

Rate Constants from Uncorrelated Single-Molecule Data

Marián Bogañá,[†] Lisen Kullman,[‡] Sergey M. Bezrukov,^{‡,§} Alexander M. Berezhkovskii,^{†,||} and George H. Weiss^{*,†}

Center for Information Technology and Laboratory of Physical and Structural Biology, NICHD, National Institutes of Health, Bethesda, Maryland 20892, St. Petersburg Nuclear Physics Institute, Gatchina 188350, Russia, and Karpov Institute of Physical Chemistry, Ul. Vorontsovo 10, 103064, Moscow K-64, Russia

Received: February 18, 2001; In Final Form: April 24, 2001

We describe a method for estimating rate constants of a two-state Markov system from data obtained by a single-molecule (SM) type of experiment corrupted by white noise. The method is effective even when the characteristic (or resolution) time of the measurement is much longer than a characteristic time of the two-state system, and the data are uncorrelated. We apply our suggested method to data from an SM experiment on the transport of sugar molecules through cell membrane channels. The method is quite accurate in analyzing this set of data. We also demonstrate that in some cases the method can be used to check whether the system is Markovian or not.

1. Introduction

Recent successes in experimentally observing the behavior of single molecules,^{1–6} enables one to study dynamic properties of a system without perturbing it from its equilibrium state. This cannot be done using traditional bulk measurements, in which one perturbs the system and then observes the decay to the equilibrium state. There is a significant difference between bulk and single-molecule (SM) experiments. The first of these deals with ensemble averages while the second monitors the behavior of single molecules out of the ensemble and provides a time average of some observable for a given “trajectory” of the molecule. Different realizations of the SM trajectory will produce different values of the time average. Repetition of the experiment a number of times provides an estimate of the probability density of the time average. Because of the property of ergodicity this density must approach a delta function form centered at the value of the bulk average as the time interval used in averaging tends to infinity. When the measurement time is finite, the density estimate will differ from the delta function form, thereby providing additional information to estimate dynamical properties of the system.

It is worth noting that different realizations of SM trajectories can be found from measurements on different molecules randomly chosen from the ensemble or can be different pieces of a long trajectory produced by a single molecule. In the first case the averages will not be correlated, while in the second case they will be. These correlations can be exploited to gain information about properties of the dynamics. Standard techniques for analyzing data from a single realization of a stochastic process are based on correlation or spectral analysis.

The time interval used in averaging, the resolution time of the SM measurement, is an important parameter that potentially restricts the applicability of spectral analysis. Spectral analysis is useful when the dynamics of interest are slow in the sense

that its characteristic time is larger than the resolution time.^{7,8} However, when the system dynamics are fast and its characteristic time is shorter than the resolution time, spectral analysis will fail because data obtained from a single trajectory will be almost uncorrelated.

In this paper we describe a method for estimating the rate constants of a two-state Markovian system that is effective even when the data are completely uncorrelated. This is done in the following section where we also discuss the sampling error inherent in these estimates. A demonstration of its efficiency will be given in section 3 by applying it to data from an SM experiment on the transport of sugar molecules through cell membrane channels.⁹ In this experiment a single trimeric maltoporin channel is used to observe the passage of single sugar molecules. When a sugar molecule is in the channel, the current through the trimer decreases.

Our approach to the estimation of rate constants for two-state Markovian systems can, in some cases, allow one to check the correctness of the assumption that the experimental system is Markovian. Simulations relating to this point are given in section 4.

2. Mathematical Framework

We analyze a system whose dynamics are defined by random transitions between two states, 1 and 2, so that the trajectory of the system can be described as an alternating sequence of sojourns in the two states. Let $\psi_i(t)$ be the probability density for a single sojourn time in state i . When the system is Markovian, so that $\psi_i(t) = k_i e^{-k_i t}$, $i = 1, 2$, where the k_i are rate constants, it can be described by the simple scheme



In a SM experiment a physical observable of the system is measured for time T . For single-channel recordings the observable is the measured current. Let I_i be the intensity of this observable when the system is in the state i . In addition, one

[†] Center for Information Technology, NIH.

[‡] Laboratory of Physical and Structural Biology, NIH.

