C}; \mathcal{C}'} dt$ . Putting all these considerations together, we obtain that the affinity of the reaction  $\mathcal{C} \leftrightarrow \mathcal{C}'$  is  $\ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C}; t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}'; t)}$  (there is no  $1/2$  because  $\sum_{\mathcal{C}, \mathcal{C}'}$  counts every unoriented edge, i.e. every reaction, twice).

If  $\mu$  is a probability distribution on  $\mathbf{C}$ , we define  $\mathbb{A}^\mu$  by

$$(\mathcal{C}'|\mathbb{A}^\mu|\mathcal{C}) = \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})\mu(\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')\mu(\mathcal{C}')},$$

so that  $(\mathcal{C}'|\mathbb{A}^\mu|\mathcal{C})$  is the affinity of the reaction  $\mathcal{C} \leftrightarrow \mathcal{C}'$  when the probability distribution of the system is  $\mu$ . For fixed  $\mu$ ,  $\mathbb{A}^\mu$  is an exchange, or 1-cocycle, and if the system is in state  $\mu$  we can write

$$(dS^{SG})_{\text{irr}} = \langle j^{\mathbb{A}^\mu} \rangle_\mu dt.$$

<sup>4</sup> The term “affinity” was coined by De Donder (see e.g. [1] for an introduction in English).

The fact, recalled above, that  $\langle j^{\mathbb{A}^\mu} \rangle_\mu$  is always  $\geq 0$  is well-documented in the literature. However, we have not found there that for  $\mu = P_{\text{st}}$  (the stationary measure) the current  $j^{\mathbb{A}^{P_{\text{st}}}}$  itself is positive:

$$j^{\mathbb{A}^{P_{\text{st}}}}(\mathcal{C}) = \sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C}) P_{\text{st}}(\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}') P_{\text{st}}(\mathcal{C}')} \geq 0 \quad \forall \mathcal{C}. \quad (2.4)$$

This fact can be proven by hand using  $-\ln x \geq 1 - x$  for  $x \geq 0$  and the definition of  $P_{\text{st}}$ :  $\sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}') P_{\text{st}}(\mathcal{C}') = \sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) P_{\text{st}}(\mathcal{C})$ . It is also a consequence of the interpretation of  $j^{\mathbb{A}^{P_{\text{st}}}}(\mathcal{C})$  for each fixed  $\mathcal{C}$ , rewritten as

$$- \sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) \ln \frac{P_{\text{st}}(\mathcal{C})^{-1} (\mathcal{C}|\mathbb{W}|\mathcal{C}') P_{\text{st}}(\mathcal{C}')}{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}$$

i.e. as a relative entropy.

One can carry a similar analysis for Markov chains: if  $(\mathbf{C}, \mathbb{P})$  is a Markov chain, and under the usual assumption of micro-reversibility, one can show that the (discrete time) variation of the entropy leads to consider

$$\ln \frac{(\mathcal{C}'|\mathbb{P}|\mathcal{C})}{(\mathcal{C}|\mathbb{P}|\mathcal{C}')} \text{ and } \ln \frac{(\mathcal{C}'|\mathbb{P}|\mathcal{C})\mu(\mathcal{C})}{(\mathcal{C}|\mathbb{P}|\mathcal{C}')\mu(\mathcal{C}')} \quad (2.5)$$

as the discrete time analogs of  $\mathbb{S}$  and  $\mathbb{A}^\mu$ .

Note that  $\mathbb{A}^\mu$  is a 1-cocycle and that  $\mathbb{A}^\mu - \mathbb{S}$  is a 1-coboundary, as already noticed by Schnackenberg in a different language [2]. So the “integrals” of  $\mathbb{A}^\mu$  and  $\mathbb{S}$  along any cycle of  $\mathbf{G}$  agree. As our interest in the sequel is mainly in such “integrals”, there is no reason to consider  $\mathbb{A}^\mu$  and  $\mathbb{S}$  as different objects. Another way to phrase this idea, maybe more familiar for physicists, is to view 1-coboundaries as gauge transformations in an abelian gauge theory, as emphasized in [12] as we learned after this work was completed.

In the (co)homology language, one regroups all 1-cocycles differing by a 1-coboundary under the name “cohomology class”. All we shall keep from the concept is the name. So we define the affinity class as the class of all exchanges that differ from  $\mathbb{S}$  by a 1-coboundary. As seen above, this class comprise  $\mathbb{A}^\mu$  whatever  $\mu$ .

The above symmetry is somewhat formal at the moment, and our aim in the next section is to show that the affinity class is preserved by many natural probabilistic constructions.

### 3 Probabilistic Constructions

#### 3.1 A “Graphical” Construction: Drag and Drop

At the basis of our interpretation of affinities lies a simple<sup>5</sup> general construction on (oriented or non-oriented) graphs. We refer the reader to Appendix 1 for basic definitions and notations for graphs. This construction has several variants. Suppose that  $\mathbf{G}$  is a graph and  $\mathbf{H}$  a subgraph of  $\mathbf{G}$ .

##### 3.1.1 First Variant

Take a walk  $X = (X_0, \dots, X_N)$  on  $\mathbf{G}$  and let  $Y_0$  be a vertex of  $\mathbf{H}$ . Construct recursively a sequence  $Y = (Y_0, \dots, Y_N)$  of vertices of  $\mathbf{H}$  as follows. The first term  $Y_0$  is already

<sup>5</sup> So simple that it is likely to have already been introduced elsewhere, but we are not aware of a reference.

defined. Suppose that, for some  $0 \leq n < N$ ,  $Y_0, \dots, Y_n$  are already defined. If  $X_n = Y_n$  and  $(X_n, X_{n+1})$  is an edge of  $\mathbf{H}$  then set  $Y_{n+1} = X_{n+1}$ , otherwise, set  $Y_{n+1} = Y_n$ . By construction,  $Y_0, \dots, Y_N$  is a sequence of vertices of  $\mathbf{H}$ , but in general is does not have to be a walk on  $\mathbf{H}$  because  $Y$  can stay at the same place even if there can be no edge having this vertex both as origin and extremity.

### 3.1.2 Second Variant

Take a walk  $X = (X_0, \dots, X_N)$  on  $\mathbf{G}$  and let  $Y_0$  be a vertex of  $\mathbf{H}$ . Construct recursively a walk  $Y = (Y_0, \dots, Y_M)$  on  $\mathbf{H}$  and a sequence  $n_0 = 0 < n_1 < \dots < n_M \leq N$  for some  $0 \leq M \leq N$  as follows. The first terms  $Y_0$  and  $n_0$  are already defined. Suppose that for some  $0 \leq m$ ,  $Y_0, \dots, Y_m$  and  $n_0 = 0 < n_1 < \dots < n_m$  are already defined. If it exists, let  $n$  be the smallest integer  $n_m \leq n < N$  such that  $X_n = Y_m$  and  $(X_n, X_{n+1})$  is an edge of  $\mathbf{H}$  and set  $Y_{m+1} = X_{n+1}$ ,  $n_{m+1} = n + 1$ . If no such  $n$  exists, set  $M = m$  and stop.

Of course, we could replace the finite walk  $X_0, \dots, X_N$  by an infinite walk  $X_0, X_1, \dots$  and do the same construction. The second variant is most useful for our analysis. By lack of a better name, we choose to call it “drag and drop” along  $\mathbf{H}$ .

The intuitive picture is that  $Y$  represents the position of an object (a box) attached to  $\mathbf{H}$  and the walker  $X$  carries  $Y$  along  $\mathbf{H}$  when they meet and  $X$  moves along  $\mathbf{H}$ . The second variant is closely related to the first one. In the first variant,  $Y$  and  $X$  are indexed by the same “clock”, in the second variant, the clock of  $Y$  ticks only when  $X$  carries  $Y$  along  $\mathbf{H}$ .

There is a special case of “drag and drop” which is already of some interest. It happens when all walks on  $\mathbf{G}$  joining two distinct vertices of  $\mathbf{H}$  involve at least one edge of  $\mathbf{H}$ .<sup>6</sup> In that case, the box is not really needed: to go from the walk  $X = (X_0, \dots, X_N)$  to the walk  $Y = (Y_0, \dots, Y_M)$  first remove all vertices up to the first appearance of  $Y_0$ , then all vertices after  $Y_0$  that do not belong to  $\mathbf{H}$ , and finally remove multiple contiguous occurrences of one and the same vertex. This simple special case is relevant when joining two distinct vertices of  $\mathbf{H}$  without going through an edge of  $\mathbf{H}$  requires rather long walks, which can be neglected to a good approximation in practice. This could be used for instance to get a crude model of the experiment in [13].

It is not unlikely that this kind of coupled motion is routinely performed by the biological molecular machinery, though the microscopic understanding is still too preliminary to know for sure [14].

In such a scenario,  $\mathbf{H}$  models for instance some hetero-polymer within a cell and  $\mathbf{G}$  models  $\mathbf{H}$  plus the relevant environment of  $\mathbf{H}$  inside the cell. There is a ligand bound at some place on  $\mathbf{H}$  but that may move along it. There is a complex that moves on  $\mathbf{G}$ . When it meets the ligand somewhere on  $\mathbf{H}$ , it binds to the ligand. If the complex moves along  $\mathbf{H}$  it carries the ligand, but if it goes somewhere else the ligand and complex unbind and the ligand waits for another encounter with the complex.

By the way, in this interpretation, there may be several complexes that cannot occupy the same position at the same time and it is trivial to generalize the above construction to this more general case. Then, the process becomes close to hitch-hiking. The hitch-hiker wants to go somewhere or make some journey, but several roads may be acceptable for him. When a car stops and the driver offers a lift to a place that is compatible with the hitch-hiker’s goals, they do a piece of journey together after which the driver drops the hitch-hiker and drives his way, while the hitch-hiker waits for another driver to continue his journey.

<sup>6</sup> For microreversible graphs, this can be stated as follows: if the edges of  $\mathbf{H}$  are removed from  $\mathbf{G}$ , distinct vertices of  $\mathbf{H}$  belong to disjoint connected component of the resulting graph.

This generalization is also appropriate to make a connection with some recent experiments in active matter with applications to nanotechnologies, see e.g. Ref.[13] entitled “Bacterial Ratchet Motors”, for which the special case of “drag and drop” is already of some relevance.

### 3.2 Drag and Drop for Markov Processes

We would like to understand “drag and drop” when the walker moves according to a finite state Markov process. We refer the reader to Appendix 3 for the relevant notations and basic constructions. To avoid some trivial situations, we assume that the number of configurations is  $\geq 2$ . We denote by  $\mathbb{W}$  the transition matrix and by  $\mathbb{M} = \mathbb{W} - \mathbb{D}$  the infinitesimal generator of a finite state Markov process, by  $\mathbb{P}$  the associated stochastic matrix satisfying  $\mathbb{M} = (\mathbb{P} - \mathbb{I})\mathbb{D}$  and by  $\mathbf{G}$  the associated graph. We let  $\mathbf{H}$  be a subgraph of  $\mathbf{G}$  and denote by  $\mathbb{W}_{\mathbf{H}}^{\text{rest}}$  the restriction of the transition matrix to  $\mathbf{H}$ , i.e.  $(\mathcal{C}'|\mathbb{W}_{\mathbf{H}}^{\text{rest}}|\mathcal{C})$  is defined only when  $\mathcal{C}, \mathcal{C}'$  are vertices of  $\mathbf{H}$  and then:  $(\mathcal{C}'|\mathbb{W}_{\mathbf{H}}^{\text{rest}}|\mathcal{C}) \equiv (\mathcal{C}'|\mathbb{W}|\mathcal{C})$  if  $(\mathcal{C}, \mathcal{C}')$  is an edge of  $\mathbf{H}$  and 0 otherwise. From  $\mathbb{W}_{\mathbf{H}}^{\text{rest}}$  viewed as a transition matrix (indexed by vertices of  $\mathbf{H}$ ) we construct as recalled in Appendix 3 an infinitesimal generator  $\mathbb{M}_{\mathbf{H}}^{\text{rest}} = \mathbb{W}_{\mathbf{H}}^{\text{rest}} - \mathbb{D}_{\mathbf{H}}^{\text{rest}}$  and a stochastic matrix  $\mathbb{P}_{\mathbf{H}}^{\text{dd}}$  (where the superscript dd stands for “drag and drop”) satisfying  $\mathbb{M}_{\mathbf{H}}^{\text{rest}} = (\mathbb{P}_{\mathbf{H}}^{\text{dd}} - \mathbb{I}_{\mathbf{H}})\mathbb{D}_{\mathbf{H}}^{\text{rest}}$ .

We can apply the general construction of “drag and drop”. Take a trajectory of the Markov process on  $\mathbf{G}$  with infinitesimal generator  $\mathbb{M}$  and initial probability distribution  $P(\mathcal{C}, 0)$ . If  $0, T, T + T', \dots$  denotes the (finite or infinite) sequence of jump times we let  $(\mathcal{C}, \mathcal{C}', \mathcal{C}'', \dots)$  denote the (finite or infinite) sequence of corresponding positions on the graph: the process is at  $\mathcal{C}$  between 0 and  $T$ , at  $\mathcal{C}'$  between  $T$  and  $T + T'$ , and so on. Keeping in mind our remarks (see Appendix 3) on the connection between a finite state Markov process and the associated finite state Markov chain, if the sequence  $(\mathcal{C}, \mathcal{C}', \mathcal{C}'', \dots)$  is finite, we turn it into an infinite one by repeating the last member again and again. Thus we have the first ingredient for “drag and drop”. The second ingredient is the initial configuration  $\bar{\mathcal{C}}$  on  $\mathbf{H}$  that will be “dragged and dropped” by the walk  $(\mathcal{C}, \mathcal{C}', \mathcal{C}'', \dots)$ . Applying the *second* variant of “drag and drop” we get a (finite or infinite) sequence  $(\bar{\mathcal{C}}, \bar{\mathcal{C}'}, \bar{\mathcal{C}''}, \dots)$  on  $\mathbf{H}$  and a (finite or infinite) sequence  $0, \bar{T}, \bar{T} + \bar{T}', \dots$  of times at which the jumps along  $\mathbf{H}$  occur. Note that these sequences may be finite even though the sequence  $(\mathcal{C}, \mathcal{C}', \mathcal{C}'', \dots)$  is infinite. In this way, we have defined a (continuous time) random process with the vertices of  $\mathbf{H}$  as configurations.

