

Unveiling the hidden structure of complex stochastic biochemical networks

Angelo Valleriani*

*Max Planck Institute of Colloids and Interfaces,
Dept. of Theory and Bio-Systems, 14424 Potsdam, Germany*

Xin Li

Department of Chemistry, Rice University, Houston, Texas 77005, USA

Anatoly B. Kolomeisky

*Department of Chemistry and Center for Theoretical Biological Physics,
Rice University, Houston, Texas 77005, USA*

(Dated: November 7, 2013)

Abstract

Complex Markov models is a widely used and powerful predictive tool to analyze stochastic biochemical processes. When, however, the network of states is unknown, it is necessary to extract information from the data to partially build the network and to give estimates about the rates. The short-time behavior of the first-passage time distributions between two states in linear chains has been shown recently to behave as a power of time with an exponent equal to the number of intermediate states. For a general Markov network system we derive here the complete Taylor expansion of the first passage time distribution in terms of absorption times. By combining algebraic methods and graph theoretical approaches it is shown that the first term of the Taylor expansion is determined by the shortest path from the initial state to the absorbing state. When this path is unique, we prove that the coefficient of the first term can be written in terms of the product of the transition rates along the path. It is argued that the application of our results to first-return times may be used to estimate the dependence of rates from external parameters in experimentally measured time distributions.

*angelo.valleriani@mpikg.mpg.de

I. INTRODUCTION

Many processes in stochastic chemical kinetics are modeled in terms of Markov chains, i.e. by chemical transitions on a network of discrete states representing the biochemical states of a complex process [1]. Prominent examples are given by enzymatic reactions leading to the movement of motor proteins [2–4] or properties of single molecule reactions on complex substrates [5, 6]. This is especially important for understanding dynamics of biological molecular motors since there are many problems associated with the determination of the time distributions until certain states have been reached by the process. These problems are closely connected with first-passage events and their distributions can be obtained in terms of absorption times. Application of this technique to analysis of kinesin mechanical steps [7, 8] and to general bimolecular chemical reactions[9] have revealed deep mechanistic details into the nature of these systems, which usually escape our attention when only mean values are considered.

Despite the tremendous advances [7, 10, 11] in the application of theoretical methods to compute first-passage times, there is a much harder problem that concerns the reconstruction of the network of states from the experimental knowledge of the time distributions. Recently, a first step towards the solution of this problem has been done when it was shown that on linear (or 1D) network of states the short time behavior of the probability density behaves like a polynomial with the exponent equal to the number of intermediate states between the initial and the absorbing state [12]. Thus, under the assumption of a linear network of states, it would be sufficient to fit the probability function of the first passage time between two experimentally accessible states to determine the number of intermediate states.

However, the assumption of a linear network of states is rather restrictive and the question arises as to whether this result remains valid also for more complex networks. It has been suggested and numerically supported that similar relations exist for general network systems [12]. In this paper, we address this problem analytically for any type of the networks by solving the forward Master equation in terms of phase-type distributions[13]. Being the exponential of the infinitesimal generator of the Markov chain, the phase type distribution allows us to explicitly write the *entire* Taylor series of the first-passage time probability density and thus to identify both the power of the first term and its coefficient, which are expressed in terms of the product of rates. It is found that the coefficients of the expansion

are associated to the walks and paths on the network of states. The first term of the expansion corresponds to the shortest path from the initial to the chosen absorbing state.

II. RESULTS

There is one important observation, which will be used throughout this manuscript, that the first-passage times can be very efficiently computed as absorption times. For the purpose of simplifying calculations, the network of states that we consider is modified in a such way that the state to be visited becomes an absorbing state of the process. While this trick does not change anything on the nature of the first-passage times it allows us to use a set of very powerful tools aimed at computing the absorption times. The first part of this section, therefore, is devoted to the derivation of the full Taylor series of the absorption time probability density with one specific absorbing state. The applications of the general result to some typical cases arising in the modeling of stochastic biochemical processes is presented in the second part.