[§] St. Petersburg Nuclear Physics Institute. Currently employed at NIH.

^{||} Karpov Institute of Physical Chemistry. Currently employed at NIH.

records noise that has to be taken into account. We can model this process as

$$I(t) = I_{\text{dic}}(t) + D^{1/2}\xi(t) \quad (2.2)$$

where $I_{\text{dic}}(t)$ is a dichotomic process randomly taking the values I_1 and I_2 , and $\xi(t)$ is zero-mean Gaussian white noise, δ -function correlated and independent of $I_{\text{dic}}(t)$. The parameter D is the amplitude of this noise term.

If the resolution time of the device is Δt , the current $I_n(\Delta t)$ measured at time $t_n = n\Delta t$ is

$$I_n(\Delta t) = \frac{1}{\Delta t} \int_{t_{n-1}}^{t_n} I(t) dt = \frac{1}{\Delta t} \int_{t_{n-1}}^{t_n} I_{\text{dic}}(t) dt + \frac{D^{1/2}}{\Delta t} \int_{t_{n-1}}^{t_n} \xi(t) dt \quad (2.3)$$

The first term on the right-hand side of this equation is related to the fraction of time, out of the total time Δt , that the system has been in state 1. This fraction will be written as $x_n(\Delta t)$. The term $\xi(t) dt$ is the differential of the Wiener process, which we express as $\xi(t) dt = dW(t)$. Therefore, we can write

$$I_n(\Delta t) = I_2 + x_n(\Delta t)\Delta I + \frac{D^{1/2}}{\Delta t} \Delta W_n \quad (2.4)$$

where $\Delta I = I_1 - I_2$ and $\Delta W_n \equiv W(t_n) - W(t_{n-1})$. The Wiener process is characterized by the properties (i) the increments are all independent, (ii) the increments have a Gaussian distribution, and (iii) the first moment of ΔW_n is equal to zero and the second moment, is $\langle \Delta W_n^2 \rangle = \Delta t$. Under these conditions it is possible to calculate the moments of $I(\Delta t)$ from (2.4). The first moment and variance can be written as

$$\langle I(\Delta t) \rangle = I_2 + \langle x(\Delta t) \rangle \Delta I \quad (2.5)$$

$$\sigma_I^2(\Delta t) = \Delta I^2 \sigma_x^2(\Delta t) + \frac{D}{\Delta t} \quad (2.6)$$

We therefore have a relation between the average and variance of the observed current $I(\Delta t)$ and the fraction of time out of Δt that the system has spent in state 1.

Expressions for the moments of $x(\Delta t)$ can be derived from results obtained in refs 10 and 11. These can be used to estimate the two rate constants. For the first moment we can write

$$\langle x(\Delta t) \rangle = p_1^{\text{eq}} = \frac{k_2}{k} \quad (2.7)$$

where $k \equiv k_1 + k_2$ and p_1^{eq} is the equilibrium probability for the system to be in state 1. The theoretical expression for the variance is^{10,11}

$$\sigma_x^2(\Delta t) = \frac{2p_1^{\text{eq}} p_2^{\text{eq}}}{k\Delta t} \left[1 - \frac{1}{k\Delta t} (1 - e^{-k\Delta t}) \right] \quad (2.8)$$

where $p_2^{\text{eq}} \equiv 1 - p_1^{\text{eq}}$. Since we are primarily interested in the limit of bad resolution, which means that Δt is greater than the characteristic time of the process, k^{-1} , we can use an asymptotic form of eq 2.8

$$\sigma_x^2(\Delta t) \approx \frac{2p_1^{\text{eq}} p_2^{\text{eq}}}{k\Delta t} \quad (2.9)$$

Values of $k\Delta t$ of the order of 10 or more generally suffice to validate this approximation.

