Estimating Single-Channel Kinetic Parameters from Idealized Patch-Clamp Data Containing Missed Events

Feng Qin, Anthony Auerbach and Frederick Sachs
Department of Biophysical Sciences, State University of New York at Buffalo, Buffalo, New York 14214 USA

ABSTRACT We present here a maximal likelihood algorithm for estimating single-channel kinetic parameters from idealized patch-clamp data. The algorithm takes into account missed events caused by limited time resolution of the recording system. Assuming a fixed dead time, we derive an explicit expression for the corrected transition rate matrix by generalizing the theory of Roux and Sauve (1985, *Biophys. J.* 48:149–158) to the case of multiple conductance levels. We use a variable metric optimizer with analytical derivatives for rapidly maximizing the likelihood. The algorithm is applicable to data containing substates and multiple identical or nonidentical channels. It allows multiple data sets obtained under different experimental conditions, e.g., concentration, voltage, and force, to be fit simultaneously. It also permits a variety of constraints on rate constants and provides standard errors for all estimates of model parameters. The algorithm has been tested extensively on a variety of kinetic models with both simulated and experimental data. It is very efficient and robust; rate constants for a multistate model can often be extracted in a processing time of approximately 1 min, largely independent of the starting values.

INTRODUCTION

Ion channels have been proposed to exist in a finite number of discrete conformation states, with transitions between states governed by a first-order Markov process (Colquhoun and Hawkes, 1983). In certain conformations, selected ions can flow through the channel, and in other conformations the channel is closed. With the technique of patch-clamp recording, we are able to record the current flowing through a single channel (Neher and Sakmann, 1992). However, these recordings provide only a reduced view of the time course of the underlying conformational changes because of the aggregation of different kinetic states into the same conductance levels. A central goal of single-channel kinetic analysis is to provide a complete description of the gating mechanism of the ion channel in question from the recordings, including estimates of all transition rates.

The traditional technique for single-channel kinetic analysis is the fitting of lifetime histograms, in which dwell times of a channel in a given conductance level are plotted in bins on the abscissa and the number of events observed in each bin is plotted on the ordinate (Colquhoun and Sigworth, 1983). The histogram, when scaled properly, is an estimate of the probability density for the lifetime at that conductance level. For Markov models, the lifetime distributions are sums of exponentials with as many components as states at a given conductance level. Thus, certain kinetic information can be derived by fitting the observed dwell time distribution directly with sums of exponentials or via the predicted distribution from a specific model. Because the cross-correlation information between adjacent events is not included in such one-dimensional data, the histogram fitting techniques can resolve only a maximal number of $2(N - 1)$ kinetic parameters where N is the number of states. This is generally inadequate for complex models with loops. Two-dimensional histogram fitting has been used to account for correlation between adjacent durations (Fredkin et al., 1985; Magleby and Weiss, 1990), but the method has been found to be less accurate than the likelihood procedure (Ball and Sansom, 1989). Perhaps more importantly, histogram methods cannot be easily applied to nonstationary data such as that resulting from inactivating or adapting channels.

To alleviate the inherent limitations of histogram fitting techniques, Horn and Lange (1983) introduced the use of a full maximal likelihood approach to estimate single-channel kinetic parameters. Given a kinetic model, the parameters are determined to maximize the probability of the observed samples. In their method, the computation of the likelihood is carried out recursively over individual samples, and the search of the likelihood surface is accomplished by a nonlinear regressive method where the required partial derivatives of the likelihood function are calculated numerically. Because the probability at every data point needs to be evaluated, the method requires considerable computing time and a large memory space. Thus, it is applicable only for the patches having a small number of channels, few transition rates, and a limited number of experimental runs. To improve speed, several authors including Horn and Lange have proposed computing the likelihood over dwell times rather than data points (Horn and Lange, 1983; Chay, 1988; Ball and Sansom, 1989). It was shown that the likelihood of a sequence of dwell times may be obtained by accumulating the transition density matrices for individual dwell times (Fredkin et al., 1985). Because there are many fewer dwell

Received for publication 13 July 1995 and in final form 27 September 1995.

Address reprint requests to Dr. Feng Qin, 120 Cary Hall, SUNY at Buffalo, Buffalo, NY 14214. Tel.: 716-829-3289; Fax: 716-829-2028; E-mail: qin@acsu.buffalo.edu.

0006-3495/96/01/264/17 $2.00

times than there are samples, the resultant algorithms have significantly reduced computation requirements.