A. First-passage time for a process starting in i and absorbing in one of several absorbing states

We consider a Markov chain $X(t)$ on the state space $\sigma \equiv \sigma_0 \cup \sigma_A$, composed of a subset $\sigma_0 = \{0, 1, \dots, n-1\}$ of transient states, and a subset $\sigma_A = \{n, n+1, \dots, N\}$ of absorbing states. The infinitesimal generator of the process is a block matrix

$$\mathbf{Q} = \begin{pmatrix} \mathbf{S}_0 & \mathbf{S}_A \\ \mathbf{0}_{a \times n} & \mathbf{0}_{a \times a} \end{pmatrix}, \quad (1)$$

where $a = N - n + 1$ is the number of absorbing states, the off-diagonal elements of the $n \times n$ matrix \mathbf{S}_0 are the transition rates between the transient states, and the $n \times a$ matrix \mathbf{S}_A contains the transition rates from the transient to the absorbing states. Therefore, the element (i, k) of the matrix \mathbf{Q} , denoted as q_{ik} , governs the infinitesimal transitions

$$\Pr\{X(t+\tau) = k \mid X(t) = i\} = q_{ik}\tau + o(\tau), \quad (2)$$

for τ infinitesimal. The diagonal elements of \mathbf{Q} are negative or zero and satisfy the usual summation property $\sum_{k \in \sigma} q_{ik} = 0$.

Hereafter we consider processes that start in the state i , i.e. $X(0) = i$ and we look for the probabilities,

$$P_{ik}(t) = \Pr\{X(t) = k \mid X(0) = i\}, \quad (3)$$

for any $i, k \in \sigma$ by solving the corresponding forward Master equation,

$$\frac{d\vec{P}}{dt} = \vec{P} \cdot \mathbf{Q}. \quad (4)$$

This equation is supplemented with the initial condition that $\vec{P}(0)$ is a row vector of length $N + 1$ defined on σ with all zeros except at the element corresponding to the state i . In terms of phase-type distributions[8, 9, 13] the general solution of the Master equation is thus given by

$$\vec{P}(t) = \vec{P}(0) \sum_{\ell=1}^{\infty} \frac{t^{\ell}}{\ell!} \mathbf{Q}^{\ell} \equiv \vec{P}(0) \exp(\mathbf{Q}t). \quad (5)$$

It can be split into the solution for the transient states σ_0 and the solution for the absorbing states σ_A due to the block form of the infinitesimal generator \mathbf{Q} .

Together with the initial condition $X(0) = i \in \sigma_0$ we consider an absorbing state $j \in \sigma_A$ (see figure 1 for an illustration). The probability $P_{ij}(t)$ to find the process in the absorbing state j at time t is given by the corresponding component of the solution vector computed in Eq. (5). By exploiting the block form of the matrix \mathbf{Q} in Eq. (5) we obtain

$$P_{ij}(t) = \vec{e}_i [\mathbf{S}_0^{-1} (\exp(\mathbf{S}_0 t) - \mathbf{1d}) \mathbf{S}_A] \vec{e}_j^{\top}, \quad (6)$$

where \vec{e}_i is a n -dimensional row vector defined on σ_0 with all zeros except at element i , whereas \vec{e}_j^{\top} is a column unit vector defined in σ_A with all zeros except at the element corresponding to state j . The matrix $\mathbf{1d}$ is a $n \times n$ unit matrix (all zeros except in the diagonal, where it has all ones).

1. Exact full Taylor series of the absorption time

The limit for $t \rightarrow \infty$ of $P_{ij}(t)$ gives the probability u_{ij} that a process starting in i is absorbed in j . By taking the limit $t \rightarrow \infty$ of the matrix in the square brackets of (6) it is easy to see that the matrix

