

CHAPTER 8

Analysis of Experimental Observables and Oscillations in Single-Molecule Kinetics

Marcel O. Vlad^{*,†} and John Ross^{*}

**Department of Chemistry, Stanford University
Stanford CA 94305-5080 USA*

*†Institute of Mathematical Statistics and Applied Mathematics
Casa Academiei Romane, Calea 13 Septembrie 13, Bucharest
050711 Romania*

1. Introduction

This chapter deals with the theoretical analysis of different types of oscillations which may occur in single-molecule kinetics and their influence on different experimental observables. The main difference between a single-molecule system and a macroscopic kinetic system is that for a macroscopic system the details of intramolecular dynamics are lost due to the overlapping of a large number of signals produced by the different molecules present in the system, whereas for a single-molecule system these microscopic signals have a direct influence on the experimental observables. In particular, oscillations of observables due to intramolecular dynamics, which in macroscopic experiments are smoothed out and do not show up in the observed data, can be observed directly in single-molecule experiments.

In this chapter we do not intend to develop a general theory of oscillations in single-molecule kinetics; instead we consider a few

models, which can be investigated in detail. We intend to show the reader how to build different stochastic models for single-molecule kinetics, with special reference to oscillations. Experiments in single-molecule kinetics¹⁻⁷ consist in studying the chemical changes of one large molecule, such as a protein or an enzyme, immobilized on a support; the process may involve either a large molecule alone or a large molecule interacting with smaller molecules; intramolecular and molecular fluctuations are large and thus the deterministic mass action laws of chemical kinetics do not hold and are replaced by probabilistic laws. Although it would be desirable to develop a microscopic description based on nonequilibrium statistical mechanics, this is an extremely difficult task and thus the approaches in single-molecule kinetics are based on a stochastic, mesoscopic description involving two different types of stochastic processes. In the following we use variations of the basic model: (a) A single-molecule can exist in different chemical states $u = 1, 2, \dots$ and the random transitions from one chemical state to another can be described by a local, Markovian master equation with time-dependent transition rates, $k_{uu'} = k_{uu'}(t)$. (b) Due to the conformational and other (energy) fluctuations in the single molecule, the rate coefficients $k_{uu'}(t)$ themselves are random functions of time. Some approaches consider directly the fluctuations of the rate coefficients, whereas other approaches assume that the stochastic properties of the rate coefficients can be represented in terms of a set of control parameters,^{3,7-9} such as total energy of the molecule or the energy corresponding to a given degree of freedom; in this chapter we use both approaches, (a) and (b). In both cases we can write a Markovian master equation with random rate coefficients for the probability $P_u(t)$ that the molecule is in the chemical state u at time t

$$\frac{\partial}{\partial t} P_u(t) = \sum_{u' \neq u} P_{u'}(t) k_{u'u}(t) - P_u(t) \sum_{u'' \neq u} k_{uu''}(t), \quad (1)$$

where $k_{uu'}(t)$ is the rate of transition (rate coefficient) from the state u to the state u' at time t . The rate coefficients are random functions

of time; thus, we need additional stochastic equations for describing the fluctuations of the rate coefficients $k_{uu'}(t)$. The simplest approximation is to neglect the fluctuations of the rate coefficients altogether and describe the kinetics of the process in terms of a master equation with constant coefficients; such an approach is similar to traditional chemical kinetics and is used as a first approximation. More sophisticated approaches make certain assumptions regarding the fluctuations of the rate coefficients based on theoretical models or experimental observations. In this chapter, we assume that the fluctuations of the rate coefficients can be described in terms of known characteristic functionals.

Among the different states $u = 1, 2, \dots$ of the molecule studied, some are fluorescent and some are not. The molecule undergoes a random walk among these states, resulting in random variations of the fluorescent signal. The direct, raw experimental observable in a single-molecule experiment is the fluorescent signal $I(t)$ as a function of time, collected from the molecule studied; since the experiments are usually carried out in a time-independent regime, the time series describing the evolution of $I(t)$ is stationary. The most commonly used approach of data analysis is based on the computation of the correlation functions¹⁰ of the fluorescence signal at times t_1, \dots, t_m

$$C_m = \langle (I(t_1) - \langle I(t_1) \rangle) \cdots (I(t_m) - \langle I(t_m) \rangle) \rangle. \quad (2)$$

The second type of observables include the on/off time distributions,^{6,7} that is, the distributions of the time intervals for which the fluorescent signal is on or off, respectively.