The parameters needed to find k_1 and k_2 are p_1^{eq} and $\sigma_x^2(\Delta t)$. These, together with the relations in eqs 2.7 and 2.9 provide estimates of the rate constants. We need the following information to do so: the intensities of the current in the two states, I_1 and I_2 ; the intensity of the noise, D ; and the experimental series $\{I_n(\Delta t)\}$ where $n = 1, \dots, N$ with $N = T/\Delta t$. The estimates of the first moment and variance are given by the standard formulas

$$\bar{I}(\Delta t) = \frac{1}{N} \sum_{n=1}^N I_n(\Delta t) \quad (2.10)$$

$$\bar{\sigma}_I^2(\Delta t) = \frac{1}{N-1} \sum_{n=1}^N [I_n(\Delta t) - \bar{I}(\Delta t)]^2 \quad (2.11)$$

where the overbars denote estimates. While $\bar{I}(\Delta t)$ is an unbiased estimate of the current, that is $\langle \bar{I}(\Delta t) \rangle \equiv I_{\text{eq}} = I_2 + \Delta I p_1^{\text{eq}}$, the estimator of the variance, $\bar{\sigma}_I^2(\Delta t)$, is unbiased only to terms of order $[N(k\Delta t)^2]^{-1}$. Because it has been assumed that both $k\Delta t$ and N are much greater than 1, this lower-order contribution will be neglected. Using the results of eqs 2.5 and the asymptotic behavior 2.9, we can write

$$p_1^{\text{eq}} = \frac{I_{\text{eq}} - I_2}{\Delta I} \quad (2.12)$$

and

$$k = \frac{2(I_1 - I_{\text{eq}})(I_{\text{eq}} - I_2)}{\sigma_I^2(\Delta t)\Delta t - D} \quad (2.13)$$

The individual rate constants can be expressed in terms of k and p_1^{eq} as

$$k_1 = (1 - p_1^{\text{eq}})k \quad k_2 = p_1^{\text{eq}}k \quad (2.14)$$

The replacement of I_{eq} and $\sigma_I^2(\Delta t)$ in eqs 2.12 and 2.13 by their estimates $\bar{I}(\Delta t)$ and $\bar{\sigma}_I^2(\Delta t)$ allows one to estimate k and p_1^{eq} and finally the k_i .

We now characterize the sampling errors in the estimates of p_1^{eq} and k due to the finite total number of observations. Consider first the error in the estimate of p_1^{eq} . Since $\bar{I}(\Delta t)$ is just the current sampled at the time $N\Delta t$, it follows that $\bar{I}(\Delta t) = I(N\Delta t)$. Therefore, the variance of $\bar{I}(\Delta t)$ is

$$\sigma_I^2 = \sigma_I^2(N\Delta t) \quad (2.15)$$

Combining this result with eq 2.6 we find an expression for the standard deviation of p_1^{eq} which we denote by $\sigma[\bar{p}_1^{\text{eq}}]$:

$$\sigma[\bar{p}_1^{\text{eq}}] = \sqrt{\sigma_x^2(N\Delta t) + \frac{D}{N\Delta t\Delta I^2}} \quad (2.16)$$

A somewhat more elaborate calculation of the standard deviation for the estimate \bar{k} yields the approximate expression¹²

$$\sigma(\bar{k}) \approx k \sqrt{\frac{2}{N}} \left\{ 1 + \frac{(\Delta I)^2 - 5(I_1 - I_{\text{eq}})(I_{\text{eq}} - I_2)}{2(I_1 - I_{\text{eq}})(I_{\text{eq}} - I_2)k\Delta t} + \sigma\left(\frac{1}{[k\Delta t]^2}\right) \right\} \quad (2.17)$$

To finish the theoretical analysis, we examine the question of determining whether the data are consistent with the assumption