$$\mathbf{U} = -(\mathbf{S}_0)^{-1} \mathbf{S}_A, \quad (7)$$

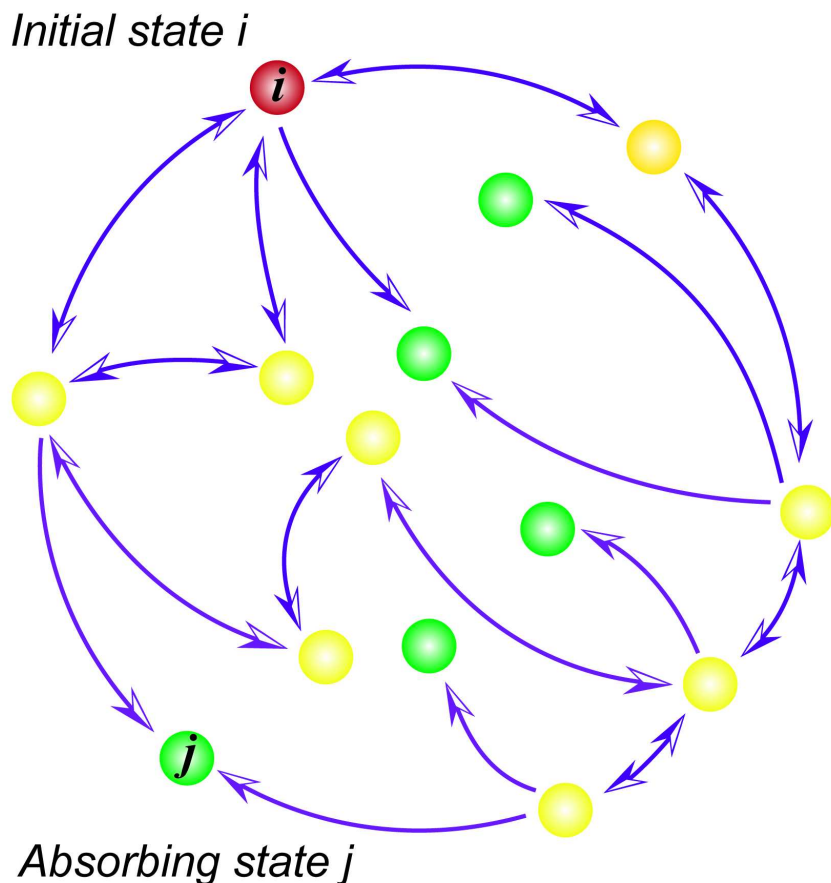


FIG. 1. A generic network with absorbing states. The green vertices represent absorbing states. The arrows pointing towards them from the transient states indicate that these transitions are irreversible. The initial state, in red, is one of the transient states, which are otherwise yellow. The transitions between the transient states are bi-directional. In this network there are several paths from the state i to the state j , but one path that has only one intermediate state is the shortest one. According to our theory, therefore, here the leading term at short time scales of the probability density function of the absorption time is linear.

contains all absorption probabilities from any transient state to any absorbing state. therefore, we obtain

$$u_{ij} = \lim_{t \rightarrow \infty} P_{ij}(t) = \vec{e}_i \mathbf{U} \vec{e}_j^T, \quad (8)$$

as the element of the matrix \mathbf{U} corresponding to the transient state i and to the absorbing state j . This indicates that the probability of absorption in j must be taken properly into account, in order to define the probability density ρ_{ij} of the random time T_{ij} that the process starting in i is conditioned to be absorbed in j . It can be shown[7] that the probability density must be normalized by the probability of absorption u_{ij} given in (8), namely

$$\rho_{ij}(t) = \frac{1}{u_{ij}} \vec{e}_i [\exp(\mathbf{S}_0 t) \mathbf{S}_A] \vec{e}_j^\top, \quad (9)$$

where we recall that \vec{e}_i is a n -dimensional row vector, \vec{e}_j^\top is a column vector of length equal to the rank of σ_A with a one corresponding to the state j , \mathbf{S}_0 is a $n \times n$ matrix, and \mathbf{S}_A is a $n \times a$ dimensional matrix. In addition, we obtain also the average absorption time from i to j as

$$E[T_{ij}] = -\frac{1}{u_{ij}} \vec{e}_i [(\mathbf{S}_0)^{-1} \mathbf{U}] \vec{e}_j^\top, \quad (10)$$

which is a simple result based only on the matrix algebra that does not require any integration.

The Taylor series of $\rho_{ij}(t)$ is given by expanding Eq. (9) as

$$\rho_{ij}(t) = \frac{1}{u_{ij}} \sum_{m=0}^{\infty} \frac{t^m}{m!} \vec{e}_i [\mathbf{S}_0^m \cdot \mathbf{S}_A] \vec{e}_j^\top, \quad (11)$$

where the explicit form of the generic terms is

$$\rho_{ij}(t) = \frac{q_{ij}}{u_{ij}} + \frac{t}{u_{ij}} \sum_{k_1=0}^{n-1} q_{ik_1} q_{k_1 j} + \dots + \frac{t^\ell}{u_{ij} \ell!} \sum_{k_1=0}^{n-1} \dots \sum_{k_\ell=0}^{n-1} q_{ik_1} q_{k_1 k_2} \dots q_{k_\ell j} + \dots \quad (12)$$

In this equation terms have a graph theoretical interpretation, which is an extremely valuable observation. It is important to stress that the expression (12) shows that each coefficient of the Taylor series can be written explicitly as a sum of the products of the transition rates.