The third type of observables which can be extracted from the experimental data include the statistical properties of the numbers of reaction events,¹¹ that is, the numbers of occurrences of different reactions occurring in a given time interval. The models used in this paper are variations of the basic model based on the master equation with random coefficients (1) supplemented by suitable descriptions for the fluctuations of the rate coefficients. In the following sections

we focus on using these models for investigating the connections among the three types of experimental observables mentioned earlier, and the possible occurrence of oscillations in single-molecule kinetics.

2. Correlation functions and oscillations

Following Ref. 9 we start out by considering a special class of systems for which the fluctuating rate coefficients obey a separability condition $k_{uu'} = k_{uu'}^0 \chi(\mathbf{s})$, that is, they are made up of the multiplicative contributions of two factors: (a) a universal factor, $\chi(\mathbf{s})$ which is fluctuating and is the same for all interaction processes and (b) process-dependent factors, $k_{uu'}^0$ which depend on the initial and final chemical states of the molecule but are not random. This separability condition makes it possible to introduce an intrinsic timescale and use the method of characteristic functionals for computing the correlation functions of the fluorescent signal. The separability condition is consistent with the condition of detailed balance for a system with a unique equilibrium state and is automatically fulfilled by a system with two chemical states and a unique equilibrium.

In Ref. 9, the theory was developed for the general case when the single molecule has an arbitrary number of chemical states and general expressions were derived for correlation functions of all orders. However, for simplicity, we begin by considering a system with two different chemical states. By applying the general theory developed in the study of Ref. 9, we obtain the following formula for the second-order correlation function:

$$\langle \Delta I(t) \Delta I(t + \tau) \rangle = \frac{K}{(K + 1)^2} \mathcal{J}(\tau), \quad (3)$$

where

$$K = k_+(t)/k_-(t) \text{ independent of } t \quad (4)$$

is the equilibrium constant of the process, which, for a single chemical equilibrium is time-independent and not random. The term

$$\mathcal{J}(\Delta t) = \left\langle \exp \left(- \int_0^{\Delta t} k_{\Sigma}(t') dt' \right) \right\rangle \quad (5)$$

is a dynamic damping factor and

$$k_{\Sigma}(t) = k_{+}(t) + k_{-}(t) \quad (6)$$

is a total fluctuating rate coefficient, which is the sum of forward and backward reaction rates $k_{+}(t)$ and $k_{-}(t)$, respectively, and $\langle \dots \rangle$ denotes a dynamic average over all possible values of the total fluctuating rate coefficient $k_{\Sigma}(t)$. For a system with two chemical states the separability condition mentioned earlier is automatically fulfilled and Eq. (3) is not subjected to any restriction.

If we assume that the cumulants $\langle \langle k_{\Sigma}(t) \rangle \rangle$, $\langle \langle k_{\Sigma}(t_1) k_{\Sigma}(t_2) \rangle \rangle \dots$ of the total rate coefficient exist and are finite, the damping factor can be expressed by a cumulant expansion. We arrive at

$$\begin{aligned} \mathcal{J}(\Delta t) = \exp \left\{ \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \right. \\ \left. \times \int_0^{\Delta t} \cdots \int_0^{\Delta t} \langle \langle k_{\Sigma}(t_1) \cdots k_{\Sigma}(t_m) \rangle \rangle dt_1 \cdots dt_m \right\}. \quad (7) \end{aligned}$$

The data can be analyzed in terms of the effective decay rate

$$\begin{aligned} k_{\text{eff}}(\Delta t) &= - \frac{\partial}{\partial \Delta t} \ln \mathcal{J}(\Delta t) \\ &= \langle k_{\Sigma} \rangle + \frac{\partial}{\partial \Delta t} \sum_{m=2}^{\infty} \frac{(-1)^{m-1}}{m!} \\ &\quad \times \int_0^{\Delta t} \cdots \int_0^{\Delta t} \langle \langle k_{\Sigma}(t_1) \cdots k_{\Sigma}(t_m) \rangle \rangle dt_1 \cdots dt_m. \quad (8) \end{aligned}$$

The effective decay rate bears information about the nature of intramolecular fluctuations. If the fluctuations of the rate of change

are of short range in time (Markovian or independent fluctuations), then in the long run, the effective decay rate is independent of the time difference Δt

$$k_{\text{eff}}(\Delta t) = \text{independent of } \Delta t \text{ as } \Delta t \rightarrow \infty. \quad (9)$$

If condition (9) is fulfilled by the experimental data, then the intramolecular fluctuations are of short range in time. If $k_{\text{eff}}(\Delta t)$ varies with Δt for large time differences Δt , then the intramolecular fluctuations are of long range. In addition, we notice that the effective rate is a better function for identifying the existence of oscillations in single-molecule kinetics than the correlation function of the fluorescent signal. In order to evaluate $k_{\text{eff}}(\Delta t)$, however, accurate measurements are necessary.