The first question that comes to mind is whether this process is Markov, and a moment thinking shows that the answer is no in general. Indeed, the waiting time between two jumps along  $\mathbf{H}$  is in general not exponential: it is a complicated mixture (we use the term in an informal way) of the exponential waiting times of the original Markov process.

However, some remnant of the Markov property is preserved. To state it properly, we apply our usual trick to the sequence of successive positions on  $\mathbf{H}$ : if the sequence  $(\bar{\mathcal{C}}, \bar{\mathcal{C}'}, \bar{\mathcal{C}''}, \dots)$  is finite, we turn it into an infinite one by repeating the last member again and again. We have the following (we do not try to give minimal hypotheses).

**Claim** Suppose that the initial point  $\bar{\mathcal{C}}$  on  $\mathbf{H}$  is recurrent for the Markov process on  $\mathbf{G}$ , i.e. with probability 1 configuration  $\bar{\mathcal{C}}$  appears infinitely many times in the sequence  $(\mathcal{C}, \mathcal{C}', \mathcal{C}'', \dots)$ . Then  $(\bar{\mathcal{C}}, \bar{\mathcal{C}'}, \bar{\mathcal{C}''}, \dots)$  is a Markov chain started at  $\bar{\mathcal{C}}$  with stochastic matrix  $\mathbb{P}_{\mathbf{H}}^{\text{dd}}$ .

**Remark** The hypothesis that  $\bar{\mathcal{C}}$  on  $\mathbf{H}$  is recurrent for the Markov process on  $\mathbf{G}$  is always fulfilled if micro-reversibility holds (in which case non-oriented graphs are the right thing to look at) and  $\mathbf{G}$  is connected.

**Sketch of proof** By hypothesis  $\bar{\mathcal{C}}$  is visited infinitely many times by the Markov process on  $\mathbf{G}$ .

- Either  $\bar{\mathcal{C}}$  is not the origin of an outgoing edge of  $\mathbf{H}$ , and then the sequence of positions on  $\mathbf{H}$  is just the repetition of  $\bar{\mathcal{C}}$  forever, which is indeed a Markov chain started at  $\bar{\mathcal{C}}$  because  $(\bar{\mathcal{C}}|\mathbb{P}_{\mathbf{H}}^{\text{dd}}|\bar{\mathcal{C}}) = 1$ ,
- Or let  $(\hat{\mathcal{C}}, \hat{\mathcal{C}}', \hat{\mathcal{C}}'', \dots)$  be the sequence of vertices of  $\mathbf{G}$  visited by jumping from  $\bar{\mathcal{C}}$ . These vertices are chosen independently and according to the probability measure  $(\cdot|\mathbb{P}|\bar{\mathcal{C}})$  on vertices of  $\mathbf{G}$ . The first time one of the jumps is along an edge of  $\mathbf{H}$ , this edge is sampled from the probability law  $(\cdot|\mathbb{P}|\bar{\mathcal{C}})$  conditioned to  $\mathbf{H}$  (see footnote 7), which is easily seen to be nothing but the probability law  $(\cdot|\mathbb{P}_{\mathbf{H}}^{\text{dd}}|\bar{\mathcal{C}})$ . Let  $\bar{\mathcal{C}'}$  be the end of the chosen edge. Now if  $\bar{\mathcal{C}}$  is visited infinitely many times by the Markov process on  $\mathbf{G}$ , then so must be  $\bar{\mathcal{C}'}$  and we can iterate the argument, which works whether or not  $\bar{\mathcal{C}'}$  is the origin of an outgoing edge of  $\mathbf{H}$ .

To conclude,  $(\bar{\mathcal{C}}, \bar{\mathcal{C}'}, \bar{\mathcal{C}''}, \dots)$  is a Markov chain started at  $\bar{\mathcal{C}}$  with transition matrix  $\mathbb{P}_{\mathbf{H}}^{\text{dd}}$ .

□

As for the times  $\bar{T}, \bar{T}', \dots$  between jumps along edges of  $\mathbf{H}$ , we content with an informal discussion. As already stated, they are not exponential in general (and their laws are quite complicated even for  $\mathbf{G}$  and  $\mathbf{H}$  of modest size). But deep independence properties survive.

**Claim** *The sequence  $((\bar{\mathcal{C}}, \bar{T}), (\bar{\mathcal{C}'}, \bar{T}'), (\bar{\mathcal{C}''}, \bar{T}'') \dots)$ , assumed to be infinite, is a Markov chain on the infinite configuration space  $\mathbb{C} \times ]0, +\infty[$  with a stochastic matrix of a very special form (observe that the transition matrix factorizes and contains no  $\bar{T}$ -dependence):*

$$(\bar{\mathcal{C}'}, \bar{T}' \mid \bar{\mathcal{C}}, \bar{T}) = (\bar{\mathcal{C}'}|\mathbb{P}_{\mathbf{H}}^{\text{dd}}|\bar{\mathcal{C}})\rho_{\bar{\mathcal{C}'}}(\bar{T}')$$

where  $\rho(t)$  is a family of probability densities on  $]0, +\infty[$  indexed by the vertices of  $\mathbf{H}$ .

This formulation is not fully correct, and without aiming at full rigour, we can be a bit more formal:

**Claim** *The sequence of pairs  $((\bar{\mathcal{C}}, \bar{T}), (\bar{\mathcal{C}'}, \bar{T}'), (\bar{\mathcal{C}''}, \bar{T}'') \dots)$  is a generalized renewal sequence.*

**Remark** By *generalized renewal sequence* we mean the following. We are given a finite set of configurations (here, the vertices of  $\mathbf{H}$ ), the stochastic matrix of some Markov chain on these configurations, (here,  $\mathbb{P}_{\mathbf{H}}^{\text{dd}}$ ) and laws  $v_., \mu_.$  (two for each configuration i.e. here, two for each vertex of  $\mathbf{H}$ ) for random variables with values in  $]0, +\infty[$ . Then starting from a configuration  $\bar{\mathcal{C}}$  (which can be chosen randomly), one waits a time  $\bar{T}$  distributed according to  $v_{\bar{\mathcal{C}}}$  and jumps to another configuration  $\bar{\mathcal{C}'}$  with probability  $(\bar{\mathcal{C}'}|\mathbb{P}_{\mathbf{H}}^{\text{dd}}|\bar{\mathcal{C}})$ . Then independently of what happened before (this is renewal) one waits at  $\bar{\mathcal{C}'}$  a time  $\bar{T}'$  distributed according to  $\mu_{\bar{\mathcal{C}'}}$  and jumps to another configuration  $\bar{\mathcal{C}''}$  with probability  $(\bar{\mathcal{C}''}|\mathbb{P}_{\mathbf{H}}^{\text{dd}}|\bar{\mathcal{C}'})$ . Then independently of what happened before one waits at  $\bar{\mathcal{C}''}$  a time  $\bar{T}''$  distributed according to  $\mu_{\bar{\mathcal{C}''}}$  and so on. Note that in general, the waiting time at the initial configuration (law  $v$ ) does not have to be distributed as the waiting time at further passages at the same configuration (law  $\mu$ ). In our case, this is relevant because we do not impose that the initial configuration  $\mathcal{C}$  of the original Markov process on  $\mathbf{G}$  be  $\bar{\mathcal{C}}$  so  $\bar{T}$  is the sum of two (independent) contributions: some time is

<sup>7</sup> If  $K, K', K'' \dots$  is a sequence of independent identically distributed random elements with law  $\mu$ , then the first element of the sequence that belongs to some measurable set  $B$  with  $\mu(B) > 0$  has law  $\mu(\cdot|B)$ , i.e.  $\mu$  conditioned on  $B$ . Indeed, if  $A \subset B$  is measurable the event that the first term belonging to  $B$  in fact belongs to  $A$  is the disjoint union  $\{K \in A\} \cup \{K \notin B, K' \in A\} \cup \{K \notin B, K' \notin B, K'' \in A\} \cup \dots$  which by independence has probability  $\mu(A) + (1 - \mu(B))\mu(A) + (1 - \mu(B))^2\mu(A) + \dots = \mu(A)/\mu(B)$ .

spent to reach  $\bar{\mathcal{C}}$  from  $\mathcal{C}$ , and then a time distributed according to  $\mu_{\bar{\mathcal{C}}}$  is spent before the jump to  $\bar{\mathcal{C}'}$ , so  $v_{\bar{\mathcal{C}}}$  is a composite object. But for all further passages at  $\bar{\mathcal{C}}$ , the waiting times to jump from  $\bar{\mathcal{C}}$  will have law  $\mu_{\bar{\mathcal{C}}}$ . Note that the density  $\rho_{\bar{\mathcal{C}}}$  above is the density of the law  $\mu_{\bar{\mathcal{C}}}$ . Note also that Markov processes form a particular class of generalized renewal sequences, those for which all waiting times are exponential.

We leave to the reader to deal with, i.e. make the appropriate conventions in the cases when the sequence is finite because one of the waiting times is infinite.

*Remark* Though we formulated the original process as a Markov process on  $\mathbf{G}$ , we could in fact have started from a generalized renewal process on  $\mathbf{G}$  as described informally above (in that case, there would be no pressing reason to singularize the first waiting time). All the arguments carry through, yielding a generalized renewal process on  $\mathbf{H}$ . This looks even more natural, because the continuous time Markov property is lost when going from  $\mathbf{G}$  to  $\mathbf{H}$  by drag and drop, but the renewal property is preserved.

We have kept the discussion quite general in this subsection, but for the application to affinities and cycles, we shall restrict to the case when  $\mathbf{H}$  is a micro-reversible cycle, i.e. a graph with  $\geq 3$  configurations arranged in cyclic order.

### 3.3 Conditioning for Markov Processes

This section is about a topic that is likely to be very classical, but we have not found a standard reference. The analogous computations for Markov chains are totally elementary because they rely solely on conditional probabilities. One could use a limiting argument to get, at least heuristically, the result for Markov processes from the one for Markov chains. The direct computations in continuous time require the use of conditional expectations.<sup>8</sup> They are lengthy and slightly more delicate, so we shall only give the results. Let us note that there are some intimate relations between conditioning and so-called taboo probabilities and taboo Green functions, see e.g. [4].

Suppose  $(\mathbf{C}, \mathbb{W})$  is a Markov process with associated Markov matrix  $\mathbb{M}$  and graph  $\mathbf{G}$ . As usual, let  $P$  denote the law of this process. Take a subgraph  $\mathbf{H}$  of  $\mathbf{G}$ . Fix  $T > 0$ . We want to describe the law of the process conditioned to move along  $\mathbf{H}$ , i.e. such that all jumps between time 0 and  $T$  are along edges of  $\mathbf{H}$ . Let  $P_{\mathbf{H},T}$  be the probability law for the conditioned process.

We recall that it is in principle easy to collect samples of  $P_{\mathbf{H},T}$ . One simply collects samples of the original process (with law  $P$ ) between 0 and  $T$  and keeps only those samples that obey the constraint that all jumps between time 0 and time  $T$  are along edges of  $\mathbf{H}$ .

The crucial ingredient is  $\Gamma_{\mathbf{H},T}(\mathcal{C}, \mathcal{C}')$  defined as the probability under the law  $P$  that a trajectory started at  $\mathcal{C}$  at time 0 ends at  $\mathcal{C}'$  at time  $T$  and jumps only along edges of  $\mathbf{H}$  in between. This is obviously 0 when either  $\mathcal{C}$  or  $\mathcal{C}'$  is not a vertex of  $\mathbf{H}$ .

**Claim** *Using standard manipulations of conditional expectations, the Markov property and homogeneity ( $\mathbb{W}$  is time-independent) of the original process, one can prove the following: for  $0 \leq s \leq t \leq T$ , the  $P_{\mathbf{H},T}$ -probability that  $\mathcal{C}_t = \mathcal{C}'$  when the process has been observed up to time  $s$  is given by*

$$P_{\mathbf{H},T}(\mathcal{C}_t = \mathcal{C}', \text{ knowing the past up to time } s) = \frac{\sum_{\mathcal{C}''} \Gamma_{\mathbf{H},t-s}(\mathcal{C}_s, \mathcal{C}') \Gamma_{\mathbf{H},T-t}(\mathcal{C}', \mathcal{C}'')}{\sum_{\mathcal{C}''} \Gamma_{\mathbf{H},T-s}(\mathcal{C}_s, \mathcal{C}'')}$$

<sup>8</sup> On conditional expectations, see the general references given at the beginning of Appendix 2.

Though the proof is not difficult, there are a number of steps and it would take us too far from our main interest. We advise the interested reader to build a proof in the context of Markov chains, instead of Markov processes.

The above formula leads to two comments:

- The right-hand side depends on the past only via  $C_s$ , meaning that the conditioned process is still Markov.
- Time  $T$  appears explicitly on the right-hand side. This may be surprising at first sight but a moment of thinking shows that this has to be so: even if we look only at conditioned trajectories up to time  $t$ , their law depends on how long the process is conditioned to move along  $\mathbf{H}$  after  $t$ .

If a trajectory has moved along  $\mathbf{H}$  between 0 and  $t$ , one computes easily the probability that between  $t$  and  $t + dt$  it will either stay where it is or jump along  $\mathbf{H}$ . Defining a matrix  $\mathbb{N}_{\mathbf{H}}$  by  $(C'|N_{\mathbf{H}}|C) = 0$  if either  $C$  or  $C'$  isn't a vertex of  $\mathbf{H}$ ,  $(C|\mathbb{N}_{\mathbf{H}}|C) = (C|\mathbb{M}|C)$  if  $C$  is a vertex of  $\mathbf{H}$  and  $(C'|\mathbb{N}_{\mathbf{H}}|C) = (C'|\mathbb{M}|C)$  if  $(C, C')$  is an edge of  $\mathbf{H}$ , we get

$$\frac{d}{dt} \Gamma_{\mathbf{H},t}(C, C') = \sum_{C''} (C'|\mathbb{N}_{\mathbf{H}}|C'') \Gamma_{\mathbf{H},t}(C, C''),$$

with initial condition  $\Gamma_{\mathbf{H},0}(C, C') = 1$  if  $C = C'$  is a vertex of  $\mathbf{H}$  and 0 otherwise.