2. Graph theoretical interpretation of the coefficients in the Taylor series

The graph \mathcal{G} associated with the Markov chain X has the states σ of the process as vertexes. The elements a_{ik} of the adjacency matrix \mathbf{A} are defined as

$$a_{ik} = \begin{cases} 1 & \text{if } q_{ik} \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

for any $i, k \in \sigma$. Notice that all diagonal elements of \mathbf{A} are equal to one. We will say that the vertex k is a neighbor of the vertex i if there is a directed edge that connects i to k , namely if $a_{ik} = 1$. Herewith we do not impose any restriction on the neighborhood relationship, but it must be noticed that by the definition of the absorbing state if $j \in \sigma_A$ then $a_{jk} = 0$ for any $k \in \sigma$. Therefore, the vertices associated to states in σ_A have only incoming edges. A *walk* starting in i and ending in j is an alternating sequence of edges and vertexes, where the same vertex and the same edge may be visited once or more times. On the network \mathcal{G} , a walk can contain self-loops as well. A *path* from i to j is instead a special walk without loops where each state visited during a path is visited only one time.

The generic term of the expansion (12),

$$q_{ik_1} q_{k_1 k_2} \cdots q_{k_\ell j} \quad (14)$$

for $k_1, \dots, k_\ell \in \sigma_0$ is different from zero only if each term in the product is different from zero. If (14) is different from zero, then there exists a walk made of $\ell + 1$ steps from state i to state j through the states k_1, \dots, k_ℓ . If instead the generic term (14) is zero, it means that at least one of the rates in the product is zero. In this case, that precise walk through the states k_1, \dots, k_ℓ does not exist. In summary, each single term in each of the coefficients of the expansion (12) corresponds to a walk on the state space. In particular, the generic term (14) describes a walk through ℓ states, which may include cycles and self loops and is directly associated to the power ℓ of the Taylor series. As a consequence, the coefficient of each ℓ -th power of the Taylor series is the sum of all contributions from walks through exactly ℓ states, some of which possibly visited more than one time.

3. *The shortest path contributes to the coefficient of the first term in the Taylor expansion*

If the process X could transit directly from state i to state j , i.e. if $q_{ij} > 0$, then the states i and j would be directly connected, i.e. $a_{ij} = 1$, and the first contribution to the Taylor series (12) of ρ_{ij} would be the zero order term. If the states i and j are not directly connected, i.e. $a_{ij} = 0$, the zeroth-order term is absent and the first order term might give the first contribution. The first-order term in the expansion of ρ_{ij} will contribute only if the state j is a neighbor of at least one state k that is a neighbor of state i , i.e. if there is at least one state k directly connected to state j and which can be reached from state i in one

step. If there is no such intermediate state, the first-order term will be zero as well.

By direct inspection of the expansion (12) we notice that the first contribution to the Taylor series of ρ_{ij} must come from a path and in particular it will be the shortest path that connects the state i with the state j . If the shortest path consists of $m + 1$ steps, it involves m intermediate states and contributes with the power t^m to the Taylor expansion. If the shortest path is unique, the coefficient of the m -th term gives the product of the rates along this path divided by the absorption probability u_{ij} , as in (12), thus indicating that at short times we have the asymptotic behavior,

$$\rho_{ij}(t) \sim \left(\frac{q_{ik_1} q_{k_1 k_2} \cdots q_{k_m j}}{u_{ij}} \right) \frac{t^m}{m!} + \dots, \quad (15)$$

by neglecting higher order terms for small t .

If the shortest path is not unique, the coefficient is a sum of the product of the rates along several distinct paths, which may share some sub-paths. If the term t^m is the first contribution to the Taylor series of ρ_{ij} then all other higher order terms t^ℓ , for $\ell > m$, do also contribute. Moreover, since the absorption probabilities can be measured experimentally independently of the absorption times, the coefficient of (15) can be probably used to estimate of the strength of the rates.