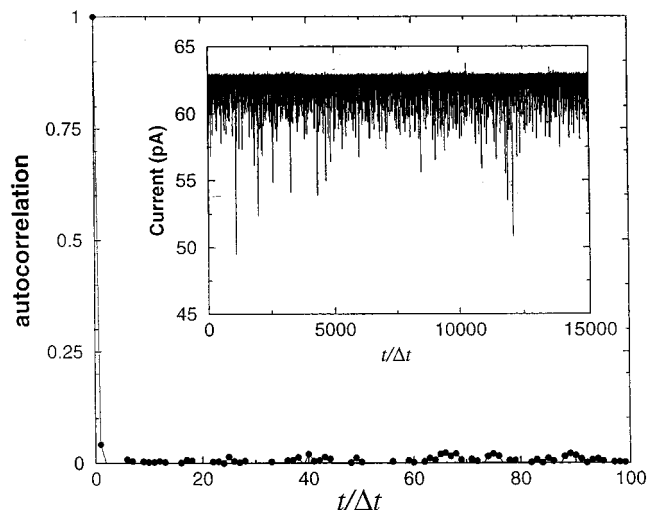


Figure 1. Set of experimental data $\{I_n(\Delta t)\}$ from the ion channel experiment (inset) sampled at $\Delta t = 30$ ms, together with the normalized autocorrelation function calculated from it. The point to notice is that the autocorrelation function drops to a value essentially equal to zero very rapidly, which would vitiate the use of a spectral analysis. The experimental data were first sampled at a rate of 50 kHz and then resampled over consecutive intervals lasting 30 ms.

that the system is Markovian or not.¹³ A solution for the probability density of the fraction of time that the system spends in one of the states of a two-state Markovian system has been derived in ref 10. Let $p(x|\Delta t)$ be the probability density for the fraction of time spent in state 1. This function has been shown to be

$$p(x|\Delta t) = p_1^{\text{eq}} e^{-k_1 \Delta t} \delta(x-1) + p_2^{\text{eq}} e^{-k_2 \Delta t} \delta(x) + \left\{ 2p_1^{\text{eq}} p_2^{\text{eq}} k \Delta t \times \right. \\ \left. I_0(y) + \frac{p_1^{\text{eq}} x + p_2^{\text{eq}} (1-x)}{2x(1-x)} y I_1(y) \right\} e^{-(p_2^{\text{eq}} x + p_1^{\text{eq}} (1-x)) k \Delta t} \quad (2.18)$$

where $y \equiv 2k\Delta t \sqrt{p_1^{\text{eq}} p_2^{\text{eq}} x(1-x)}$ and $I_0(y)$ and $I_1(y)$ are modified Bessel functions of order 0 and 1.¹⁴ Using $p(x|\Delta t)$ and taking into account the fact that the noise term in eq 2.2 represents Gaussian white noise, one can calculate the probability density for the random current $I(\Delta t)$ by

$$g(I|\Delta t) = \frac{\Delta t}{2\pi D} \int_0^1 p(x|\Delta t) e^{-(\Delta t/2D)(I-I_2-\Delta t x)^2} dx \quad (2.19)$$

The last step in the analysis, whether the system is Markovian or not, is to compare the experimental probability density obtained directly from the set of data with the theoretical result in this equation, replacing the rate constants by their estimates.

3. Analysis of Experimental Data

The SM experiments whose data have been analyzed are able to resolve single sugar molecule translocation through a single transmembrane protein channel.⁹ From these high-resolution measurements of the channel conductance it has been found that, at low sugar concentrations, sugar binding is a two-state Markovian process in which the channel is partly blocked when a single sugar molecule is found in the channel. Two experiments, differing in duration (463 and 163 s) were run, with results as described below.

In generating the data shown in the inset of Figure 1, we used a low sugar concentration (1 μM maltohexaose) to induce transient blocking of ionic conductance of a single maltoporin

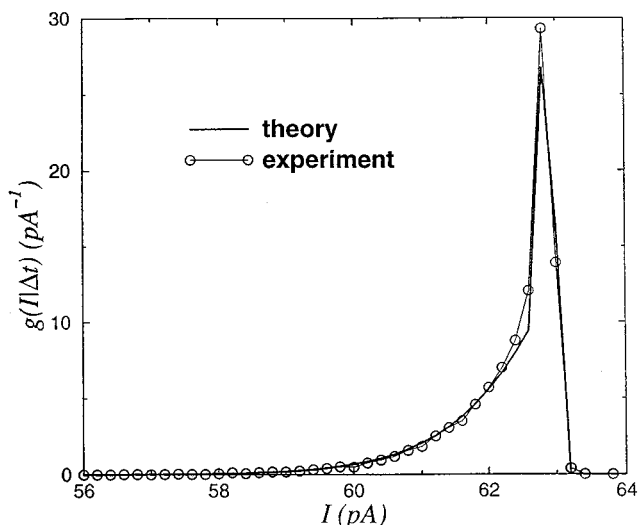


Figure 2. Comparison of the estimate of the probability density obtained from the data in Figure 1 from experiment 1 (463 s) with the theoretical expression in (2.19). The circles in this and the following figure are found from the data. The thin line is drawn through the centers of the circles. The thicker line is generated from the theoretical expression in (2.19) for the same values of the current used in representing experimental results.