Inherent in methods based on dwell time analysis is the issue of missed events. Because of limited time resolution of the recording system, very short transitions cannot be detected. The missing dwell times will lead to overestimation of apparent dwell times; for example, consecutive openings separated by undetected closures will appear as a single opening. A great deal of work has been done in an attempt to predict the effects of missed events on observed channel lifetime distributions (see, for example, Sachs et al., 1982; Roux and Sauve, 1985; Blatz and Magleby, 1986a; Ball and Sansom, 1988; Milne et al., 1988; Crouzy and Sigworth, 1990; Hawkes et al., 1992; Magleby and Weiss, 1990). Unfortunately, all results are restricted to the case of binary conductances. Furthermore, no attempt has been made to analyze the effects of missed events on the likelihood function.

In this paper, we present a general approach for estimating single-channel kinetic parameters from idealized patch-clamp data. The approach employs the joint probability density of the idealized dwell time series as the likelihood and makes use of analytically calculated derivatives for maximizing the likelihood. It has the explicit capacity of dealing with multiple conductance levels and multiple channels. It takes into account the effects of missed events. Assuming a fixed dead time, we derive the explicit expression for the corrected transition rate matrix by generalizing the theory of Roux and Sauve (1985) to the case of multiple conductance levels. The method allows data sets obtained under different experimental conditions, e.g., concentration, voltage, and force, to be fit as a single data set. It also permits typical constraints on rate constants, e.g., detailed balance constraints. The method has been tested extensively on a variety of kinetic models with both simulated and actual experimental data. It is shown to be exceedingly efficient; e.g., a set of rate constants for a multistate model can often be extracted from a few seconds of real time data in a processing time of less than 1 min.

The paper is structured as follows. The following section presents a brief review of the maximal likelihood estimation of an aggregated Markov process without missed events. The third section describes how to correct the likelihood method for the effects of missed events. The fourth section discusses some practical issues encountered in single-channel data analysis, such as global fitting, constraints, and multiple channels. Simulation and testing of the method are presented in the fifth section. Finally, some discussions about the advantages and disadvantages of the method are given in the last section.

THE LIKELIHOOD FUNCTION: ITS EVALUATION AND MAXIMIZATION

We consider a channel with N kinetic states. The kinetics of the channel is modeled as a continuous-time Markov process. The rate constants for transitions between states i and j, $i \neq j$, are the elements, q_{ij}, of the transition rate matrix Q. The elements of Q have the dimension of reciprocal time, and the diagonal elements, q_{ii}, are defined so that the row sums to zero; thus $-1/q_{ii}$ is the mean lifetime of a dwell in state i.

Patch-clamp data report the current going through a channel at a given moment. Thus, the conductances of different kinetic states determine the distinguishability of kinetic states. Suppose the channel has M ($M \leq N$) conductances. Throughout the paper, we denote a conductance, or equivalently, the set of the kinetic states in a conductance, by the letters a, b, c, etc. Suppose there are n_a kinetic states in conductance a. Partition the Q matrix into submatrices, Q_{ab}, each of which are of size n_a by n_b and collect all transition rates between the states in conductance a and b. Then the transitions of the channel from one conductance to another in a given duration can be described by a probability density matrix (Colquhoun and Hawkes, 1981)

$$G_{ab}(t) = \exp(Q_{aa}t)Q_{ab}, \tag{1}$$

where the (i, j)th element represents the probability density that the channel enters conductance a from its ith state, stays in a throughout the time interval $[0, t]$, and then exits into the jth state of conductance b.

A set of patch-clamp data can be described as a series of dwell times during which a channel is in a given conductance. Let $t = t_1, t_2, \cdots t_L$ be the list of time intervals and $a = a_1, a_2, \cdots a_L$ the list of conductances. Let θ designate the set of the unknown parameters of the channel. Then the likelihood of obtaining t and a given the model θ can be formulated as (Fredkin et al., 1985)

$$f(\theta) = \pi_{a_1} \prod_{j=1}^{L} G_{a_j a_{j+1}}(t_j)\mathbf{1}, \tag{2}$$

where π_a is a row vector of the entry probabilities into the states in conductance a, a_{L+1} refers to the subset of the states that do not belong to a_L, and the $\mathbf{1}$ denotes a column vector of ones. In other words, the likelihood is obtained by the entry probability vector, multiplied by the probability density matrix for the first dwell time, multiplied by the probability density matrix for the second dwell time, and so on.

Maximization of $f(\theta)$ with respect to θ yields the maximal likelihood estimates of the kinetic parameters of the channel for a given data set. We have recently developed an efficient procedure for the maximization of $f(\theta)$ (F. Qin, A. Auerbach, and F. Sachs, submitted for publication). One of the most attractive features of the method is that it exploits analytically calculated derivatives for search of the likelihood surface. In the rest of this section, we present a brief outline of the procedure.