B. First-passage and first-return times

In the previous section we have shown the general theory concerning the absorption times. In this section we argue instead that the technique with absorbing states is useful to compute several important properties such as the distributions of the first passage times and of the first-return times. The starting point here is a Markov chain Y without absorbing states, which has a regular behavior including a well defined steady state. Let \mathbf{Q} be the infinitesimal generator of Y and let $\sigma_0 = \{0, 1, \dots, n - 1\}$ be the state space of the process. All processes considered here will start from the initial state $i \in \sigma_0$.

1. First passage time to $j \neq i$

Let $j \neq i$ be another state for which we would like to know the distribution of the first passage time starting from i .

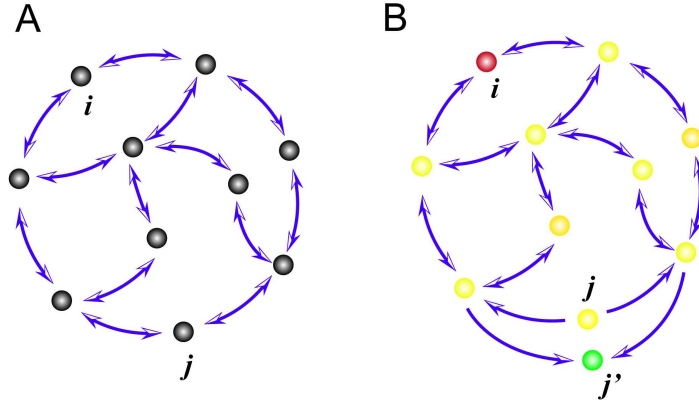


FIG. 2. A network transformation to compute the first-passage times as absorption times. In panel A we have a regular network of states without the absorbing states. To compute the distribution of the first-passage times from i to j , the incoming transitions into the state j are redirected towards a new state j' , thus leading to the network in panel B. The first-passage time from i to j on the network A has the same distribution as the absorption time from i to j' in the network B.

To study this problem, we enlarge the state space of the process by adding one absorbing state j' and by redirecting all rates intended to j towards j' , instead. An illustration of this procedure is shown in figure 2. This defines a new Markov chain X , whose infinitesimal generator \mathbf{Q}' , has the same form as (1) with $\sigma_A = \{j'\}$ and $a = 1$. The probability density of the first passage time $f_{ij}(t)$ is thus given by the probability density $\rho_{ij'}$ of the absorption time in j' starting from i . Given that $q'_{kj'} = q_{kj}$ and using the results obtained above we therefore have

$$f_{ij}(t) = q_{ij} + t \sum_{k_1=0}^{n-1} q_{ik_1} q_{k_1j} + \dots + \frac{t^\ell}{\ell!} \sum_{k_1=0}^{n-1} \dots \sum_{k_\ell=0}^{n-1} q_{ik_1} q_{k_1k_2} \dots q_{k_\ell j} + \dots, \quad (16)$$

which is almost identical to (12) with the difference that the probability of absorption in j' is one. If the shortest path from i to j' is unique and has length m , then the first term of the expansion of f_{ij} is given by

$$f_{ij}(t) \sim (q_{ik_1} q_{k_1k_2} \dots q_{k_mj}) \frac{t^m}{m!} + \dots, \quad (17)$$

by neglecting higher order terms for small t . Thus it indicates that the first coefficient is strictly proportional to the product of the intermediate rates.

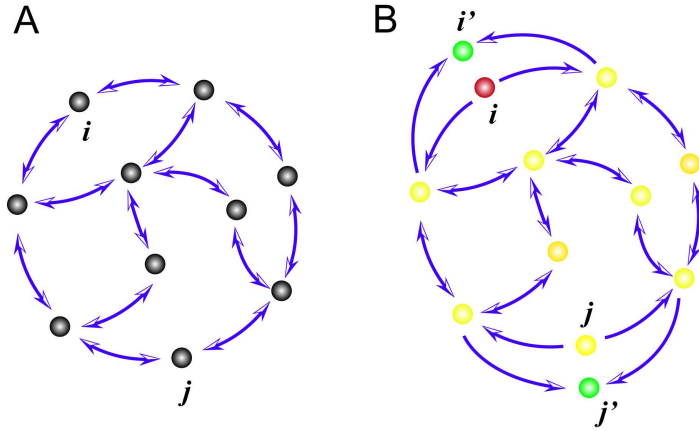


FIG. 3. A network transformation to compute the first-passage times as absorption times by considering only trajectories that are not allowed to return to the initial state i . In panel A we have a regular network of states without absorbing states. To compute distribution of the first-passage time in j excluding i , one needs to exclude all trajectories returning to i . To do this, the incoming transitions into state j are redirected towards a new state j' and the incoming transitions into i are redirected towards an additional absorbing state i' , thus leading to the network in panel B. The condition that the process should not return to i is taken into account by computing the probability that the process is absorbed in j' . This restriction changes the weight of all terms in the sum, which needs to be divided by the absorption probability in j' .