In order to express the contribution of intramolecular fluctuations to the effective decay rate, we evaluate the difference

$$\begin{aligned} \Delta k_{\text{eff}}(\Delta t) &= k_{\text{eff}}(\Delta t) - \langle k_{\Sigma}(t) \rangle \\ &= \frac{\partial}{\partial \Delta t} \sum_{m=2}^{\infty} \frac{(-1)^{m-1}}{m!} \\ &\quad \times \int_0^{\Delta t} \cdots \int_0^{\Delta t} \langle \langle k_{\Sigma}(t_1) \cdots k_{\Sigma}(t_m) \rangle \rangle dt_1 \cdots dt_m. \end{aligned} \quad (10)$$

In the particular case of Gaussian fluctuations of the total rate coefficient, all cumulants of order higher than 2 vanish and the difference $\Delta k_{\text{eff}}(\Delta t)$ is simply given by

$$\begin{aligned} \Delta k_{\text{eff}}(\Delta t) &= -\frac{1}{2} \frac{\partial}{\partial \Delta t} \int_0^{\Delta t} \int_0^{\Delta t} \langle \langle k_{\Sigma}(t_1) k_{\Sigma}(t_2) \rangle \rangle dt_1 dt_2 \\ &= -k_{\Sigma} \frac{\partial}{\partial \Delta t} \int_0^{\Delta t} (\Delta t - x) g(x) dx, \end{aligned} \quad (11)$$

where

$$g(|t_2 - t_1|) = \frac{1}{\langle k_{\Sigma} \rangle^2} \langle \langle k_{\Sigma}(t_1) k_{\Sigma}(t_2) \rangle \rangle \quad (12)$$

is the relative value of the correlation function of the total rate coefficient. The relative correlation function can be evaluated from experimental data by solving Eq. (11) for $g(\Delta t)$, resulting in

$$g(\Delta t) = -\frac{1}{k_{\Sigma}} \frac{\partial}{\partial \Delta t} \Delta k_{\text{eff}}(\Delta t). \quad (13)$$

This simplified model makes it possible to discuss the possible existence of damped oscillations in single-molecule kinetics. We start out by analyzing the oscillations due to the intramolecular fluctuations. For simplicity we limit ourselves to the case of Gaussian fluctuations. In this case it is easy to show that if the relative correlation function of the rate of change, $g(\Delta t)$, which expresses the intramolecular fluctuations, displays damped oscillations, then damped oscillations may also occur in the correlation functions of the fluorescent signal. According to the normal mode theory,¹² the function $g(\Delta t)$ can be expressed as

$$g(|t_1 - t_2|) = \sum_q c_q \exp[-\varepsilon_q |t_1 - t_2|] + \int_q c(q) \exp[-\varepsilon(q) |t_1 - t_2|] dq. \quad (14)$$

In this equation, c_q , $c(q)$, ε_q , $\text{Re}(\varepsilon_q) > 0$ and $\varepsilon(q)$, $\text{Re}(\varepsilon(q)) > 0$ are amplitude and frequency factors attached to the different normal modes. Since, in general, both c_q and ε_q are complex, their values must be chosen in such a way that the corresponding Gaussian process is physically consistent. For a purely discrete mode spectrum in Eq. (14) the integral term is missing, and the stochastic process, even though generally nonMarkovian, has short memory. The Markovian memory corresponds to a single exponential, that is, to a single mode. For a discrete spectrum, the Markovian approximation is accurate for large time differences, because in this case the main contribution to the sum in Eq. (14) is given by a single exponential which corresponds to the frequency with the smallest absolute value. If the mode spectrum has a continuum branch, then the tail of the correlation function

may obey a scaling law of the inverse power type and the system may display long memory. Our analysis in this chapter is limited to the case of short-range fluctuations, for which Eq. (14) contains only the contribution of the discrete spectrum.