TABLE 1: Rate Constants Found Using Data for Two Monitoring Times

	exp 1 ($T = 463$ s)	exp 2 ($T = 163$ s)
I_1 (pA)	62.90	60.67
I_2 (pA)	42.19	41.67
D (msec \times pA ²)	0.19	0.19
p_1^{eq}	0.970 ± 0.001	0.968 ± 0.002
k_1 (ms ⁻¹)	$(3.34 \pm 0.16) \times 10^{-2}$	$(3.07 \pm 0.25) \times 10^{-2}$
k_2 (ms ⁻¹)	1.07 ± 0.05	0.93 ± 0.07

channel reconstituted in a bilayer membrane formed from dioleoylphosphatidylcholine. The membrane bathing aqueous solution contained 1 M KCl, 1 mM CaCl₂ and 10 mM Tris at pH 7.4. Conductance recordings were obtained at a 200 mV membrane potential difference, positive from the side of protein addition. The original sampling frequency was 50 kHz, and the resampling to produce the data shown in Figure 1 was performed by data averaging over consecutive intervals of 30 ms.

We have created nearly completely uncorrelated data sets by resampling high-resolution recordings of channel conductance using a resolution time much longer than the characteristic correlation time of the system. Figure 1 shows the averaged data (in the inset) as well as the correlation function calculated from the raw data. The falloff in this function as a function of dimensionless time is evident. A comparison of the estimates obtained from two experiments is shown in the Table 1, which was produced using the method outlined above, where the error estimates have been calculated using (2.16) and (2.17). The terms after the \pm signs have been set equal to $3\bar{\sigma}$ since the distribution function is essentially Gaussian.

The two rate constants for both experiments are in good agreement with rate constants found by conventional spectral analysis applied to high resolution data.⁹ Using subsets of the two experiments taken with high-resolution data of 10 s duration, we found $k_1 = (3.38 \pm 0.44) \times 10^{-2}$ ms⁻¹ and $k_2 = 1.17 \pm 0.15$ ms⁻¹ for the longer experiment. For the shorter experiment the rate constants were found to be $k_1 = (3.56 \pm 0.48) \times 10^{-2}$ ms⁻¹ and $k_2 = 1.15 \pm 0.16$ ms⁻¹. We used the parameters in Table 1 to calculate the probability density of the current for two experiments according to eq 2.19. Figures 2 and 3 compare

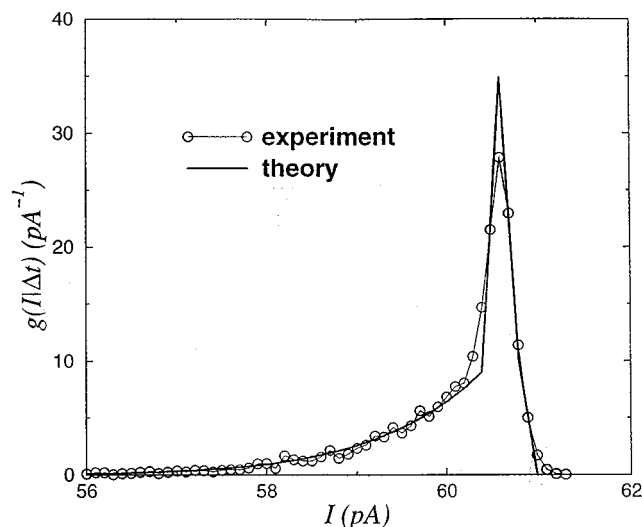


Figure 3. Same comparison for experiment 2 (163 s). In both cases the agreement is seen to be quite good. Notice that the probability density is still quite far from the asymptotic Gaussian limit.

the experimental and theoretical probability densities, showing good agreement between theory and experiment.