Restricting  $\Gamma_{\mathbf{H},t}$  and  $\mathbb{N}_{\mathbf{H}}$  to vertices of  $\mathbf{H}$ , we have the formula

$$\Gamma_{\mathbf{H},t}(C, C') = (C' | e^{t\mathbb{N}_{\mathbf{H}}} | C).$$

Thus we have:

$$P_{\mathbf{H},T}(C_t = C', \text{ knowing the past up to time } s) = \frac{\sum_{C''} (C'' | e^{(T-s)\mathbb{N}_{\mathbf{H}}} | C') (C' | e^{(t-s)\mathbb{N}_{\mathbf{H}}} | C_s)}{\sum_{C''} (C'' | e^{(T-s)\mathbb{N}_{\mathbf{H}}} | C_s)}.$$

By the probabilistic interpretation, it is clear that all eigenvalues of  $\mathbb{N}_{\mathbf{H}}$  have real part  $\leq 0$ . Under appropriate conditions, Perron-Frobenius theory will provide the existence of a unique eigenvalue with maximal real part, say  $\lambda$  which will be real and correspond to left and right eigenvectors  $(\eta|$  and  $|\mu)$  with strictly positive components. In that case  $e^{T\mathbb{N}_{\mathbf{H}}} \sim e^{\lambda T} |\mu\rangle (\eta|$  for  $T \rightarrow +\infty$ , and

$$\begin{aligned} \lim_{T \rightarrow +\infty} P_{\mathbf{H},T}(C_t = C', \text{ knowing the past up to time } s) \\ = e^{-\lambda(t-s)} (\eta|C') (C' | e^{(t-s)\mathbb{N}_{\mathbf{H}}} | C_s) (\eta|C_s)^{-1}. \end{aligned}$$

This means precisely that conditioning the Markov process  $(\mathbf{C}, \mathbb{W})$  to move (forever) along the subgraph  $\mathbf{H}$  leads to a Markov process with transition Matrix  $\mathbb{W}_{\mathbf{H}}^{\text{cond}}$  (indexed by vertices of  $\mathbf{H}$ ) such that

$$(C' | \mathbb{W}_{\mathbf{H}}^{\text{cond}} | C) = (\eta|C') (C' | \mathbb{W} | C) (\eta|C)^{-1} \text{ if } (C, C') \text{ is an edge of } \mathbf{H} \text{ and } 0 \text{ otherwise.} \quad (3.1)$$

Indeed, it is easy to check that the Markov matrix  $\mathbb{M}_{\mathbf{H}}^{\text{cond}}$  associated to  $\mathbb{W}_{\mathbf{H}}^{\text{cond}}$  is such that  $(C' | e^{t\mathbb{M}_{\mathbf{H}}^{\text{cond}}} | C) = e^{-\lambda t} (\eta|C') (C' | e^{t\mathbb{N}_{\mathbf{H}}} | C) (\eta|C)^{-1}$ .

Note that Perron-Frobenius theory applies if  $\mathbf{H}$  is connected. We do not prove this but observe that if  $\mathbf{H}$  is connected an argument similar to the one given at the end of Appendix 3 for  $e^{t\mathbb{M}}$  for Markov processes shows that  $e^{t\mathbb{N}_{\mathbf{H}}}$  has, for every  $t > 0$ , strictly positive matrix elements between vertices of  $\mathbf{H}$ .

If  $\mathbf{H}$  splits as a disjoint union of connected components, Perron-Frobenius theory will apply to each component separately and a limiting conditioned process on  $\mathbf{H}$  will exist for

$T \rightarrow +\infty$ : the main difference is that  $\lambda$  will vary from one connected component to another. But the formula (3.1) will survive for an appropriate collection  $(\eta|\mathcal{C})$  made by gathering the dominant eigenvectors for each connected component.

### 3.4 A Summary of Affinity Class Invariances

We give a (non-exhaustive) list of probabilistic constructions that preserve the affinity class. Most relevant for us are “drag and drop” and conditioning.

#### 3.4.1 From a Markov Process to a Markov Chain

As recalled in Appendix 3, to every Markov process  $(\mathbf{C}, \mathbb{W})$  one can associate a Markov chain  $(\mathbf{C}, \mathbb{P})$  such that  $\mathbb{M} = (\mathbb{P} - \mathbb{I})\mathbb{D}$  where  $-\mathbb{D}$  is the diagonal part of  $\mathbb{M} = \mathbb{W} - \mathbb{D}$ . Then  $\mathbb{W}$  and  $\mathbb{P}$  define the same graph, and when defined (i.e. for edges)

$$\frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')} = \frac{(\mathcal{C}'|\mathbb{P}|\mathcal{C})(\mathcal{C}|\mathbb{D}|\mathcal{C})}{(\mathcal{C}|\mathbb{P}|\mathcal{C}')(C'|\mathbb{D}|\mathcal{C}')}.$$

Taking the logarithm on both sides shows that affinity and its discrete analog (2.5) define the same class.

Our next aim is the behavior of the affinity class on subgraphs. We start from a Markov process  $(\mathbf{C}, \mathbb{W})$  with associated graph  $\mathbf{G}$ , and take a subgraph  $\mathbf{H}$  of  $\mathbf{G}$ , both assumed to be micro-reversible.

#### 3.4.2 Restriction

This construction is trivial, but has no particular probabilistic meaning: let  $\mathbb{W}_{\mathbf{H}}^{\text{rest}}$  be the restriction of the transition matrix to  $\mathbf{H}$ . So  $\mathbb{W}_{\mathbf{H}}^{\text{rest}}$  is a matrix indexed by the vertices of  $\mathbf{H}$  and

$$(\mathcal{C}'|\mathbb{W}_{\mathbf{H}}^{\text{rest}}|\mathcal{C}) = (\mathcal{C}'|\mathbb{W}|\mathcal{C}) \text{ if } (\mathcal{C}, \mathcal{C}') \text{ is an edge of } \mathbf{H} \text{ and } 0 \text{ otherwise.}$$

From  $\mathbb{W}_{\mathbf{H}}^{\text{rest}}$  we can define the exchange  $\mathbb{S}_{\mathbf{H}}^{\text{rest}}$ , which is the restriction of  $\mathbb{S}$  to edges of  $\mathbf{H}$ . Trivially, the integral of  $\mathbb{S}_{\mathbf{H}}^{\text{rest}}$  or  $\mathbb{S}$  along any walk on  $\mathbf{H}$  gives the same result.

#### 3.4.3 Invariance Under “Drag and Drop”

The intuitive definition of “drag and drop” (defined formally in Section 3.1, which should be consulted for reference) is the following. At  $t = 0$  put a box at some vertex of  $\mathbf{H}$  and start a walker on  $\mathbf{G}$ . The box remains at its vertex until the walker gets there. When the walker and the box are at the same vertex and the next jump of the walker is along an edge of  $\mathbf{H}$ , the walker drags the box along. This goes on as long as the jumps of the walker are along  $\mathbf{H}$ . But as soon as the walker prepares for a jump along an edge in  $\mathbf{G}$  but not  $\mathbf{H}$  he drops the box. Then the box remains immobile and waits for the next encounter with the walker and so on. So by successive drags and drops, the box itself describes a walk on  $\mathbf{H}$ . There is an analogy that the reader may find useful between hitch-hiking and a variant of “drag and drop” (see the end of Sect. 3.1).

As explained in detail in Sect. 3.2, under the procedure of “drag and drop”, if the walker samples a Markov process on  $\mathbf{G}$ , the trajectory of the box samples a generalized renewal process on  $\mathbf{H}$  and in particular a Markov chain on  $\mathbf{H}$ . The corresponding stochastic matrix  $\mathbb{P}_{\mathbf{H}}^{\text{dd}}$  is the Markov chain associated to the Markov process associated to the restriction  $\mathbb{W}_{\mathbf{H}}^{\text{rest}}$ .

of  $\mathbb{W}$  to  $\mathbf{H}$ . Restriction preserves edge affinities on  $\mathbf{H}$ , and going from a Markov process to its Markov chain may change edge affinities, but preserves the affinity class.

So the cycle affinities are left invariant under “drag and drop”.

### 3.4.4 Invariance Under Conditioning

As explained in detail in Appendix 2.3, if the Markov process  $(\mathbf{C}, \mathbb{W})$  is conditioned to have all its jumps along edges of  $\mathbf{H}$ , one ends up with a new Markov process, the configuration space being made of the vertices of  $\mathbf{H}$  and the transition matrix  $\mathbb{W}_{\mathbf{H}}^{\text{cond}}$  being defined by (3.1) which we repeat here for convenience:  $\mathbb{W}_{\mathbf{H}}^{\text{cond}}$  is a matrix indexed by the vertices of  $\mathbf{H}$

$$(\mathcal{C}'|\mathbb{W}_{\mathbf{H}}^{\text{cond}}|\mathcal{C}) = (\eta|\mathcal{C}')(\mathcal{C}'|\mathbb{W}|\mathcal{C})(\eta|\mathcal{C})^{-1} \text{ if } (\mathcal{C}, \mathcal{C}') \text{ is an edge of } \mathbf{H} \text{ and } 0 \text{ otherwise,}$$

where  $(\eta|\mathcal{C})$  is a strictly positive function on the set of vertices of  $\mathbf{H}$ . It is plain that  $\mathbb{W}_{\mathbf{H}}^{\text{rest}}$  and  $\mathbb{W}_{\mathbf{H}}^{\text{cond}}$  define the same affinity class.

So the cycle affinities are left invariant under conditioning. Let us recall that to collect sample of the conditional probability one simply collects samples of the original process (with law  $P$ ) and keeps only those samples that obey the constraint that all jumps are along edges of  $\mathbf{H}$ .

## 3.5 Discussion

Let us pause for a moment to say a few words about the respective merits of conditioning and “drag and drop”. Conditioning seems at first sight to be more natural, as it requires no extraneous structure (drag and drop introduces another object, the box) and is defined for general micro-reversible  $\mathbf{H}$ . However, conditioning is meant above as holding at all times (from 0 to  $+\infty$ ): it is obtained via a limiting procedure from finite time conditioning over  $[0, T]$  (see the details in 3.3). For finite time conditioning, there are corrections to the time homogeneous description. As  $T \rightarrow +\infty$  these corrections are generically exponentially small, but the measure of the trajectories staying on  $\mathbf{H}$  is exponentially small as well. So the number of samples of the unconditioned system from which one selects the ones fulfilling the constraint must be exponentially large. Hence the perspectives of measuring the conditioned quantities are unclear (unless there is an unbiased physical way to implement the constraint instead of simply discarding the samples which do not fulfill them). On the other hand, “drag and drop” needs to identify physically an extra structure (the box), and the “drag and drop” motion is slower than the original motion, but does not suffer of the exponential penalty of conditioning. However, the “drag and drop” motion is described by a renewal process only when the vertices of  $\mathbf{H}$  have certain recurrence properties for the original Markov process trajectories. We have mentioned in Sect. 3.1 that “drag and drop” possibly occurs naturally in biological systems, and we could imagine many other situations, in work-flows for instance. But we must admit that we have no natural interpretation in the case of general non-equilibrium systems. However, one may hope that for certain systems in which the dynamics is related to exchanges with specific reservoirs, these exchanges give a clue to identify one or several objects for which “drag and drop” is naturally observable.

We do not pursue this route here, and simply view conditioning and “drag and drop” as important examples spelled in detail, showing that the affinity class, and the corresponding cycle affinities, are remarkably robust observables that remain invariant under a number of natural probabilistic constructions. These two examples allow to observe processes or chains on subgraphs in terms of processes or chains on graphs. Among all subgraphs, the simplest, most basic but somehow most interesting for non equilibrium physics, subgraphs are cycles.

In this context, we call random motion on a cycle a “norria”. Indeed, when a particle moves along a cycle, one may keep track of its position, but also of its winding number, i.e. one can count how many times the cycle has been described in a given time interval, i.e. how many turns the norria has made. We shall see that (cycle) affinity is the crucial quantity that describes the efficiency of the norria, i.e. how fast the turns are made in average, in which direction, and what is the size of fluctuations. To insist on the existence of fluctuations, we might call it more precisely a mesoscopic norria. We now turn to their detailed study, i.e. assume that the subgraph  $\mathbf{H}$  is a cycle.

## 4 A Probabilistic Interpretation of Cycle Affinities in norias

In this section, we shall use repeatedly, and often without even mentioning it, that the strong Markov property (for a precise definition, see the general references at the beginning of Appendix 2) holds for finite state Markov chains and finite state Markov processes. Let us just describe informally what this statement means.

The Markov property says that knowing the present (i.e. the configuration at a given time  $t$ ,  $t = 0, 1, \dots$  for chains,  $t \in [0, +\infty[$  for processes) the trajectory before  $t$  and after  $t$  are independent.

But what if  $t$  is replaced by a random time  $T$ ? For example, fix a configuration  $\mathcal{C}$  and let  $T$  be the first time the trajectory visits  $\mathcal{C}$ , i.e.  $T$  is the smallest  $s$  such that  $\mathcal{C}_s = \mathcal{C}$ . The strong Markov property states that  $\mathcal{C}_{T+t}$ , ( $t = 0, 1, \dots$  for chains,  $t \in [0, +\infty[$  for processes) has exactly the same law as the original process started at  $\mathcal{C}$ , i.e. with  $v(\mathcal{C}_0) = \delta_{\mathcal{C}_0, \mathcal{C}}$ .

More generally, for a strong Markov process, this identity in law holds whenever  $T$  is a stopping time (again, for a precise definition, see the general references at the beginning of Appendix 2). Intuitively, a stopping time  $T$  is a random time for which one can decide if  $T \leq s$  by looking at the trajectory  $\mathcal{C}_t$ ,  $t \in [0, s]$ , i.e. without knowledge of the future of  $s$ . A deterministic time is a stopping time, so the strong Markov property is in principle stronger than the Markov property.

It is a theorem (and not a triviality) that finite state Markov chains and finite state Markov processes do have the strong Markov property.

### 4.1 Inhomogeneous Random Walks on Cycles

The general setting we need to study cycle affinities is a system with  $M \geq 3$  configurations labeled  $1, \dots, M$  (configuration  $m$  and  $m + M$  are identified in all subsequent formulæ). We choose arbitrarily an orientation of the cycle: a jump from  $m$  to  $m + 1$  (resp.  $m - 1$ ) is said to be clockwise (resp. anti-clockwise). We view the process as the motion of a particle along the cycle, jumping from time to time from a site to one of its two neighbors. As established in the previous subsection, whatever the initial Markov process was, as far as the cycle  $\mathbf{H}$  in the initial graph is concerned, the rules are of the following kind.