2. First passage time to j without returns to i

One could notice that the expansion (16) includes all terms that start in i and finish in j' . So it includes also those walks on the network of states which visit state i once or more times before arriving at the state j' . Although these walks are not involved in the shortest path from i to j' , their weight in the normalization has an influence on the coefficient of the first term (17) of the expansion.

To consider only those realizations of the process that visit state j' before returning to state i , we need to proceed as follows. We create a second absorbing state i' towards which we redirect all transitions that are originally directed towards state i and we consider a process X with two absorbing states $\sigma_A = \{j', i'\}$, with initial condition in i (see figure 3 for an illustration). At this point, the first-passage time in j conditioned that the process is

not allowed to return to i is given has a probability density function $f_{ij \setminus i}$ given by a simple application of Eq. (12)

$$f_{ij \setminus i} = \frac{q_{ij}}{u_{ij'}} + \frac{t}{u_{ij'}} \sum_{k_1=0}^{n-1} q_{ik_1} q_{k_1j} + \dots + \frac{t^\ell}{u_{ij'} \ell!} \sum_{k_1=0}^{n-1} \dots \sum_{k_\ell=0}^{n-1} q_{ik_1} q_{k_1k_2} \dots q_{k_\ell j} + \dots, \quad (18)$$

where all intermediate states $k_\ell \neq i$ and $u_{ij'}$ is the absorption probability in j' , which is the same as the probability to visit the state j before returning to the state i .

3. First-return times in i

We consider here again the condition discussed at the beginning of the section IIB for the Markov chain Y without absorbing states with the infinitesimal generator \mathbf{Q} on the state space $\sigma_0 = \{0, 1, \dots, n-1\}$. We let $i \in \sigma_0$ be the initial condition of the process and we are interested in the computation of the probability density of the first-return times in i . This is accomplished by considering the arguments from the section IIB by setting $j = i$. The absorbing state i' receives the edges that are otherwise directed towards i (see figure 4 for the illustration). Therefore, the first-return time probability density r_i is equal to the absorption time probability density $\rho_{ii'}$ according to Eq. (12) by setting $u_{ii'} = 1$ and $q_{ki'} = q_{ki}$.

Under the physically relevant assumption that there are no irreversible transitions, however, the first term of the expansion of r_i is certainly linear. Indeed, a transition back to i from any neighbor of i must be possible. This gives the first term of the expansion

$$r_i(t) \sim \left(\sum_k q_{ik} q_{ki} \right) t + \dots, \quad (19)$$

where the non-zero contribution to the sum comes exclusively from those states that are neighbor of i . If the state i is accessible experimentally under a variety of external conditions, the short time behavior of the return time probability density may reveal important dependencies of the rates on the varying parameters.

III. SUMMARY AND CONCLUSIONS

It is shown that for a general Markov chain the distribution of the first-passage times from a state i to any other state j can be computed as the distribution of absorption times after a

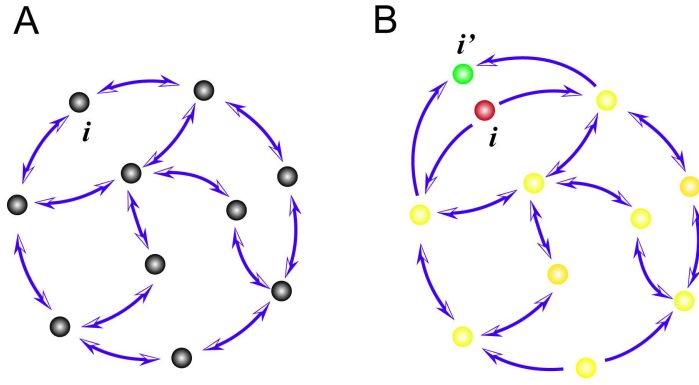


FIG. 4. A network transformation to compute the first-return times as absorption times. Starting from a generic network in panel A, the incoming transitions into state i are redirected towards a new state i' , which becomes a new absorbing state. This leads to the network in panel B. All processes starting in i will eventually be absorbed in i' . The first-return time has the same distribution as the absorption time from i to i' . From this example it is also clear that the first contribution to the Taylor series of the return times is given by paths of length two, through the neighbors of the state i .