4. A Test of the Markov Property

In the previous section we found that our suggested procedure works well in estimating the rate constants for a two-state Markov system from nearly uncorrelated data. We also found good agreement between the predicted probability density for the current with the density estimated from the data. This suggests the further problem of whether a lack of agreement between data and the theoretical result found from eq 2.19 can be used to distinguish between a two-state Markov system and a two-state non-Markov system. The latter case indicates a more complicated reaction scheme than that in eq 2.1.¹⁵ This possibility was tested using simulated data. Accordingly, a set of data, $\{I_n(\Delta t)\}$, was generated using most of the same parameters as the experimentally determined ones summarized in Table 1. In generating the data, we inserted the same resolution time, the same observation time, the same intensities I_1 and I_2 , and the same intensity of the white noise term. We also assumed that the kinetics of the reaction $1 \rightarrow 2$ was first-order with the same value of k_1 as in Table 1 for the longer experiment. For the reaction $2 \rightarrow 1$ we used three different non-Markovian probability densities, $\psi_2(t)$, having the common property that the mean lifetime is equal to k_2^{-1} where k_2 is the rate constant given in Table 1 for the longer experiment. The three densities used in the simulations are

$$\begin{aligned}\psi_2^{(a)}(t) &= \frac{3}{8}k_2 \left[\exp\left(-\frac{3k_2}{4}t\right) + 2 \exp\left(-\frac{3k_2}{2}t\right) \right] \\ \psi_2^{(b)}(t) &= \frac{\pi}{2}k_2^2 t \exp\left(-\frac{\pi k_2^2}{4}t^2\right) \\ \psi_2^{(c)}(t) &= \sqrt{\frac{k_2}{2t}} \exp(-\sqrt{2k_2}t) \quad (4.2)\end{aligned}$$

For each of the densities we generated a set of currents $\{I_n(\Delta t)\}$ following which we applied the procedure described in section 2 to derive two rate constants under the assumption that the system is Markovian. These rate constants were then used to calculate the probability density $g(I|\Delta t)$ according to the relation in eq 2.19. To check whether the system is

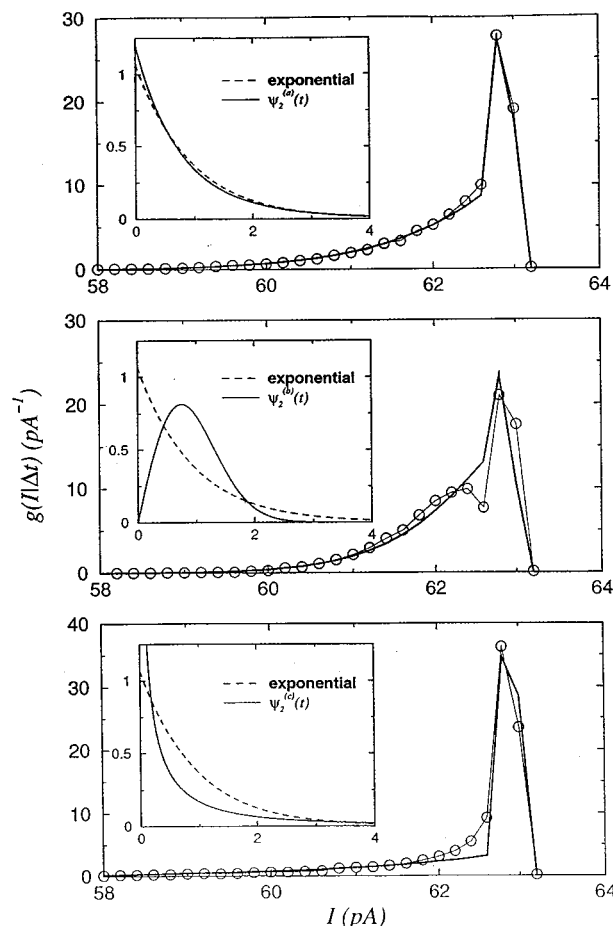


Figure 4. Comparison of estimates of the probability density obtained by using the three forms of $\psi_2^{(i)}(t)$, $i = a, b, c$ in (4.2). The density found from the data is given by the circles, while the solid line is the density calculated from eq 2.19. The first of these produces an estimate quite close to that found by using eq 2.19 and does not suggest that one could distinguish between these results and those produced by a true two-state Markov system, while the remaining two do suggest a deviation from this type of system. The three probability densities in eq 4.2 are shown in the insets plotted as a function of the time in milliseconds.