At  $t = 0$  the particle starts at some  $m$  on the cycle. We make the convention that the first waiting time does not play a special role. This is automatic for conditioning, but not for drag and drop (in that case, either we start the walker at  $m$ , or start time at the first passage of the walker at  $m$ ). The particle waits at  $m$  for a (random) time  $T$  (whose law depends only on  $m$ ) and then jumps to  $m' = m \pm 1$  with probability  $p_m^\pm$ . It waits there for a random time  $T'$  (whose law depends on the past only via the position  $m'$ ) and at  $T + T'$  jumps to  $m'' = m' \pm 1$  with probability  $p_{m'}^\pm$ . It waits there for a random time  $T''$  (whose law depends on the past only via the position  $m''$ ) and so on.

In particular the successive positions  $m, m', m'', \dots$  form a Markov chain with a stochastic matrix whose only non-vanishing elements are  $(m \pm 1 | \mathbb{P} | m) = p_m^\pm$  (use the periodic boundary conditions when necessary) with  $p_m^\pm > 0$  and  $p_m^+ + p_m^- = 1$  for  $m = 1, \dots, M$ . The law of the random time spent at  $m$  could in principle be computed for each  $m$  from the knowledge of the original Markov process, a rather simple task for conditioning (where all waiting times are exponential and with easily computable parameters), but a very involved one for “drag and drop”.

#### 4.1.1 Winding Numbers

We define a sequence  $W_k$  as follows:  $MW_k$  is the number of clockwise jumps minus the number of anti-clockwise jumps among the  $k$  first jumps. The process  $(MW_k)_{k=0,1,\dots}$  is an inhomogeneous random walk on the integers, the inhomogeneity being periodic in space with period  $M$ . If the initial position was  $m$ ,  $m + MW_k$  (taken modulo  $M$ ) is the position just after the  $k^{\text{th}}$  jump. But  $W_k$  carries more information: by construction,  $W_k$  changes by  $-1, 0$  or  $1$  between two successive passages at  $m$  for  $m = 1, \dots, M$ , so it keeps a memory of the number of turns. This is the reason why we call  $W_k$  the winding number up to (and including) the  $k^{\text{th}}$  jump.

There are two natural ways to measure time: time in the original Markov process,  $t$ , and time as number of steps in the cycle Markov chain,  $k$ . If  $t$  belongs to the interval between jump  $k$  and jump  $k + 1$ , so that the winding number is  $W_k$ , we set  $\tilde{W}_t \equiv W_k$ . Observe that  $\tilde{W}_t, t \in [0, +\infty]$  and by extension  $W_k, k = 0, 1, 2, \dots$  are exchange processes as defined in 2.2.

The first question we shall ask is the probability that the winding number will ever reach a certain value. This question does not involve a time parameterisation, and has a simple answer.

The second question will be about the distribution of the time it takes to reach a certain winding number, and this question, closely related to the efficiency of the noria, depends on the time parameterisation. The answer is more complicated and less explicit, but exhibits some simple general symmetries, strongly reminiscent of out-of equilibrium relations.

The quantity  $p_m^+ / p_{m+1}^-$  quantifies the relative tendency to traverse edge  $(m, m + 1)$  in one direction or another. In the rest of this section, we set

$$e^A \equiv \prod_{m=1}^M \frac{p_m^+}{p_m^-}$$

and see directly that  $A$  is the affinity along the cycle oriented clockwise. Thereafter we use the simple name “affinity” for the quantity  $A$ .

#### 4.1.2 Algebraic Preliminary

We start with an algebraic preliminary which is the crucial ingredient for the following elaborations.

We suppose given a collection of indeterminates  $x_m, m = 1, \dots, M$ . We look for the solution(s) of the system

$$f_m^- = x_m(p_m^- + p_m^+ f_{m+1}^- f_m^-), \quad (4.1)$$

with periodic boundary conditions, for the unknown  $f_m^-, m = 1, \dots, M$ , and for the solution(s) of the system

$$f_m^+ = x_m(p_m^+ + p_m^- f_{m-1}^+ f_m^+), \quad (4.2)$$

with periodic boundary conditions, for the unknown  $f_m^+, m = 1, \dots, M$ .

Later, we shall give a probabilistic interpretation of these systems, where the indeterminates  $x_m$  will be real numbers in  $[0, 1]$ , or complex numbers in the closed unit disc. We set  $F^- \equiv \prod_{m=1}^M f_m^-$  and  $F^+ \equiv \prod_{m=1}^M f_m^+$ .

**Claim** *Each system (4.1) or (4.2) has generically two solutions. In particular,  $F^+$  as well as  $F^-$  can take two values. Moreover, either  $F^- F^+ = 1$  or  $F^- = F^+ e^{-A}$ .*

*Proof* We start with system (4.1). This system is a discrete analog of a Riccati differential equation, and it can be linearized by the standard trick. Let  $g_0$  be arbitrary and define iteratively  $g_m \equiv f_m^- g_{m-1}$  for  $m = 1, \dots, M, M + 1$ . Observe that  $g_M = F^- g_0$  and  $g_{M+1} = F^- g_1$ : the sequence  $f_m$  is periodic by construction but the sequence  $g_m$  is not, and  $F^-$  appears as an holonomy. Then (4.1) turns into

$$g_m = x_m(p_m^- g_{m-1} + p_m^+ g_{m+1}) \text{ for } m = 1, \dots, M. \quad (4.3)$$

This two-terms recursion relation can be transformed in a vector one-term recursion relation:

$$\begin{pmatrix} g_{m+1} \\ g_m \end{pmatrix} = \mathbb{F}_m \begin{pmatrix} g_m \\ g_{m-1} \end{pmatrix}$$

for  $m = 1, \dots, M$ , where

$$\mathbb{F}_m \equiv \begin{pmatrix} 1/(x_m p_m^+) & -p_m^-/p_m^+ \\ 1 & 0 \end{pmatrix}. \quad (4.4)$$

We set

$$\mathbb{F} = \begin{pmatrix} \mathbb{F}_{11} & \mathbb{F}_{12} \\ \mathbb{F}_{21} & \mathbb{F}_{22} \end{pmatrix} \equiv \mathbb{F}_M \cdots \mathbb{F}_1.$$

One finds by iterating the above formula that

$$F^- \begin{pmatrix} g_1 \\ g_0 \end{pmatrix} = \begin{pmatrix} g_{M+1} \\ g_M \end{pmatrix} = \mathbb{F} \begin{pmatrix} g_1 \\ g_0 \end{pmatrix} \quad (4.5)$$

so that  $F^-$  is an eigenvalue of the transfer (or Bloch-Floquet or  $\dots$  depending on the community) matrix  $\mathbb{F}$ , i.e. a solution of

$$(F^-)^2 - F^- \operatorname{Tr} \mathbb{F} + \operatorname{Det} \mathbb{F} = 0. \quad (4.6)$$

This formula shows clearly that  $F^-$  is a cyclic invariant, because  $\operatorname{Tr} \mathbb{F}$  and  $\operatorname{Det} \mathbb{F}$  are. The trace  $\operatorname{Tr} \mathbb{F}$  has a complicated expression in general, but the determinant  $\operatorname{Det} \mathbb{F} = \prod_{m=1}^M \operatorname{Det} \mathbb{F}_m = \prod_{m=1}^M p_m^-/p_m^+ = e^{-A}$ , a simple function of the cycle affinity.

Eq. (4.6) is a quadratic equation, which has generically two solutions. If one is chosen, the ratio of the components of the corresponding eigenvector  $g_1/g_0 = f_1^-$  is fixed, namely  $f_1^- = (F^- - \mathbb{F}_{11})/\mathbb{F}_{21}$ , and then all other  $f_i^-$ 's as well using (4.1).

To summarize, we have shown that the system (4.1) has generically exactly two solutions.

The system (4.2) can be solved in an analogous way. It is easily seen that if one takes an arbitrary  $g_{M+1}$  and defines iteratively  $g_m \equiv f_m^+ g_{m+1}$  for  $m = M, \dots, 1, 0$  the  $g$ 's do again satisfy (4.3) (though they may differ from the  $g$ 's introduced using the  $f^-$ 's). This time  $g_0 = F^+ g_M$  and  $g_1 = F^+ g_{M+1}$  and we infer that

$$(F^+)^{-2} - (F^+)^{-1} \operatorname{Tr} \mathbb{F} + \operatorname{Det} \mathbb{F} = 0. \quad (4.7)$$

So  $F^-$  and  $1/F^+$  are roots of the same quadratic equation, and there are two possibilities: either  $F^- F^+ = 1$  or  $F^- = F^+ \operatorname{Det} \mathbb{F}$ . As  $\operatorname{Det} \mathbb{F} = e^{-A}$  the proof is completed.

□

We shall see in the sequel (see Sects. 4.1.4 and 4.1.5) that for our probabilistic aims, the second possibility, namely  $F^- = F^+ \text{Det } \mathbb{F}$ , is the relevant one. Then an explicit computation yields

$$\frac{f_1^-}{f_0^+} = -\frac{\mathbb{F}_{12}}{\mathbb{F}_{21}}. \quad (4.8)$$

The ratios  $f_m^-/f_{m-1}^+$  are given by analogous formulae obtained by replacing the product  $\mathbb{F} \equiv \mathbb{F}_M \cdots \mathbb{F}_1$  by the appropriate circular permutation. Apart from the fact that these ratios involve no square root and are plain rational functions of all the arguments, the explicit expressions are complicated in general.

We return now to our primary interest, the cyclic Markov chain. The strategy in the following sections is always the same: use the (strong) Markov property to obtain recursion relations of the type studied above among the quantities of interest, then use the periodicity along the cycle to solve the recursion relations.

#### 4.1.3 Clockwise and Anti-Clockwise Cycles

The easiest question to answer is very classical and its solution can be found in textbooks (though the proof we give is not totally standard): will the winding number will ever reach  $\pm 1$ ?

**Claim** *If  $A \geq 0$  (resp.  $\leq 0$ ) the probability to reach winding number 1 is 1 (resp.  $e^A$ ) and the probability to reach winding number  $-1$  is  $e^{-A}$  (resp. 1).*

*The complete proof has to wait until 4.1.4, but for orientation we give a very simple argument leading to a slightly weaker result. The (strong) Markov property is the crucial ingredient.*

*Sketch of proof* Let  $\pi_m^-$  be the probability that the winding number of a trajectory started at  $m$  ever reaches the value  $-1/M$ . Then by the Markov property

$$\pi_m^- = p_m^- + p_m^+ \pi_{m+1}^- \pi_m^-. \quad (4.9)$$

The meaning of this equation is clear : either the first jump is anti-clockwise (probability  $= p_m^-$ ) and the winding number reaches its target  $-1/M$  or the first jump is clockwise (probability  $= p_m^+$ ), and then the particle has “lost” a winding  $1/M$  so it has to go from  $m+1$  to  $m$  with winding number  $-1/M$  to compensate (probability  $\pi_{m+1}^-$ ), and take a new chance.

Notice that (4.9) is simply (4.1) with all  $x_m$ ’s replaced by 1. In this simple special case, an ad hoc argument works because  $f_m^- = 1$  for  $m = 1, \dots, M$  is obviously a solution. The probability that starting from  $m$  the winding number reaches  $-1$  is, by the Markov property again,  $\prod_{l=0}^{M-1} \pi_{m-l}^-$ , which is independent of  $m$ . So we denote this probability simply by  $\Pi^- \equiv \prod_{l=0}^{M-1} \pi_{m-l}^-$ . We rewrite (4.9) as

$$p_m^- (1 - \pi_m^-) = p_m^+ \pi_m^- (1 - \pi_{m+1}^-). \quad (4.10)$$

A first consequence is that if  $\pi_m^- = 1$  for some  $m$  then also  $\pi_{m+1}^- = 1$  and so on, so that all  $\pi_m^-$ ’s are equal to one, and the probability to reach winding number  $-1$  is unity. On the other hand, in the case when no  $\pi_m^-$  equals one,

$$\Pi^- = \prod_{m=1}^M \frac{p_m^-}{p_m^+} = e^{-A}. \quad (4.11)$$

Indeed, in (4.10) take the product over all  $m$ 's in the cycle and simplify both sides by  $\prod_{m=1}^M (1 - \pi_m^-) \neq 0$  to get  $\prod_{m=1}^M p_m^- = \prod_{m=1}^M p_m^+ \prod_{m=1}^M \pi_m^-, i.e. \Pi^- = \prod_{m=1}^M p_m^- / p_m^+$ . As  $\Pi^-$  is a probability, this is possible only if  $\prod_{m=1}^M p_m^- / p_m^+ \leq 1$ .

We have already proved the following: if  $A \leq 0$  then  $\Pi^- = 1$ , i.e. winding number  $-1$  is reached with probability 1, and if  $A \geq 0$  then  $\Pi^- \in \{1, e^{-A}\}$ . We could reproduce the above argument with  $\pi_m^+$ , the probability that the winding of a trajectory started at  $m$  ever reaches the value  $1/M$ , and  $\Pi^+ \equiv \prod_{l=0}^{M-1} \pi_{m+l}^+$ . We would get : if  $A \geq 0$  then  $\Pi^+ = 1$ , i.e. winding number 1 is reached with probability 1 and if  $A \leq 0$  then  $\Pi^+ \in \{1, e^A\}$ . In particular, if  $A = 0$  then the probability to reach winding numbers  $-1$  and 1 is unity, and by the (strong) Markov property, the probability to reach any winding number an infinite number of times is also unity: the winding number  $W_k$  will oscillate and take arbitrarily large positive and negative values as  $k \rightarrow +\infty$ .

□

These arguments suggest that in fact

$$\Pi^- = \min\{1, e^{-A}\} \text{ and } \Pi^+ = \min\{1, e^A\}. \quad (4.12)$$

Though it is intuitive that if  $A > 0$  there is a systematic drift towards positive winding number, hence a finite probability to never reach winding number  $-1$ , it takes a deeper argument to get a proof. This is our next aim.

#### 4.1.4 First Passage Times: Markov Chain Case

We generalize a bit the previous discussion. We let  $F^-$  (resp.  $F^+$ ) be the generating function for the number of steps it takes to reach winding number  $-1$  (resp. 1), the parameter being  $z$ . By definition,  $F^- \equiv \sum_{k \geq 1} F_k^- z^k$  where  $F_k^-$  is the probability that it takes exactly  $k$  steps to reach winding number  $-1$ . In the same way  $F^+ \equiv \sum_{k \geq 1} F_k^+ z^k$  where  $F_k^+$  is the probability that it takes exactly  $k$  steps to reach winding number 1. Note that  $F^\pm(z=1) = \Pi^\pm$ .