Central to the method is a forward-backward recursive procedure, as follows:

$$\alpha_k^\tau = \alpha_{k-1}^\tau G_{a_k a_{k+1}}(t_k), \quad k = 1, 2, \cdots, L \tag{3}$$

$$\beta_k = G_{a_k a_{k+1}}(t_k)\beta_{k+1}, \quad k = L, L-1, \cdots, 1, \tag{4}$$

where the initial conditions are $\alpha_0^\tau = \pi_{a1}$ and $\beta_{L+1} = \mathbf{1}$. The forward vector α_k is essentially a vector of the probability densities of the partial dwell time series from the beginning to the kth dwell time and then leaving for the states in conductance a_{k+1}. Similarly, the backward vector β_k represents the probability densities of the partial dwell time series from the kth dwell to the end given the starting states being in conductance a_k.

The probability density matrix $G_{ab}(t)$ involved in the forward and backward recursions can be efficiently calculated by using the spectral decomposition of the matrix Q_{aa} (Colquhoun and Hawkes, 1977):

$$G_{ab}(t) = \left[\sum_{i=1}^{n_a} A_{a,i}\exp(\lambda_{a,i}t)\right]Q_{ab}, \tag{5}$$

where $\lambda_{a,i}$ is the ith eigenvalue of Q_{aa} and $A_{a,i}$ is a matrix determined from the corresponding left and right eigenvectors. It should be noticed that the quantities $\lambda_{a,i}$ and $A_{a,i}$ do not vary with t and thus need to be evaluated only once for a given θ.

From the forward and backward vectors as well as the spectral decomposition results, it is possible to determine the likelihood and its partial derivatives analytically. The likelihood is simply the sum of the components of the final forward vector α_L; i.e.,

$$f(\theta) = \alpha_L^\tau \mathbf{1}. \tag{6}$$

The derivatives of the likelihood function with respect to the elements in the Q matrix are given by

$$\frac{\partial f}{\partial Q_{ab}} = \sum_{i=1}^{n_a} A_{a,i}\left[\sum_{\substack{k=1 \\ a_k=a \\ a_{k+1}=b}}^{L} \alpha_{k-1}\beta_{k+1}^\tau \exp(\lambda_{a,i}t_k)\right] \tag{7}$$

$$\frac{\partial f}{\partial Q_{aa}} = \sum_{i=1}^{n_a}\sum_{j=1}^{n_a} A_{a,i}^\tau\left[\sum_{\substack{k=1 \\ a_k=a}}^{L} \alpha_{k-1}\beta_{k+1}^\tau \gamma_{a,i,j}(t_k)\right]A_{a,j}^\tau, \tag{8}$$

where $\gamma_{a,i,j}(t)$ is a scalar function of t defined by

$$\gamma_{a,i,j}(t) = \begin{cases} te^{\lambda_{a,i}t} & i = j \\ \dfrac{e^{\lambda_{a,i}t} - e^{\lambda_{a,j}t}}{\lambda_{a,i} - \lambda_{a,j}} & i \neq j. \end{cases} \tag{9}$$

Maximization of the likelihood is accomplished by the variable metric method of David, which uses the Fletcher-Powell approximation to the inverse of the Hessian matrix of second derivatives and an exact line search with adaptive step sizes (Fletcher, 1981). This optimizer has several advantages. First, it converges rapidly, and the convergence becomes quadratic when the parameters are sufficiently close to the maximum point. Secondly, a programmed version of the method exists in most numerical libraries, making it almost universally available. Finally, it has been successfully applied to solve a wide range of problems, demonstrating its practicability and reliability.

The standard errors for the parameter estimates are obtained from the curvature of the likelihood surface at its maximum (Cox and Miller, 1965):

$$\mathrm{Cov}(\theta) \approx -H^{-1}(\theta), \tag{10}$$

where Cov is the covariance matrix, and H is the Hessian matrix of second derivatives. The exact Hessian matrix may be calculated numerically but involves intensive computing. A faster, although less accurate, way is to directly use the approximate inverse Hessian matrix generated by the optimizer. This is another reason why the variable metric method is chosen for optimization.

Finally, we notice that the forward and backward recursions in Eqs. 3 and 4 may run into numerical problems. Because the probability density matrix for a dwell time often has elements less than one, the forward vectors usually decay with increasing k and the backward vectors with decreasing k. For sufficiently large k, the dynamic range of these vectors will exceed the machine precision range. Fortunately, this problem can be circumvented by making use of an ongoing scaling procedure.