Markovian or not, we compared $g(I|\Delta t)$ with the corresponding histogram obtained from the data set $\{I_n(\Delta t)\}$. The simulations produced results shown in Figure 4, where it is seen that there are noticeable discrepancies from the prediction based on eq 2.19 for $\psi_2^{(b)}(t)$ and $\psi_2^{(c)}(t)$ whereas for $\psi_2^{(a)}(t)$ the results are relatively close to the theoretical prediction in the single exponential case. The curves in Figure 4 make the obvious point that when the functional form for $\psi_2(t)$ is not close to a negative exponential (cases b and c) it is possible to distinguish between Markovian and non-Markovian systems.

5. Conclusions

In this paper we have presented a method for obtaining rate constants of a two-state Markovian system from SM type experiments when the resolution is so bad that the time correlations are lost. We have shown that in such situations it is still possible to get the rate constants with good accuracy even when the data set of time averages is randomized. Moreover, in some cases, it is possible to distinguish between Markovian and non-Markovian systems.

Acknowledgment. L.K. gratefully acknowledges support by the Swedish Foundation for International Cooperation in Research and Higher Education.

References and Notes

- (1) Moerner, W. E., Ed. *Acc. Chem. Res.* **1996**, *29*, 561–613 (a special issue entirely devoted to SMS).
- (2) Nie, S.; Zare, R. N. *Annu. Rev. Biophys. Struct.* **1997**, *26*, 567–597.
- (3) Xie, X. S.; Trautman, J. K. *Annu. Rev. Phys. Chem.* **1998**, *49*, 441–480.
- (4) *Science* **1999**, *283*, No. 5408, 1667–1695 (contains a collection of articles on SMS).
- (5) *Chem. Phys.* **1999**, *247*, 1–189 (a special issue entirely devoted to SMS).
- (6) Bezrukov, S. M. *J. Membr. Biol.* **2000**, *174*, 1–13.
- (7) Neher, E.; Stevens, C. F. *Annu. Rev. Biophys. Bioeng.* **1977**, *49*, 345–381.
- (8) DeFelice, L. J. *Introduction to Membrane Noise*; Plenum Press: New York, 1981.
- (9) Bezrukov, S. M.; Kullman, L.; Winterhalter, M. *FEBS Lett.* **2000**, *476*, 224–228.
- (10) Berezhkovskii, A. M.; Szabo, A.; Weiss, G. H. *J. Chem. Phys.* **1999**, *110*, 9145–9150.
- (11) Boguñá, M.; Berezhkovskii, M.; Weiss, G. H. *Physica A* **2000**, *282*, 475–485.
- (12) Boguñá, M.; Berezhkovskii, M.; Weiss, G. H. To appear in *J. Phys. Chem.*
- (13) Even when the system is non-Markovian, it is possible to calculate the mean relaxation time of the decay of equilibrium fluctuations, τ , from moments of the single sojourn times in each of the states. Let t_i be a single sojourn time in state i . It has been shown in ref 11 that τ is

$$\tau = \frac{\sigma_1^2 \langle t_2 \rangle^2 + \sigma_2^2 \langle t_1 \rangle^2}{2 \langle t_1 \rangle \langle t_2 \rangle (\langle t_1 \rangle + \langle t_2 \rangle)}$$

in which $\sigma_i^2 = \langle t_i^2 \rangle - \langle t_i \rangle^2$. When the system is Markovian one finds that $\tau = (k_1 + k_2)^{-1} = k^{-1}$.

(14) Abramowitz, M.; Stegun, I. A. *Handbook of Mathematical Functions*; Dover: New York, 1972.

(15) Boguñá, M.; Weiss, G. H. *Physica A* **2000**, *282*, 486–494.