Defining  $f_m^-$  to be the generating function for the number of steps it takes to reach winding number  $-1/M$  starting from  $m$ , the following holds.

**Claim** *The  $f_m^-$ 's solve the system (4.1) for all  $x_m$ 's equal to  $z$ , namely*

$$f_m^- = z(p_m^- + p_m^+ f_{m+1}^- f_m^-) \text{ for } m = 1, \dots, M \text{ with periodic boundary conditions,} \quad (4.13)$$

and  $F^- = \prod_{m=1}^M f_m^-$ .

*Proof* That the  $f_m^-$ 's are periodic modulo  $M$  is included in their very definition. Consider the right-hand side of (4.13). The factor  $z$  is the weight of the first step. If the first step is anti-clockwise (probability  $p_m^-$ ) winding number  $-1/M$  is reached after one step. Otherwise (probability  $p_m^+$ ) the first step takes to winding number  $1/M$  and then the number of steps to reach winding number  $-1/M$  is, by the (strong) Markov property, the sum of two independent contributions, one with generating function  $f_{m+1}^-$  and one with generating function  $f_m^-$ . Finally,  $F^- = \prod_{m=1}^M f_m^-$ , because  $F^-$  is the generating function for a sum of  $M$  independent (the strong Markov property again) random variables with generating functions  $f_m^-$ ,  $m = 1, \dots, M$ .

□

Of course an analogous results holds for  $F^+$ . For the rest of this discussion we set  $x_m = z$  for all  $m$ 's in the formulæ from 4.1.2.

### Claim of Our Main Result (Markov Chain Case)

The generating functions  $F^-$  and  $F^+$  satisfy  $F^- = F^+e^{-A}$ , and in particular (4.12) holds.

*Proof* The definition of  $F^-$  as a normally convergent sum in the closed unit disc  $|z| \leq 1$  shows immediately that  $F^-$  is in fact an analytic function in the open unit disc  $|z| < 1$ , continuous up to the boundary i.e. in the closed unit disc  $|z| \leq 1$  and the same holds for each  $F_m^-$ . Moreover  $F^-(z=1) = \Pi^-$ , while by definition  $f_m^-(z) \sim p_m^- z$  at small  $z$ .

As  $F^-$  is real for  $z \in ]0, 1]$ , obviously the discriminant of the quadratic equation (4.6) has to be  $\geq 0$  in this parameter range.

Now, at small  $z$ , the upper left corner is the dominant term in each  $\mathbb{F}_m$  (defined in (4.4) where  $z$  is substituted for  $x_m$ ) so that  $\text{Tr } \mathbb{F} \sim z^{-M}/(\prod_{m=1}^M p_m^+)$  at small  $z$ . Taking  $z$  small and positive, we see that to prevent explosion the branch in the solution of (4.6) has to be

$$F^- = \frac{\text{Tr } \mathbb{F} - \sqrt{(\text{Tr } \mathbb{F})^2 - 4\text{Det } \mathbb{F}}}{2}, \quad (4.14)$$

where the determination of the square-root is the analytic continuation of the positive square root at small  $z$ .

The generating function  $F^+$  for the number of steps it takes to reach winding number 1, the parameter being  $z$ , is related to (4.2) in the same way. Now  $F^+$  is small at small  $z$ , so that the relevant solution of (4.7) at small  $z$  is

$$\frac{1}{F^+} = \frac{\text{Tr } \mathbb{F} + \sqrt{(\text{Tr } \mathbb{F})^2 - 4\text{Det } \mathbb{F}}}{2}. \quad (4.15)$$

where the determination of the square-root is the analytic continuation of the positive square root at small  $z$ . In particular we find that

$$\frac{F^-}{F^+} = \text{Det } \mathbb{F} = e^{-A}$$

is valid everywhere by analytic continuation. As a consequence, taking  $z \rightarrow 1^-$  one finds

$$\frac{\Pi^-}{\Pi^+} = \text{Det } \mathbb{F} = e^{-A},$$

which together with the results in 4.1.3 proves (4.12). □

A few remarks are in order:

- The proof that we have given for the identity  $F^- = F^+e^{-A}$  is analytic (via generating functions and singularity analysis), but it would deserve a *good* combinatorial proof. It is clear that if we compared the generating functions for winding number  $\pm 1$  (without the restriction that it is the first passage) we would obtain a one-to-one correspondence preserving length (i.e. the number of steps) by simply reversing paths, and the ratio would obviously be  $\prod_{m=1}^M p_m^-/p_m^+$ . In the same way, if we compared generating functions for winding number  $-1/M$  starting at  $m$  and for winding number  $1/M$  starting at  $m-1$  (without the restriction that it is the first passage) the ratio would be  $p_m^-/p_{m-1}^+$ . Whereas the ratio remains simple (and the same) in the first case (full cycle) when we restrict to first passages, the ratio becomes more involved in the second case (piece of cycle) when we restrict to first passages: by (4.8), the ratios  $f_m^-/f_{m-1}^+$  are given by explicit but tedious formulæ, with one salient feature however, they are rational functions of the parameters of the Markov chain.

- The proof we have given that if  $\prod_{m=1}^M p_m^- / p_m^+ < 1$  there is finite probability to never reach winding number  $-1$  is also of analytic nature, but would deserve a purely probabilistic proof. This could be obtained either as a consequence of the individual ergodic theorem or of the renewal theorem, or of other refinements of the strong law of large numbers.

#### 4.1.5 First Passage Times: General Case

In the beginning of 4.1 we have explained why concentrating on a cycle for general finite state Markov processes leads to a Markov chain on the cycle and to waiting times between the jumps which depend only on the position on the cycle but are in general not exponential.

So take  $x_m \equiv x_m(\lambda)$  as the Laplace transform of the law of the time during which the particle remains at  $m$  before jumping. If the probability that this time belongs to  $[0, t]$  is  $u_m(t)$  then  $x_m(\lambda) \equiv \int_0^{+\infty} e^{-\lambda t} du_m(t)$ . As the  $u_m$ 's are derived (though in a complicated way for “drag and drop”) from the exponential waiting times of a finite state Markov process, they are quite nice, i.e. continuous, vanishing at 0 and going to 1 exponentially fast at large  $t$  so that  $x_m(\lambda)$  has finite derivatives of all orders (giving the moments) at  $\lambda = 0$  and  $\lim_{\lambda \rightarrow +\infty} x_m(\lambda) = 0$ .

Let  $T^-$  be the (random) time it takes to reach winding number  $-1$  and  $T^+$  be the (random) time it takes to reach winding 1. We do not exclude *a priori* that for some trajectories, the particle never reaches winding number  $-1$  or  $+1$  and on such events we define  $T^-$  or  $T^+$  to be  $+\infty$ .

Let  $f_m^- \equiv f_m^-(\lambda)$  (resp.  $F^- \equiv F^-(\lambda)$ ) stand for the Laplace transform of the law of the time it takes to reach winding number  $-1/M$  starting from  $m$  (resp. to reach winding number  $-1$ ), and let  $f_m^+ \equiv f_m^+(\lambda)$  (resp.  $F^+ \equiv F^+(\lambda)$ ) stand for the Laplace transform of the law of the time it takes to reach winding number  $1/M$  starting from  $m$  (resp. to reach winding number 1). In equations,  $F^-(\lambda) \equiv \int_0^{+\infty} e^{-\lambda t} dP(T^- < t)$  and  $F^+(\lambda) \equiv \int_0^{+\infty} e^{-\lambda t} dP(T^+ < t)$ . In these definitions, time stands for the real time of the original Markov process, not for the number of steps.

##### *Claim of Our Main Result (General Case)*

The functions  $f_m^-(\lambda)$  [resp.  $f_m^+(\lambda)$ ] solve the system (4.1) (resp. (4.2)) with data  $x_m(\lambda)$ . The functions  $F^-(\lambda)$  and  $F^+(\lambda)$  are related by  $F^-(\lambda) = F^+(\lambda)e^{-A}$ .

*Proof* To show that the functions  $f_m^-(\lambda)$  (resp.  $f_m^+(\lambda)$ ) solve the system (4.1) (resp. (4.2)) with data  $x_m(\lambda)$ , the same probabilistic argument as before can be repeated word for word.

Because  $\lim_{\lambda \rightarrow +\infty} x_m(\lambda) = 0$ , the same simple argument as before allows to choose the appropriate branch for the quadratic equation. Hence the following relation holds between the Laplace transforms  $F^\pm(\lambda)$  of the laws of the time it takes to reach winding number  $\pm 1$ :

$$\frac{F^-(\lambda)}{F^+(\lambda)} = \text{Det } \mathbb{F} = e^{-A}, \quad (4.16)$$

so that the cycle affinity still governs the relationship between the distribution of first passage times at winding numbers  $-1$  and  $1$ . The ratio is independent from the  $x_m$ 's, namely from the distributions of the waiting times at the various positions on the cycle. □

By (4.8), the ratios  $f_m^-(\lambda)/f_{m-1}^+(\lambda)$  are more complicated. They are given by explicit but tedious formulae, with one salient feature however: they are rational functions of the parameters of the Markov chain.

Relation (4.16) can be viewed as a random time analog (or dual) of the usual finite time out-of-equilibrium relations. The usual question would be to ask whether there is a relation

between the probabilities to see a winding number  $w$  versus  $-w$  at a fixed time  $t$ : time is fixed,  $\tilde{W}_t$  (as defined in Sect. 4.1.1) is random and one compares in some way  $p(\tilde{W}_t = w)$  and  $p(\tilde{W}_t = -w)$ . But here,  $T^+$  is the time to reach winding number 1,  $T^-$  is the time to reach winding number  $-1$  and we compare the laws of  $T^+$  and  $T^-$ .

To make it look even more familiar, we may define  $T^{(w)}$  for an arbitrary integer  $w$  as the time needed to reach winding number  $w$ , and  $F^{(w)}(\lambda)$  as the Laplace transform of the law of  $T^{(w)}$ , so that  $F^\pm = F^{(\pm 1)}$ . The renewal property entails that  $F^{(w)}$  is simply the  $w^{\text{th}}$  (resp.  $(-w)^{\text{th}}$ ) power of  $F^+$  (resp.  $F^-$ ) for  $w \geq 0$  (resp.  $w \leq 0$ ). Then, for each integer  $w$

$$\frac{F^{(-w)}(\lambda)}{F^{(w)}(\lambda)} = e^{-wA}.$$

Note that this is an exact relation valid for finite winding numbers and for any cycle in a more general system of transitions as long as the procedure to define the process of the cycle preserves the affinity class, as shown for instance for “drag and drop” and conditioning. However, the relation is established under the assumption that the first waiting time plays no special role. This is natural for conditioning but not for “drag and drop” for instance. Then letting  $\hat{F}^{(w)}(\lambda)$  denote the “real” quantity (vaguely speaking, taking into account the time it takes to reach the cycle) the law of large numbers leads to the fact that  $[\hat{F}^{(w)}(\lambda)]^{1/w}$  behaves for large  $w \rightarrow +\infty$  (resp.  $w \rightarrow -\infty$ ) as  $F^+(\lambda)$  (resp.  $F^-(\lambda)$ ) so that relation (4.16) is observable as an asymptotic result in general:

$$\lim_{w \rightarrow \pm\infty} \frac{[\hat{F}^{(-w)}(\lambda)]^{1/w}}{[\hat{F}^{(w)}(\lambda)]^{1/w}} = e^{-A}.$$

*Remark* There is another interpretation of the identity  $F^-(\lambda) = F^+(\lambda)e^{-A}$ . Let  $P^+$  (resp.  $P^-$ ) be the law on  $[0, +\infty]$  of the time  $T^+$  (resp.  $T^-$ ) it takes to reach winding number 1 (resp.  $-1$ ), so that  $F^+(\lambda)$  (resp.  $F^-(\lambda)$ ) is nothing but the Laplace transform of  $P^+$  (resp.  $P^-$ ). Then  $P^+$  is absolutely continuous with respect to  $P^-$  and

$$P^- = e^{-A} P^+ + (1 - e^{-A})\delta_{+\infty},$$

i.e. the Radon-Nykodim derivative  $dP^+/dP^-$  is equal to the constant  $e^A$ .

#### 4.1.6 Mean First Passage Time and Efficiency

The previous arguments have lead to a general symmetry relation between clockwise and anti-clockwise first passage times, but as stated above the computation of the law of the individual clockwise or anticlockwise first passage times is complicated.

We illustrate this first level of complexity by giving a formula for the mean first passage time. This is the most straightforward measure of the efficiency of the cycle, i.e. the average speed at which cycles are performed by the noria.

We assume here that  $A > 0$  so that  $\Pi^+ = 1$ ,  $\Pi^- = e^{-A} < 1$  and the winding number grows in average. This suggests strongly that the mean first passage times at positive winding numbers are finite. We shall not give a proof of this (this is standard and not really difficult, but would take us too far) but we give the formula.

We let  $l_m$  denote the mean waiting time at  $m$  for  $m = 1, \dots, M$  and  $\tau_m^+$  be the mean first passage time at winding  $1/M$  starting from  $m$ , so that, by the strong Markov property, the mean first passage time at winding 1 is, whatever the starting point,  $\sum_{m=1}^M \tau_m^+ \equiv \langle T^+ \rangle$ . For an arbitrary generalized renewal process, the finiteness of  $l_m$  is not guaranteed and should be

included as an hypothesis. For the constructions we use in this work (conditioning or drag and drop on a subgraph, starting from a finite state Markov process), it is automatic.

**Claim** *The equation*

$$\tau_m^+ = p_m^+ l_m + p_m^- (l_m + \tau_{m-1}^+ + \tau_m^+) \quad (4.17)$$

holds.