We consider the following physical constraints:

- (1) For a time difference equal to zero, the autocorrelation function is equal to the dispersion of the total relative rate, $v(t) = k_{\Sigma}(t)/\langle k_{\Sigma}(t) \rangle$ at time t , $\langle \langle v^2(t) \rangle \rangle$, which, by definition, must be non-negative.
- (2) Since the characteristic frequency is a real function of time, the modes with complex frequencies ε_q must occur in conjugated pairs.
- (3) For large times the autocorrelation function of the relative total rate must decay to zero.

We keep in Eq. (14) only the contribution of the discrete spectrum and express the contribution of real eigenvalues $\varepsilon_q^{(\text{real})}$ and of complex eigenvalues $\varepsilon_q^{(\text{compl})} = \mu_q \pm i\sigma_q$. After some calculations, we obtain

$$\begin{aligned}
 g(|t_1 - t_2|) &= \sum_{\text{real values}} c_q^{(\text{real})} \exp[-\varepsilon_q^{(\text{real})}|t_1 - t_2|] \\
 &\quad + \sum_{\text{complex values}} c_q^{(\text{compl})} \exp[-\varepsilon_q^{(\text{compl})}|t_1 - t_2|] \\
 &= \sum_{\text{real values}} c_q^{(\text{real})} \exp[-\varepsilon_q^{(\text{real})}|t_1 - t_2|] \\
 &\quad + \sum_{\text{complex values}} 2\{a_q \cos[\sigma_q|t_1 - t_2|] \\
 &\quad + b_q \sin[\sigma_q|t_1 - t_2|]\} \exp[-\mu_q|t_1 - t_2|], \quad (15)
 \end{aligned}$$

where a_q and b_q are the real and imaginary parts of the complex amplitude factors, and $c_q^{(\text{compl})} = a_q \pm ib_q$. In order that the constraints (1)–(3) be valid we introduce the following restrictions for

the parameters in Eq. (15):

$$c_q^{(\text{real})}, a_q > 0, \quad b_q > 0 \quad (16)$$

$$\varepsilon_q^{(\text{real})}, \mu_q > 0. \quad (17)$$

The restrictions (16) ensure that the dispersion of the characteristic frequency is non-negative, whereas the restrictions (17) are necessary in order that the autocorrelation function tends to zero for large time differences. The damping factor of the correlation function $\mathcal{J}(\Delta t)$ can be expressed as

$$\mathcal{J}(\Delta t) = \exp[-k_\Sigma \Delta t + k_\Sigma \Theta(\Delta t)], \quad (18)$$

where $\Theta(\Delta t)$ is a phase factor given by

$$\begin{aligned} \Theta(\Delta t) = \Delta t & \left[\sum_{\text{real values}} \frac{c_q^{(\text{real})}}{\varepsilon_q} + \sum_{\text{complex values}} \frac{2(a_q \mu_q + b_q \sigma_q)}{(\mu_q)^2 + (\sigma_q)^2} \right] \\ & - \sum_{\text{real values}} \frac{c_q^{(\text{real})}}{(\varepsilon_q)^2} \\ & + 2 \sum_{\text{complex values}} \left[\frac{a_q [(\sigma_q)^2 - (\mu_q)^2] - 2b_q \mu_q \sigma_q}{[(\mu_q)^2 + (\sigma_q)^2]^2} \right] \\ & + \sum_{\text{real values}} \frac{c_q^{(\text{real})}}{(\nu_q)^2} \exp(-\varepsilon_q \Delta t) + \frac{2}{[(\mu_q)^2 + (\sigma_q)^2]^2} \\ & \times \sum_{\text{complex values}} \{a_q [(\mu_q)^2 - (\sigma_q)^2] \cos(\sigma_q \Delta t) + 2b_q \mu_q \sigma_q\} \\ & \times \exp(-\mu_q \Delta t) + \frac{2}{[(\mu_q)^2 + (\sigma_q)^2]^2} \\ & \times \sum_{\text{complex values}} \{b_q [(\mu_q)^2 - (\sigma_q)^2] \sin(\sigma_q \Delta t) - 2a_q \mu_q \sigma_q\} \\ & \times \exp(-\mu_q \Delta t). \quad (19) \end{aligned}$$