suitable modification of the state space. Therefore focused our attention on the absorption time distribution from an initial state to a chosen absorbing state in the presence of other absorbing states. There, we have been able to exactly compute the complete Taylor series of the absorption time distribution. It has been found that each coefficient of the series can be written explicitly as a sum of the products of the transition rates.

By employing graph theoretical methods, we also determined that each single term in the coefficients of the expansion corresponds to a walk that includes cycles and self-loops in the state space. This allowed us to show that the number of states visited during each walk corresponds exactly to the power of the corresponding term of the Taylor series. Furthermore, it was argued that the first term of the Taylor series must correspond to a specific path. In particular, since the first term of the Taylor series corresponds to the smallest power of the series, the coefficient corresponds to the shortest path on the network between the initial and the absorbing state. If the shortest path is unique, we have shown that the coefficient of the first term of the Taylor series can be written in terms of the product of the rates along this path. In addition, it was found that this general framework can be easily adapted

to several other situations. It includes the first-passage time from a recurrent to another recurrent state, with or without visiting the initial state, and the return time to any given experimentally accessible state.

It is important to note that our results are independent on the network's topology. We believe that the explicit expression of the coefficients in terms of the rates could be practically used to selectively determine the functional dependency of the rates on the external parameters from the experimental data. In this respect, a particularly useful result concerns the return time to a fixed initial condition. In this case, indeed, there is always only one intermediate state and the short time expansion of the probability density must be linear so that the only unknown is the coefficient. It will be very valuable if these theoretical predictions could be tested in single-molecule experimental measurements for complex for complex biochemical systems.

-
- [1] J. Goutsias and G. Jenkinson. Markovian dynamics on complex reaction networks. *Phys. Rep.*, 529:199–264, 2013.
 - [2] N. J. Carter and R. A. Cross. Mechanics of the kinesin step. *Nature*, 435:308–312, 2005.
 - [3] Anatoly B. Kolomeisky and Michael E. Fisher. Molecular motors: a theorist's perspective. *Annu Rev Phys Chem*, 58:675–695, 2007.
 - [4] S. Liepelt and R. Lipowsky. Kinesin's network of chemomechanical motor cycles. *Phys. Rev. Lett.*, 98, 2007.
 - [5] S. N. Xie. Single-molecule approach to enzymology. *Single Molecules*, 2:229–236, 2001.
 - [6] B. P. English, W. Min, A. M. van Oijen, K. T. Lee, G. B. Luo, H. Y. Sun, B. J. Cherayil, S. C. Kou, and X. S. Xie. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat. Chem. Biol.*, 2:87–94, 2006.
 - [7] A. Valleriani, S. Liepelt, and R. Lipowsky. Dwell time distributions for kinesin's mechanical steps. *Epl*, 82, 2008.
 - [8] P. Keller, S. Roelly, and A. Valleriani. A Quasi Random Walk to Model a Biological Transport Process. *Methodology and Computing in Applied Probability*, 2013.
 - [9] Peter Keller and Angelo Valleriani. Single-molecule stochastic times in a reversible bimolecular reaction. *J Chem Phys*, 137(8):084106, Aug 2012.

- [10] A Szabo, K Schulten, and Z Schulten. 1st passage time approach to diffusion controlled reactions. *The Journal of chemical physics*, 72(8):4350–4357, 1980.
- [11] Anatoly Kolomeisky, Evgeny Stukalin, and Alex Popov. Understanding mechanochemical coupling in kinesins using first-passage-time processes. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 71(3 Pt 1):031902, 2005.
- [12] Xin Li and Anatoly B. Kolomeisky. Mechanisms and topology determination of complex chemical and biological network systems from first-passage theoretical approach. *The Journal of Chemical Physics*, 139(14):-, 2013.
- [13] Marcel F. Neuts. Probability distributions of phase type. *Liber Amicorum Prof. Emeritus H. Florin*, 1975.