*Proof* We give two arguments:

- The first one is related to the generating functions in the parameter  $\lambda$  introduced in 4.1.5, because average times are first moments, so they can be computed by taking a derivative with respect to  $\lambda$  at  $\lambda = 0$ . Thus  $l_m$  (resp.  $\tau_m^+$ ) is nothing but  $-\frac{dx_m(\lambda)}{d\lambda}(\lambda = 0)$  (resp.  $-\frac{df_m^+(\lambda)}{d\lambda}(\lambda = 0)$ ), so (4.17) can be obtained by taking the derivative of (4.2) with respect to  $\lambda$  at  $\lambda = 0$ . Notice that  $x_m(\lambda = 0) = 1$  is always true, but  $f_m^+(\lambda = 0) = 1$  holds only because  $\Pi^+ = 1$ .
- The second argument, a direct probabilistic derivation relying on the strong Markov property, is even more transparent: it takes in average  $l_m$  units of time to make the first step; if it is clockwise, (probability =  $p_m^+$ ), we are done, but if the first step is anti-clockwise (probability =  $p_m^-$ ) reaching winding number  $1/M$  takes in average, on top of  $l_m$ , time  $\tau_{m-1}^+$  to hit winding number 0 and time  $\tau_m^+$  to hit winding number  $1/M$ .  $\square$

**Claim** *The mean first passage time at winding number 1,  $\langle T^+ \rangle$ , is given by*

$$\left( \prod_{m=1}^M p_m^+ - \prod_{m=1}^M p_m^- \right) \langle T^+ \rangle = \sum_{m=1}^M \sum_{k=1}^M \left( \prod_{1 \leq i < k} p_{m+i}^+ \right) l_{m+k} \left( \prod_{k < j \leq M} p_{m+j}^- \right). \quad (4.18)$$

*This formula is still elegant, but already complicated. The other moments of the first passage time  $T^+$  can be computed by analogous tricks but require more efforts, and the final expressions require more space.*

*Proof* Equation (4.17) can be rewritten:

$$\tau_m^+ = \frac{l_m}{p_m^+} + \frac{p_m^-}{p_m^+} \tau_{m-1}^+.$$

Using this relation repeatedly, we get  $\tau_M^+$  in terms of  $\tau_{M-1}^+$ , then in terms of  $\tau_{M-2}^+$  and so on, until we finally get a closed equation because  $\tau_0^+ = \tau_M^+$ . This leads to:

$$\left( \prod_{m=1}^M p_m^+ - \prod_{m=1}^M p_m^- \right) \tau_M^+ = \sum_{k=1}^M \left( \prod_{1 \leq i < k} p_i^+ \right) l_k \left( \prod_{k < j \leq n} p_j^- \right).$$

Note that  $\prod_{m=1}^M p_m^+ - \prod_{m=1}^M p_m^- > 0$ , which is equivalent to  $A > O$ , is our working hypothesis, leading to growth of the winding number. Under the opposite hypothesis, the formula would be inconsistent as it would lead to negative mean first passage times. The other  $\tau_m^+$ 's are easily obtained by a cyclic permutation, finally leading to (4.18).  $\square$

The mean first passage time at winding number  $w = 0, 1, 2, \dots$  is  $w\langle T^+ \rangle$ . The strong law of large numbers will imply that the first passage time at winding number  $w$ , which is a sum of  $w$  independent random variables with mean  $\langle T^+ \rangle$  will differ from its mean  $w\langle T^+ \rangle$

by an  $o(w)$  with probability going to 1 as  $w \rightarrow +\infty$ . By a standard trick this results can be expressed in an equivalent way as:

$$\lim_{t \rightarrow +\infty} \tilde{W}_t/t = 1/\langle T^+ \rangle \text{ with probability 1.} \quad (4.19)$$

So when the noria works for a long time fluctuations are suppressed and a “classical” behavior emerges : the noria turns at speed  $1/\langle T^+ \rangle + o(1)$  and can thus be fully characterized by its efficiency.

Formula (4.19) can be checked against a particular case: if we assume that the cycle is described by a Markov process, i.e. that the waiting times are exponential, we may define the current  $\tilde{j}_t = \tilde{j}(\mathcal{C}_t)$  associated to the exchange process  $\tilde{W}_t$ . The generic formula (2.1) specializes to  $\tilde{j}(m) = \frac{1}{M} l_m^{-1} (p_m^+ - p_m^-)$ . The average  $\langle \tilde{j}_t \rangle_{\text{st}}$  of  $\tilde{j}_t$  in the stationary state is thus  $\langle \tilde{j}_t \rangle_{\text{st}} = \frac{1}{M} \sum_m l_m^{-1} (p_m^+ - p_m^-) P_{\text{st}}(m)$ . The equation for  $P_{\text{st}}(m)$  is

$$l_m^{-1} P_{\text{st}}(m) = l_{m+1}^{-1} p_{m+1}^- P_{\text{st}}(m+1) + l_{m-1}^{-1} p_{m-1}^+ P_{\text{st}}(m-1),$$

which by a simple rearrangement (using  $1 = p_m^+ + p_m^-$  on the left-hand side) can be rewritten:

$$l_m^{-1} p_m^+ P_{\text{st}}(m) - l_{m+1}^{-1} p_{m+1}^- P_{\text{st}}(m+1) = l_{m-1}^{-1} p_{m-1}^+ P_{\text{st}}(m-1) - l_m^{-1} p_m^- P_{\text{st}}(m).$$

Thus the quantity  $l_m^{-1} p_m^+ P_{\text{st}}(m) - l_{m+1}^{-1} p_{m+1}^- P_{\text{st}}(m+1)$  is  $m$ -independent, i.e. constant along the cycle, and we call it  $\Delta^+$ . Note that summing over  $m$  one finds that  $\Delta^+ = \langle \tilde{j}_t \rangle_{\text{st}}$ . Moreover, treating  $\Delta^+$  as a parameter, the one term recursion relation can be solved as before: one iterates until a closed equation is obtained because  $P_{\text{st}}(0) = P_{\text{st}}(M)$ . Then  $\Delta^+$  is obtained by the normalization condition  $\sum_m P_{\text{st}}(m) = 1$ . One finds without surprise that the formula for  $\langle \tilde{j}_t \rangle_{\text{st}}$  is exactly the formula obtained in (4.18) for  $1/\langle T^+ \rangle$ , so that  $\langle \tilde{j}_t \rangle_{\text{st}} = 1/\langle T^+ \rangle$  in agreement with the more general result  $\lim_{t \rightarrow +\infty} \tilde{W}_t/t = 1/\langle T^+ \rangle$  obtained above because  $\langle \tilde{W}_t \rangle \sim t \langle \tilde{j}_t \rangle_{\text{st}}$  for large  $t$ .

#### 4.1.7 Illustration in the Degenerate Case $M = 1$

Though real cycles have  $M \geq 3$  vertices, there is nothing that prevents defining the winding number process for  $M = 1$ , so that all indices can be suppressed. At each jump the winding number grows by  $\pm 1$  with probability  $p^\pm$ . So the corresponding Markov chain is nothing but the simple asymmetric random walk, and the renewal process is obtained by using a sequence of independent identically distributed waiting times separating the jumps. Assume for definiteness that  $p^+ \geq p^- = 1 - p^+$  so that  $A = \ln(p^+/p^-) \geq 0$ . Trivially

$$\mathbb{F} = \begin{pmatrix} 1/(xp^+) - p^-/p^+ \\ 1 \\ 0 \end{pmatrix}.$$

Applying the general formulæ obtained above, one recovers very classical results:

- The probability to reach winding number  $w$  is 1 for  $w = 0, 1, 2, \dots$  while it is  $e^{wA}$  for  $w = 0, -1, -2, \dots$ .
- The generating functions for the number of steps it takes to reach winding number  $w = \pm 1$  are  $F^\pm(z)$  given by

$$F^+(z) = \frac{1 - \sqrt{1 - 4z^2 p^+ p^-}}{2zp^-}, \quad F^-(z) = \frac{1 - \sqrt{1 - 4z^2 p^+ p^-}}{2zp^+},$$

where the square root is the standard branch ( $\sqrt{1} = 1$ ), and the small  $z$  expansion

$$\frac{1 - \sqrt{1 - 4z^2 p^+ p^-}}{2zp^-} = \sum_{k \geq 0} (p^+)^{k+1} (p^-)^k \frac{(2k)!}{k!(k+1)!} z^{2k+1}$$

allows to retrieve the familiar fact that the number of walks that start from 0 and reach 1 for the first time after  $2k + 1$  steps is given by the Catalan number  $\frac{(2k)!}{k!(k+1)!}$ , each of these walks having weight  $(p^+)^{k+1} (p^-)^k$ .

- If  $x(\lambda)$  is the Laplace transform of the distribution of the waiting time between two jumps, the Laplace transform of the distribution of the first passage time at  $\pm 1$  is  $F^\pm(\lambda)$ , where

$$F^+(\lambda) = \frac{1 - \sqrt{1 - 4x(\lambda)^2 p^+ p^-}}{2x(\lambda)p^-}, \quad F^-(\lambda) = \frac{1 - \sqrt{1 - 4x(\lambda)^2 p^+ p^-}}{2x(\lambda)p^+},$$

- The average of the first passage time at  $-1$  is infinite, but the average first passage time at  $1$  is

$$\langle T^+ \rangle = \frac{l}{p^+ - p^-},$$

where  $l$  is the average time between two jumps, and  $\tilde{W}_t \sim \frac{p^+ - p^-}{l} t$  at large  $t$ .

## 5 Conclusions

In this article, we have proved a new general fluctuation relation dual to the ones usually exhibited in the literature.

The usual question is to ask whether there is a relation between the probabilities that an observable  $\tilde{W}_t$  (here, the winding number) takes a certain value  $w$  versus  $-w$  at a fixed time  $t$ : time is fixed,  $\tilde{W}_t$  is random and one compares in some way  $p(\tilde{W}_t = w)$  and  $p(\tilde{W}_t = -w)$ . In our approach, we have defined  $T^{(w)}$  as the random time it takes for the observable to reach a certain value  $w$  and we compared the laws of  $T^{(w)}$  and  $T^{(-w)}$  and derived a corresponding non-equilibrium fluctuation relation. Technically, the main probabilistic tool is the strong Markov property. It leads to recursion relations among the quantities of interest, typically generating functions. Then periodicity along the cycle is used to solve the recursion relations.

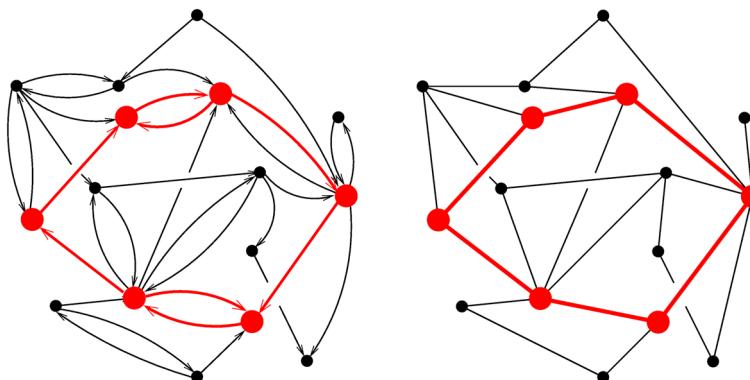
Though this fluctuation relation is established for special systems with the geometry of a cycle, we have argued that it can be observed in any Markovian system because the quantity involved in the fluctuation relation is the (cycle) affinity, which we have shown to be invariant under many probabilistic contractions, including conditioning but also “drag and drop”, a new procedure we have introduced.

One obvious useful generalization would be the study of correlations of random times between several cycles.

## Appendix 1: A Short Reminder on Graph Theory

In this section, which is meant to be self-contained (see the first chapters of e.g. [15–18] for much more material), we recall basic definitions, sometimes adapted to our needs, from elementary graph theory.

A graph  $\mathbf{G}$  is a couple  $(\mathbf{V}, \mathbf{E})$  where  $\mathbf{V}$  is an arbitrary non-empty set and  $\mathbf{E}$  is an arbitrary subset of  $\mathbf{V} \times \mathbf{V}$  disjoint from the diagonal  $\{(v, v), v \in \mathbf{V}\}$ . Elements of  $\mathbf{V}$  are called vertices



**Fig. 1** Two examples of subgraphs within a graph. *Left* a general graph with a subgraph. *Right* a non-oriented graph with a subgraph. The subgraph is *marked* with larger vertices and thicker edges. The oriented graph on the *left* is not connected, but the subgraph is

and elements of  $\mathbf{E}$  are called edges: if  $(v, v') \in \mathbf{E}$  (which implies that  $v$  and  $v'$  are distinct) we say that there is an edge from  $v$  to  $v'$ , or that  $v'$  is adjacent to  $v$ . In pictures, the vertices are represented as points, and the edges as arrows joining vertices. For our purposes, the set of vertices will always be a finite set.

A non-oriented graph  $\mathbf{G}$  is a couple  $(\mathbf{V}, \mathbf{E})$  where  $\mathbf{V}$  is an arbitrary non-empty set and  $\mathbf{E}$  is a subset of the set, sometimes denoted  $S^2(\mathbf{V})$ , of pairs of elements of  $\mathbf{V}$ , i.e. the set  $\{(v, v'), v, v' \in \mathbf{V}, v \neq v'\}$ . In that case, edges are represented as line segments because in  $\{v, v'\}$ ,  $v$  and  $v'$  play symmetric roles.

Graphs such that  $\mathbf{E}$  is symmetric, i.e. such that  $(v, v')$  is an edge if and only if  $(v', v)$  is an edge, are said to be *micro-reversible*. This is not standard terminology, but is motivated by statistical mechanics considerations. Micro-reversible graphs are in one-to-one correspondence with non-oriented graphs. In pictures, a line segment between two vertices representing the edge  $\{v, v'\}$  corresponds to a pair of arrows, one from  $v$  to  $v'$  for  $(v, v')$  and one from  $v'$  to  $v$  for  $(v', v)$ .

A graph  $\mathbf{H}$  is a subgraph of a graph  $\mathbf{G}$  if the set of vertices of  $\mathbf{H}$  is a subset of the set of vertices of  $\mathbf{G}$ , and the set of edges of  $\mathbf{H}$  is a subset of the set of edges of  $\mathbf{G}$  joining vertices of  $\mathbf{H}$ .

A walk on a graph  $\mathbf{G}$  is a sequence  $X_0, \dots, X_N$  of vertices of  $\mathbf{G}$  such that  $(X_{i-1}, X_i)$  is an edge of  $\mathbf{G}$  for  $i = 1, \dots, N$ . The number  $N$  is the length of the walk, and single vertices count as walk of length 0.

We say that a graph is connected if there is a walk joining any two vertices. Caution: this is not the standard terminology, one of the reasons being that micro-reversible graphs  $\mathbf{G}$  can be split in a unique way into disjoint connected components, but this is not true for general graphs, see Fig. 1 for illustration.