The phase factor $\Theta(\Delta t)$ has the following asymptotic behavior:

$$\Theta(\Delta t) \sim \begin{cases} \mathcal{M}\Delta t^2 & \text{as } \Delta t \rightarrow 0 \\ \mathfrak{S}\Delta t & \text{as } \Delta t \rightarrow \infty \end{cases}, \quad (20)$$

where the proportionality factors \mathcal{M} and \mathfrak{S} are given by

$$\mathcal{M} = \frac{1}{2} \sum_{\text{real values}} c_q^{(\text{real})} + \sum_{\text{complex values}} a_q > 0, \quad (21)$$

$$\mathfrak{S} = \sum_{\text{real values}} \frac{c_q^{(\text{real})}}{\varepsilon_q} + \sum_{\text{complex values}} \frac{2(a_q\mu_q + b_q\sigma_q)}{(\mu_q)^2 + (\sigma_q)^2} > 0. \quad (22)$$

In this case the variation of the effective rate coefficient has the asymptotic behavior

$$\Delta k_{\text{eff}}(\Delta t) \sim k_{\Sigma} \begin{cases} -2\mathcal{M}\Delta t & \text{as } \Delta t \rightarrow 0 \\ -\mathfrak{S} & \text{as } \Delta t \rightarrow \infty \end{cases}. \quad (23)$$

For large as well as short time differences, the variation $\Delta k_{\text{eff}}(\Delta t)$ is negative: as expected, for large time differences the variation is constant, as it should be for short-range intramolecular fluctuations.

According to Eq. (19) the phase factor $\Theta(\Delta t)$ may display damped oscillations in the time difference Δt . From Eq. (18) it follows that the same type of damped oscillation must be displayed by the damping factor $\mathcal{I}(\Delta t)$. We did a numerical study of the possible occurrence of damped oscillations in the correlation functions of the fluorescent signal, due to the presence of damped oscillations in intramolecular dynamics, represented by the complex eigenmodes in Eq. (19). In order for the damped oscillations to show up in the correlation functions on the fluorescent signal, it is necessary that the timescale of the chemical process be of the same order of magnitude as the timescale of intramolecular dynamics. Figure 1 shows such a damped oscillating behavior for the second-order correlation function of the fluorescent signal $C_2(\tau) = \langle \Delta I(t + \tau) \Delta I(t) \rangle$, which is similar to the oscillations observed in the experiments of Edman and Rigler.¹⁰ Similar behavior is displayed by the correlation functions of higher order.

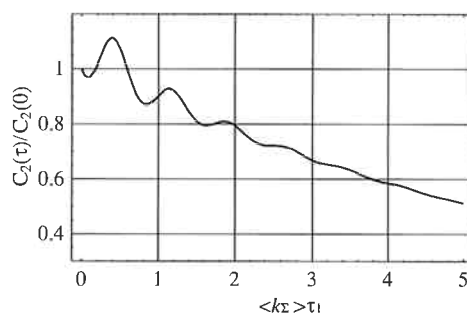


Fig. 1. Graphical representation of the absolute value of the second-order correlation function $C_2(\tau)$ versus the time difference $\tau = |t_2 - t_1|$ for a two-state Gaussian model. (The relative correlation function $g(\Delta t)$ is represented by a linear combination of complex exponential terms.)

It has been suggested that the damped oscillations of the correlation functions of high order can be used for the characterization of the nonMarkovian nature of the two-state fluorescent process.¹⁰ A non-Markovian function (NMF) has been defined in terms of the second- and third-order correlation functions of the fluorescent signal:

$$\text{NMF}(\tau_1, \tau_2) = p_f \left[\frac{C_3(\tau_1, \tau_2)}{C_2(\tau_2)} - C_2(\tau_1) \right], \quad (24)$$

where p_f is the stationary probability that the molecule is in a fluorescent state. Figure 2 shows a typical memory landscape for the NMF computed by applying our approach from Eqs. (3)–(5), (18), and (19), and the expression

$$\langle \Delta I(t) \Delta I(t + \tau_1) \Delta I(t + \tau_1 + \tau_2) \rangle = \mathcal{J}(\tau_1 + \tau_2) \frac{K(K-1)}{(K+1)^3}, \quad (25)$$

for the third-order correlation function, computed by applying our theory presented in Ref. 9. The computed landscape displays the same type of damped oscillations as the ones observed in the experiments of Edman and Rigler.¹⁰