## Appendix 2: A Short Reminder on Finite State Markov Chains

A reference dedicated to (finite or countable state space) Markov chains and processes, mostly self-contained and at an accessible level of sophistication, is [19]. Markov chains and processes are also covered in a number of general textbooks. We have found, [20–24], (at

increasing level of complexity and abstraction) well-adapted to our tastes. But this appendix should be easily readable with a very modest background in finite probability spaces, and the above references are needed only for those who want to learn more. The only deep point, namely that the law for infinite trajectories can be constructed as a limit of laws for finite trajectories (a special case of Kolmogorov's theorem, see the references above), can be admitted.

The starting point is a finite set  $\mathbf{C}$  of configurations or states, and a stochastic matrix  $\mathbb{P}$  on  $\mathbf{C}$ .

A stochastic matrix on  $\mathbf{C}$  is a collection  $(\mathcal{C}'|\mathbb{P}|\mathcal{C})_{\mathcal{C}', \mathcal{C} \in \mathbf{C}}$  such that  $(\mathcal{C}'|\mathbb{P}|\mathcal{C}) \geq 0$  for  $\mathcal{C}, \mathcal{C}' \in \mathbf{C}$ , and  $\sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{P}|\mathcal{C}) = 1$  for  $\mathcal{C} \in \mathbf{C}$ . So each column of  $\mathbb{P}$  is a probability measure on  $\mathbf{C}$ .

To these data, we associate a Markov chain on  $\mathbf{C}$  as follows: we give an initial probability distribution  $v$  on  $\mathbf{C}$  at time  $n = 0$ , and if at time  $n$  the system is in configuration  $\mathcal{C}$  then (whatever happened before  $n$ ) at time  $n + 1$  the system will be in configuration  $\mathcal{C}'$  with probability  $(\mathcal{C}'|\mathbb{P}|\mathcal{C})$ . One can show (using Kolmogorov's theorem for instance, see references above) that this defines uniquely a probability measure on sequences  $(\mathcal{C}, \mathcal{C}', \mathcal{C}'', \dots)$  of elements of  $\mathbf{C}$ . The Markovian character is transparent in this description: to know the fate of the system at time  $n + 1$ , only the knowledge of the system at time  $n$  is needed.

It is convenient to have a compact notation and we sometimes write “Consider a Markov chain  $(\mathbf{C}, \mathbb{P})$ ” or if the initial probability distribution is important “Consider a Markov chain  $(\mathbf{C}, \mathbb{P}, v)$ ”. We also sometimes write  $\mathcal{C}_n$  ( $n = 0, 1, \dots$ ) for the configuration of the chain at time  $n$ , i.e. after  $n$  steps.

## Appendix 2.1: Description of Trajectories

There are (at least) two natural descriptions of a trajectory:

- A first one is: the system is at  $\mathcal{C}$  at time 0, at  $\mathcal{C}'$  at time 1, at  $\mathcal{C}''$  at time 2, and so on. The corresponding probabilities are easy to compute. For instance the probability to be at  $\mathcal{C}$  at time 0, at  $\mathcal{C}'$  at time 1, and at  $\mathcal{C}''$  at time 2 is

$$(\mathcal{C}''|\mathbb{P}|\mathcal{C}')(\mathcal{C}'|\mathbb{P}|\mathcal{C})(\mathcal{C}|\mu)$$

where  $(\mathcal{C}|\mu)$  is the initial probability distribution.

- A second one is: the system is at  $\mathcal{C}$  at time 0 and stays at  $\mathcal{C}$  until time  $n$ , but jumps between time  $n$  and time  $n + 1$  to  $\mathcal{C}' \neq \mathcal{C}$  and stays at  $\mathcal{C}'$  up to time  $n + n' + 1$ , but jumps between time  $n + n' + 1$  and time  $n + n' + 2$  to  $\mathcal{C}'' \neq \mathcal{C}'$  and so on. The corresponding probabilities are again easy to compute. For instance the probability up to the second jump is

$$(\mathcal{C}''|\mathbb{P}|\mathcal{C}')(\mathcal{C}'|\mathbb{P}|\mathcal{C})^{n'}(\mathcal{C}'|\mathbb{P}|\mathcal{C})(\mathcal{C}|\mathbb{P}|\mathcal{C})^n(\mathcal{C}|\mu).$$

The second description is related to the following construction. Define a new stochastic matrix  $\check{\mathbb{P}}$  by the formulæ:

- If  $(\mathcal{C}|\mathbb{P}|\mathcal{C}) = 1$  then  $(\mathcal{C}'|\check{\mathbb{P}}|\mathcal{C}) = (\mathcal{C}'|\mathbb{P}|\mathcal{C})$  for every  $\mathcal{C}'$  (i.e.  $(\mathcal{C}|\check{\mathbb{P}}|\mathcal{C}) = 1$  and  $(\mathcal{C}'|\check{\mathbb{P}}|\mathcal{C}) = 0$  for  $\mathcal{C}' \neq \mathcal{C}$ ).
- If  $(\mathcal{C}|\mathbb{P}|\mathcal{C}) < 1$  then  $(\mathcal{C}|\check{\mathbb{P}}|\mathcal{C}) = 0$  and  $(\mathcal{C}'|\check{\mathbb{P}}|\mathcal{C}) = \frac{(\mathcal{C}'|\mathbb{P}|\mathcal{C})}{1 - (\mathcal{C}|\mathbb{P}|\mathcal{C})}$  for  $\mathcal{C}' \neq \mathcal{C}$ .

Also let  $\mathbb{K}$  be the diagonal matrix whose entries are those of the diagonal of  $\mathbb{P}$ , i.e.  $(\mathcal{C}'|\mathbb{K}|\mathcal{C}) \equiv \delta_{\mathcal{C}, \mathcal{C}'}(\mathcal{C}|\mathbb{P}|\mathcal{C})$ . Then

$$\check{\mathbb{P}}(\mathbb{I} - \mathbb{K}) + \mathbb{K} = \mathbb{P}.$$

Say that a random variable  $N$  obeys a geometric distribution of parameter  $k \in [0, 1]$  if  $p(N \geq n) = k^n$  for  $n = 0, 1, \dots$ , i.e.  $p(N = n) = k^n(1 - k)$ . Then the probabilistic interpretation of the second description is the following. Starting at  $\mathcal{C}$  at time 0 if  $(\mathcal{C}|\mathbb{K}|\mathcal{C}) = 1$  stay there forever, otherwise stay at  $\mathcal{C}$  during a geometric time  $N$  with parameter  $(\mathcal{C}|\mathbb{K}|\mathcal{C})$  and between  $N$  and  $N + 1$  jump to  $\mathcal{C}'$  with probability  $(\mathcal{C}'|\mathbb{P}|\mathcal{C})$ ; if  $(\mathcal{C}'|\mathbb{K}|\mathcal{C}') = 1$  stay there forever, otherwise stay at  $\mathcal{C}'$  during a geometric time  $N'$  with parameter  $(\mathcal{C}'|\mathbb{K}|\mathcal{C}')$  and between  $N + N' + 1$  and  $N + N' + 2$  jump to  $\mathcal{C}''$  with probability  $(\mathcal{C}''|\mathbb{P}|\mathcal{C}')$ ; and so on.

Note that certain (infinite) trajectories for the Markov chain associated to  $\mathbb{P}$  may lead to a finite number of jumps. However, if  $(\mathcal{C}|\mathbb{K}|\mathcal{C}) = 1$  then also  $(\mathcal{C}|\check{\mathbb{P}}|\mathcal{C}) = 1$  so staying forever at  $\mathcal{C}$  is also what is predicted by  $\check{\mathbb{P}}$ , and if this is done, any infinite trajectory for the Markov chain associated to  $\mathbb{P}$  leads to an infinite trajectory for the Markov chain associated to  $\check{\mathbb{P}}$ .

## Appendix 2.2: The Graph Associated to a Markov Chain

To any stochastic matrix  $\mathbb{P}$  we can associate a graph. The vertices are the configurations in  $\mathbf{C}$ , and (oriented) edges are indexed by the possible transitions: there is an oriented edge from  $\mathcal{C}$  to  $\mathcal{C}'$  if and only if  $\mathcal{C} \neq \mathcal{C}'$  and  $(\mathcal{C}'|\mathbb{P}|\mathcal{C}) \neq 0$ . In that case, we talk of the edge  $(\mathcal{C}, \mathcal{C}')$  and we say that  $\mathcal{C}'$  is adjacent to  $\mathcal{C}$  (not a symmetric relation in general). We shall denote the corresponding graph by  $\mathbf{G}$ : as recalled above,  $\mathbf{G}$  is a couple of sets: the set of vertices,  $\mathbf{C}$ , and the set of edges, a subset of  $\mathbf{C} \times \mathbf{C}$  disjoint from the diagonal.

Recall that a graph  $\mathbf{G}$  is said to be connected if one can go from any vertex to any other by a sequence of adjacent vertices. In that case, general theorems state that:

- All configurations are recurrent with probability 1, i.e. for almost every trajectory, every configuration appears infinitely many times.
- There is a single probability measure  $P_{\text{st}}$  on  $\mathbf{C}$ , called the stationary measure, such that  $\mathbb{P}P_{\text{st}} = P_{\text{st}}$  i.e.  $\sum_{\mathcal{C}} (\mathcal{C}'|\mathbb{P}|\mathcal{C}) P_{\text{st}}(\mathcal{C}) = P_{\text{st}}(\mathcal{C}')$  for each  $\mathcal{C}'$ . Moreover,  $P_{\text{st}}(\mathcal{C}) > 0$  for each  $\mathcal{C} \in \mathbf{C}$ .

However, due to some possible arithmetic symmetries, it is not always true that any initial probability distribution on  $\mathbf{C}$  converges at large times to  $P_{\text{st}}$ .

Motivated by the graphical interpretation, we shall repeatedly use the image that a trajectory is a random walker on  $\mathbf{C}$  jumping from time to time along edges from one vertex to another. Note that  $\mathbb{P}$  and the associated  $\check{\mathbb{P}}$  correspond to the same graph.

We shall often make the assumption of micro-reversibility, i.e.  $(\mathcal{C}'|\mathbb{P}|\mathcal{C})$  and  $(\mathcal{C}|\mathbb{P}|\mathcal{C}')$  are simultaneously either  $= 0$  or  $\neq 0$ . Then, the above (oriented) graph carries the same information as a non-oriented one, for which we keep the same name, and we shall say that  $\{\mathcal{C}, \mathcal{C}'\}$  (unordered) is an edge of  $\mathbf{G}$ .

## Appendix 3: A Short Reminder on Finite State Markov Processes

This appendix should be easily readable by anyone with minimal familiarity with the master equation approach. Our discussion closely parallels the description of Markov chains. The references given at the beginning of Appendix 2 explain what is stated here and much more.

The starting point is a finite set  $\mathbf{C}$  of configurations or states, and a transition matrix  $\mathbb{W}$  with vanishing diagonal matrix elements and whose non-diagonal matrix elements  $(\mathcal{C}'|\mathbb{W}|\mathcal{C}) \geq 0$ ,  $\mathcal{C}, \mathcal{C}' \in \mathbf{C}$ ,  $\mathcal{C} \neq \mathcal{C}'$ , describe transition rates from  $\mathcal{C}$  to  $\mathcal{C}'$ . To  $\mathbb{W}$  is associated a Markov matrix  $\mathbb{M}$  defined by  $(\mathcal{C}'|\mathbb{M}|\mathcal{C}) \equiv (\mathcal{C}'|\mathbb{W}|\mathcal{C})$  for  $\mathcal{C}, \mathcal{C}' \in \mathbf{C}$ ,  $\mathcal{C} \neq \mathcal{C}'$ , and  $(\mathcal{C}|\mathbb{M}|\mathcal{C}) \equiv -\sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}'|\mathbb{W}|\mathcal{C})$  for  $\mathcal{C} \in \mathbf{C}$ . We let  $\mathbb{D}$  denote the (diagonal) matrix such that  $\mathbb{M} = \mathbb{W} - \mathbb{D}$ .

To these data, we associate a Markov process on  $\mathbf{C}$  as follows: we give an initial probability distribution  $v$  on  $\mathbf{C}$  at time  $t = 0$ , and if at time  $t$  the system is in configuration  $\mathcal{C}$  then (whatever happened before  $t$ ) at time  $t + \Delta t$  the system is still in  $\mathcal{C}$  with probability  $1 + (\mathcal{C}|\mathbb{M}|\mathcal{C})\Delta t + o(\Delta t)$  i.e.  $1 - (\mathcal{C}|\mathbb{D}|\mathcal{C})\Delta t + o(\Delta t)$ , and is in configuration  $\mathcal{C}' \neq \mathcal{C}$  with probability  $(\mathcal{C}'|\mathbb{M}|\mathcal{C})\Delta t + o(\Delta t)$  i.e.  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})\Delta t + o(\Delta t)$ . Taking an infinitesimal  $dt$  instead of a finite but small  $\Delta t$ , we see that  $\mathbb{M}$  appears as the infinitesimal generator of the Markov process. The Markovian character is transparent in this description: to know the fate of the system at time  $t + dt$ , only the knowledge of the system at time  $t$  is needed. The mathematical construction of the law of a Markov process from these data is only slightly more complicated than for Markov chains.

### Appendix 3.1: Relation with the Master Equation

We write  $\mathcal{C}_t$  ( $t \in [0, +\infty[$ ) for the configuration of the process at time  $t$ . Thus a trajectory is a (random) function from  $[0, +\infty[$  to  $\mathbf{C}$ . The mathematical construction of the Markov process amounts to the construction of a probability measure, which we denote by  $P$  in this article, consistent with the heuristic description of transitions given above, on an appropriate set of functions from  $[0, +\infty[$  to  $\mathbf{C}$ . One shows that it is possible to concentrate on functions which are right-continuous with left limits: if there is a jump at time  $t$ ,  $\mathcal{C}_t$  is the position after the jump so  $\mathcal{C}_t \neq \mathcal{C}_{t-}$  if and only if there is a jump at time  $t$ . Moreover one can concentrate on functions that have only a finite number of jumps during any bounded time interval. So if  $\mathbb{Q}$  is any matrix indexed by  $\mathbf{C}$  and with vanishing diagonal elements, ( $\mathbb{W}$  is a typical example), the sum  $\sum_{t \in ]0, T]} (\mathcal{C}_t | \mathbb{Q} | \mathcal{C}_{t-})$ , to which only jump times contribute, is well-defined. This simple construction plays an important role in the construction of cumulative processes in sect. 2.1.

It is convenient to have a compact notation and we sometimes write “Consider a Markov process  $(\mathbf{C}, \mathbb{W})$ ” or if the initial probability distribution is important “Consider a Markov process  $(\mathbf{C}, \mathbb{W}, v)$ ”.