Another possible cause for the occurrence of the damped oscillations of the correlation functions is the interaction between chemical kinetics and intramolecular dynamics. This cause was suggested by

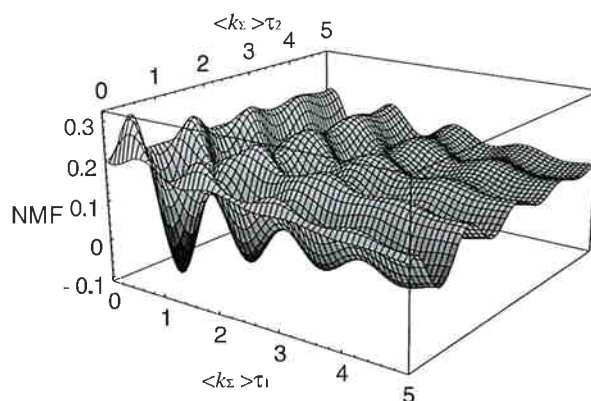


Fig. 2. Graphical representation of the non-Markovian function $NMF(\tau_1, \tau_2)$ of Edman and Rigler¹⁰ versus the time differences τ_1 and τ_2 for a two-state Gaussian model. (The relative correlation function $g(\Delta t)$ is represented by a linear combination of complex exponential terms.)

Edman and Rigler¹⁰ in order to explain their experimental data on the oxidation reaction involving a single molecule of immobilized horseradish peroxidase. These authors neglected the random fluctuations of the rate coefficients and assumed that the kinetics of the process can be described by a simplified form of the master equation (1), where the rate coefficients $k_{u'u}$ are constant. They have chosen sets of rates $k_{u'u}$, which correspond to closed loops of states and violate detailed balance. It has been theoretically proven that a master equation with rates $k_{u'u}$, which violate the detailed balance, are capable of producing damped oscillations.⁹ The model of Edman and Rigler¹⁰ seems to contradict the principles of statistical mechanics, because they evaluate time invariant, equilibrium correlation functions by using a model which violates the detailed balance. However, we show shortly that this is not necessarily the case. A single molecule is not a macroscopic system; therefore, it does not have to obey equilibrium statistical mechanics; however, a single molecule is not isolated, but connected to its environment, and in most experiments, the ensemble molecule plus environment are at statistical equilibrium. Actually, most experimental studies of single-molecule

kinetics involve the measurement of the regression of the equilibrium fluctuations of the fluorescent signal.

We consider a different approach, which shows that the model of Edman and Rigler¹⁰ may be correct and do not violate detailed balance. We assume that the intramolecular dynamics, expressed in terms of the control parameters $\mathbf{s}(t) = (s_1(t), s_2(t), \dots)$, can be described by a Markovian stochastic process. We denote by $\mathcal{R}(\mathbf{s}; t)d\mathbf{s}$ the probability that at time t the vector of control parameters is between \mathbf{s} and $\mathbf{s} + d\mathbf{s}$ and assume that its time evolution is described by a linear evolution equation:

$$\frac{\partial}{\partial t} \mathcal{R}(\mathbf{s}; t) = \mathbb{L} \mathcal{R}(\mathbf{s}; t), \quad (26)$$

where \mathbb{L} is a Markovian operator of the Fokker–Planck, master or Liouville type. We introduce the joint probability density $B_u(\mathbf{s}; t)$ for the chemical state of the molecule u and the control vector \mathbf{s} . This joint probability density is the solution of a compound stochastic Liouville equation^{13,14}:

$$\begin{aligned} \frac{\partial}{\partial t} B_u(\mathbf{s}; t) = & \mathbb{L} B_u(\mathbf{s}; t) + \sum_{u' \neq u} B_{u'}(\mathbf{s}; t) k_{u'u}(\mathbf{s}) \\ & - B_u(\mathbf{s}; t) \sum_{u' \neq u} k_{uu'}(\mathbf{s}). \end{aligned} \quad (27)$$

We are interested in the evaluation of the marginal probability

$$P_u(t) = \int B_u(\mathbf{s}; t) d\mathbf{s}, \quad (28)$$

in terms of which we can compute the experimental observables, the correlation functions of the fluorescence signal. A simple way would be to derive an approximate equation for the marginal probability $P_u(t)$ by eliminating the stochastic vector \mathbf{s} from Eq. (27). This is a standard topic in statistical physics,^{13,14} which is usually referred to