The probability  $P(\mathcal{C}; t)$  to be in configuration  $\mathcal{C}$  at time  $t$ , a shorthand for  $P(\mathcal{C}_t = \mathcal{C})$ , is obtained by solving the so-called master equation

$$\frac{d}{dt} P(\mathcal{C}; t) = \sum_{\mathcal{C}'} (\mathcal{C}|\mathbb{M}|\mathcal{C}') P(\mathcal{C}'; t).$$

with initial condition  $P(\mathcal{C}; 0)$ . This can also be written as

$$\frac{d}{dt} P(\mathcal{C}; t) = \sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}|\mathbb{W}|\mathcal{C}') P(\mathcal{C}'; t) - \sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) P(\mathcal{C}; t) \equiv \sum_{\mathcal{C}'} \dot{P}_{\mathcal{C}; \mathcal{C}'}(t)$$

which has a simple interpretation:  $P(\mathcal{C}; t)$  varies with time because of positive contribution due to jumps to  $\mathcal{C}$  and of negative contributions due to jumps from  $\mathcal{C}$ , and for the infinitesimal time balance this leads to the above formula. By construction  $\dot{P}_{\mathcal{C}; \mathcal{C}'}(t) \equiv (\mathcal{C}|\mathbb{W}|\mathcal{C}') P(\mathcal{C}'; t) - (\mathcal{C}'|\mathbb{W}|\mathcal{C}) P(\mathcal{C}; t)$ , which contains the contributions of transition from  $\mathcal{C}'$  to  $\mathcal{C}$  to the variation of  $P(\mathcal{C}, t)$ , is anti-symmetric.

### Appendix 3.2: The Graph Associated to a Markov Process

To any Markov matrix  $\mathbb{M}$  we can associate a graph. The vertices are the configurations in  $\mathbf{C}$ , and oriented edges are indexed by the possible transitions: there is an oriented edge from  $\mathcal{C}$  to  $\mathcal{C}'$  if and only if  $(\mathcal{C}'|\mathbb{W}|\mathcal{C}) \neq 0$ . In that case, we talk of the edge  $(\mathcal{C}, \mathcal{C}')$  and we say that  $\mathcal{C}'$  is adjacent to  $\mathcal{C}$  (not a symmetric relation in general). We shall denote the corresponding graph

by  $\mathbf{G}$ . As usual  $\mathbf{G}$  is a couple of sets: the set of vertices,  $\mathbf{C}$ , and the set of edges, a subset of  $\mathbf{C} \times \mathbf{C}$ .

Remember the graph  $\mathbf{G}$  is said to be connected if one can go from any vertex to any other by a sequence of adjacent vertices. In that case, a general theorem guarantees that there is a single probability measure  $P_{\text{st}}$  on  $\mathbf{C}$ , called the stationary measure, such that  $\mathbb{M}P_{\text{st}} = 0$  i.e.  $\sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{M}|\mathcal{C}) P_{\text{st}}(\mathcal{C}') = 0$  for each  $\mathcal{C}$ . Moreover,  $P_{\text{st}}(\mathcal{C}) > 0$  for each  $\mathcal{C}$  and whatever  $P(\mathcal{C}; 0)$ ,  $\lim_{t \rightarrow \infty} P(\mathcal{C}; t) = P_{\text{st}}(\mathcal{C})$ . Hence the stationary measure is unique, charges every vertex, and is the infinite time limit of every initial probability distribution.

Motivated by the graphical interpretation, we shall repeatedly use the image that a trajectory is a random walker on  $\mathbf{C}$  jumping from time to time along edges from one vertex of  $\mathbf{G}$  to an adjacent vertex.

We shall often make the assumption of micro-reversibility, i.e.  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$  and  $(\mathcal{C}|\mathbb{W}|\mathcal{C}')$  are simultaneously either  $= 0$  or  $\neq 0$ . Then, the above (oriented) graph carries the same information as a non-oriented one, for which we keep the same name.

### Appendix 3.3: Description of Trajectories

There is a more explicit description of trajectories which is useful for our purpose (and for numerical simulations of trajectories as well). It is the continuous time analog of the second description of trajectories for Markov chains, and it goes as follows. Recall that  $\mathbb{D}$  is the diagonal matrix whose entries are those of the diagonal of  $\mathbb{M}$ .

- The configuration at  $t = 0$  is sampled according to the initial probability distribution  $\nu$  on  $\mathbf{C}$ . Say the configuration at  $t = 0$  is  $\mathcal{C}$ .
- If  $(\mathcal{C}|\mathbb{D}|\mathcal{C}) = 0$  stay in  $\mathcal{C}$  forever, otherwise wait an exponential time  $T$  with parameter  $(\mathcal{C}|\mathbb{D}|\mathcal{C})$  (i.e.  $p(T > t) = e^{-(\mathcal{C}|\mathbb{D}|\mathcal{C})t}$ ) and at  $t = T$  jump to the configuration  $\mathcal{C}' \neq \mathcal{C}$  with probability  $\frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{D}|\mathcal{C})}$ .
- If  $(\mathcal{C}'|\mathbb{D}|\mathcal{C}') = 0$  stay in  $\mathcal{C}'$  forever, otherwise wait an exponential time  $T'$  with parameter  $(\mathcal{C}'|\mathbb{D}|\mathcal{C}')$  and at  $t = T + T'$  jump to configuration  $\mathcal{C}'' \neq \mathcal{C}'$  with probability  $\frac{(\mathcal{C}''|\mathbb{W}|\mathcal{C}')}{(\mathcal{C}'|\mathbb{D}|\mathcal{C}')}$ .
- ...

To the Markov matrix  $\mathbb{M}$  we can associate a stochastic matrix  $\mathbb{P}$  (i.e. the generator for a discrete time Markov chain) as follows :

- If  $(\mathcal{C}|\mathbb{M}|\mathcal{C}) = 0$  then  $(\mathcal{C}|\mathbb{P}|\mathcal{C}) = 1$  and  $(\mathcal{C}'|\mathbb{P}|\mathcal{C}) = 0$  for  $\mathcal{C}' \neq \mathcal{C}$ .
- Else,  $(\mathcal{C}'|\mathbb{P}|\mathcal{C}) = \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{D}|\mathcal{C})}$ ; in particular  $(\mathcal{C}|\mathbb{P}|\mathcal{C}) = 0$ .

Note that

$$\mathbb{M} = (\mathbb{P} - \mathbb{I})\mathbb{D}$$

and that  $\mathbb{P}$  and  $\mathbb{M}$  define the same graph. One important consequence of the above description is that the sequence  $(\mathcal{C}, \mathcal{C}', \mathcal{C}'', \dots)$  is a sample of the Markov chain associated to  $\mathbb{P}$  (with the same innocent trick as in Appendix 2 in force: if  $(\mathcal{C}, \mathcal{C}', \mathcal{C}'', \dots)$  is a finite sequence, then one turns it into an infinite one by repeating its last term over and over).

### Appendix 3.4: Perturbative Expansion

For the reader unfamiliar with the above trajectory description of the Markov process, we can offer a poor man's heuristic version, reminiscent of Feynman's sum over histories, and which is again a continuous time analog of the Markov chain case. Let  $\mathbb{U}(t) = e^{\mathbb{M}t}$  be the

evolution operator, i.e. the matrix solution of

$$\frac{d}{dt} \mathbb{U} = \mathbb{M}\mathbb{U}$$

with initial condition  $\mathbb{U}(0) = \mathbb{I}$ . Using the identity  $\mathbb{M} = (\mathbb{P} - \mathbb{I})\mathbb{D}$ , it is easy to see that these two equations can be rephrased as a single integral equation:

$$\mathbb{U}(t) = e^{-\mathbb{D}t} + \int_0^t ds e^{-\mathbb{D}(t-s)} \mathbb{P}\mathbb{D}\mathbb{U}(s).$$

We can iterate this equation, i.e. inject  $U(s) = e^{-\mathbb{D}s} + \int_0^s dr e^{-\mathbb{D}(s-r)} \mathbb{P}\mathbb{D}\mathbb{U}(r)$  in the right-hand side, and go on. This gives a series expansion

$$\mathbb{U}(t) = e^{-\mathbb{D}t} + \int_0^t ds e^{-\mathbb{D}(t-s)} \mathbb{P}\mathbb{D}e^{-\mathbb{D}s} + \int_0^t ds \int_0^s dr e^{-\mathbb{D}(t-s)} \mathbb{P}\mathbb{D}e^{-\mathbb{D}(s-r)} \mathbb{P}\mathbb{D}e^{-\mathbb{D}r} + \dots \quad (8.1)$$

The zeroth order term (where  $\mathbb{P}$  does not appear) describes trajectories that make no jump in  $[0, t]$ , the first order term (a single integral where  $\mathbb{P}$  appears once) "sums" over trajectories that make one jump in  $[0, t]$ , the jump time being  $s$ , the second term (a double integral where  $\mathbb{P}$  appears twice) "sums" over trajectories that make two jumps in  $[0, t]$ , the jump times being  $r$  and then  $s$ , and so on. This is exactly the prediction of the trajectory description in terms of jumps governed by  $\mathbb{P}$  and exponential waiting times described by  $\mathbb{D}$ . Note that, in agreement with dimensional analysis, there is one  $\mathbb{D}$  for each time integration variable: this is because if the random variable  $T$  is such that  $p(T > t) = e^{-\lambda t}$  then the density of  $T$  is  $dt \lambda e^{-\lambda t}$ .

This representation gives an easy proof of an important property of  $P(\mathcal{C}; t)$  when  $\mathbf{G}$  is connected: whatever  $P(\mathcal{C}; 0)$ ,  $P(\mathcal{C}; t) > 0$  for every  $\mathcal{C}$  and  $t > 0$ , and in particular  $P_{st}(\mathcal{C}) > 0$  for every  $\mathcal{C}$ . Indeed, each term on the right-hand side of (8.1) has non-negative matrix elements, and if one can go from  $\mathcal{C}$  to  $\mathcal{C}'$  via  $n$  jumps, the  $n^{\text{th}}$  order term (the one with  $n$  occurrences of  $\mathbb{P}$ ) has a strictly positive matrix element between  $\mathcal{C}$  and  $\mathcal{C}'$ . So when  $\mathbf{G}$  is connected, for every  $t > 0$  every matrix element of  $\mathbb{U}(t)$  is  $> 0$ .

To conclude this rapid overview, note that for any  $\varepsilon > 0$   $\mathbb{U}(\varepsilon)$  is a stochastic matrix, and that the sequence  $\mathcal{C}_{n\varepsilon}$ ,  $n = 0, 1, \dots$  is a sample of the Markov chain  $(\mathbf{C}, \mathbb{U}(\varepsilon))$ . In this way, many properties of Markov processes can be (at least heuristically) proven by proving an analog for Markov chains and letting  $\varepsilon \rightarrow 0$ .

## References

1. Prigogine, I.: Introduction to Thermodynamics of Irreversible Processes. Wiley, New York (1968)
2. Schnakenberg, J.: Network theory of microscopic and macroscopic behavior of master equation systems. Rev. Mod. Phys. **48**, 571 (1976)
3. Andrieux, D., Gaspard, P.: Fluctuation Theorem for Currents and Schnakenberg Network Theory. J. Stat. Phys. **127**, 107 (2007)
4. Jiang, D.-Q., Qian, M., Qian, M.-P.: Mathematical Theory of Nonequilibrium Steady States. No. 1833 in Lecture Notes in Mathematics. Springer, New York (2004)
5. Kalpazidou, S.-L.: Cycle representations of Markov Processes, vol 28 of Stochastic Modelling and Applied Probability. Springer, New York (2006)
6. Bauer, M., Cornu, F. In: Preparation.

7. Lebowitz, J.L., Spohn, H.: A Gallavotti-Cohen type symmetry in the large deviation functional for stochastic dynamics. *J. Stat. Phys.* **95**, 333 (1999)
8. Derrida, B.: Non-equilibrium steady states: fluctuations and large deviations of the density and of the current. *J. Stat. Mech.*, P07023 (2007).
9. Cornu, F., Bauer, M.: ‘Thermal Contact I. Symmetries Ruled by Exchange Entropy Variations. [arXiv: 1302.4538](https://arxiv.org/abs/1302.4538) [cond-mat.stat-mech] (2013).
10. Maes, C., Netocný, K., Wynants, B.: On and beyond entropy production: the case of Markov jump processes. *Markov Process. Relat. Fields* **14**(3), 445–464 (2008)
11. Muratore-Ginanneschi, P., Mejía-Monasterio, C., Peliti, L.: Heat release by controlled continuous-time Markov jump processes. *J. Stat. Phys.* **150**, 181–203 (2013)
12. Polettini, M.: Nonequilibrium thermodynamics as a gauge theory. *Europhys. Lett.* **97**, 30003 (2012)
13. Di Leonardo, R., et al.: Bacterial ratchet motors. *PNAS* **107**(21), 9541–9545 (2010)
14. Kerfoot, M-C.: Private Communication.
15. Van Lint, J.H., Wilson, R.M.: *A Course in Combinatorics*. University Press, Cambridge (1992)
16. Bollobás, B.: *Modern Graph Theory*. Graduate texts in mathematics. Springer, Heidelberg (1998)
17. Korte, B., Vygen, J.: *Combinatorial Optimization*, vol 21 of Algorithms and Combinatorics, 4th edn. Springer, New York (2008)
18. Heineman, G.T., Pollice, G., Selkow, S.M.: *Algorithms in a nutshell—a desktop quick reference*. O'Reilly, Sebastopol (2009)
19. Stroock, D.W.: *An Introduction to Markov Processes*. No. 230 in Graduate Texts in Mathematics. Springer, New York (2005)
20. Grimmett, G., Stirzaker, D.: *Probability and Random Processes*, 3rd edn. Oxford University Press, Oxford (2001)
21. Koralov, L.B., Sinai, Y.G.: *Theory of Probability and Random Processes*. Universitext, 2nd edn. Springer, New York (2007)
22. Klenke, A.: *Probability Theory. A Comprehensive Course*, Universitext. Springer, New York (2008)
23. Fristedt, B., Gray, L.: *A Modern Approach to Probability Theory. Probability and Its Applications*. Birkhäuser (1997)
24. Kallenberg, O.: *Foundations of Modern Probability. Probability and Its Applications*, 2nd edn. Springer, New York (2001)