MISSED EVENT CORRECTION

In practice, very short transitions may not be detected as a result of limited time resolution of the recording system. The missed events result in distortion in the dwell time likelihood as an observed dwell time in a conductance level may actually contain transitions into other conductance levels. In this section, we describe how to correct the likelihood for such distortion. In particular, we show that an approximate correction may be made by appropriately modifying the transition rate matrix of the channel. We also show how to efficiently calculate the corrected transition rate matrix as well as its sensitivities to the original transition rate matrix.

The corrected Q matrix

We make the usual assumption that the missed events can be characterized by a constant dead time, t_d, such that all dwell times shorter than t_d are not detected, but all longer dwell times are detected. It should be emphasized that the specification of a dead time depends on the detection procedures. If noise-free data are idealized via threshold crossing (Sachs et al., 1982), then t_d is the event duration that just reaches the threshold and can be calculated precisely from

the system response properties. However, the noise in the record makes the t_d value imprecise because some events that are actually shorter than t_d will cross the threshold whereas others that are longer than t_d will go undetected. If the data are idealized using hidden Markov modeling (Chung et al., 1990; Fredkin and Rice, 1992b) or other nonthreshold techniques (Draber and Schultze, 1994), the dead time again cannot be precisely defined. Given these ambiguities, and to satisfy the assumption of an absolute dead time that is required for the derivation below, we enforce a fixed dead time by concatenating any dwell time below t_d.

If the channel has a buzz mode with many short transitions in a row, two problems occur: (1) idealization of band-limited data becomes difficult and (2) the approximation used below, i.e., the total duration of missed events is small compared with the observed duration, becomes invalid. Thus, the method presented here is not accurate for such data and other approaches such as the point likelihood methods (Walsh and Sigworth, 1992; Fredkin and Rice, 1992a; Qin et al., 1994; Albertsen and Hansen, 1994) and the iterative simulation method (Magleby and Weiss, 1990) are preferable.

A general theory for analyzing the effect of missed events on channel dwell time distributions was first established by Roux and Sauve (1985). Here we extend the theory to correct the effects of missed events on the likelihood function and allow for multiple conductance levels. From Eq. 2 we see that the likelihood of a series of dwell times is determined by the probability density matrix $G_{ab}(t)$ of individual dwell times. Consequently, the key to correct the likelihood is to correct $G_{ab}(t)$ for missed events. Below we denote by ${}^eG_{ab}(t)$ the corrected probability density matrix of an apparent dwell time, i.e., the matrix of the probability densities that the channel stays in conductance a for a duration t without any observable leaving and then exits into conductance b.

The corrected probability density matrix ${}^eG_{ab}(t)$ is essentially the sum of the probability density matrices of all possible transition paths that yield the observed dwell time. Thus, to derive ${}^eG_{ab}(t)$, we need to implement all these paths by including the underlying missed transitions. For a path containing n undetected leavings, it may be implemented in two exclusive ways, as shown in Fig. 1, one with all n leavings occurring inside the observed dwell time and the other with $n - 1$ leavings occurring inside and one leaving at the end of the dwell time. In each implementation, the first stay in a must be longer than t_d so that the entire dwell time is observed, and all leavings from a must be shorter than t_d so that they are not detected. In other words, the implementations must satisfy

$$\sum \tau_i = t \qquad \tau_1 \geq t_d \qquad \tau_{2i} \leq t_d, \quad 1 \leq i \leq n. \tag{11}$$

The probability density matrices for such implementations are respectively given by

$$\left[\prod_{i=1}^{n} G_{a\bar{a}}(\tau_{2i-1}) G_{\bar{a}a}(\tau_{2i})\right] G_{ab}(\tau_{2n+1}) \tag{12}$$

and

$$\left[\prod_{i=1}^{n-1} G_{a\bar{a}}(\tau_{2i-1}) G_{\bar{a}a}(\tau_{2i})\right] G_{ac}(\tau_{2n-1}) G_{cb}(\tau_{2n}). \tag{13}$$

Thus, ${}^eG_{ab}(t)$ can be obtained by integrating Eqs. 12 and 13 over all possible τ_j subject to the constraints and then summing the results over all $n \geq 0$. The final result may be stated as (see Appendix for the detailed derivation)

$$\begin{aligned} {}^eG_{ab}(t) = -\frac{1}{2\pi} e^{Q_{aa}t_d} \int_{-\infty}^{\infty} e^{-j\omega(t-t_d)} \\ \cdot [j\omega I + Q_{aa} - Q_{a\bar{a}}(I - e^{(j\omega I + Q_{\bar{a}\bar{a}})t_d})(j\omega I + Q_{\bar{a}\bar{a}})^{-1} Q_{\bar{a}a}]^{-1} \\ \cdot [Q_{ab} - Q_{ac}(I - e^{(j\omega I + Q_{cc})t_d})(j\omega I + Q_{cc})^{-1} Q_{cb}]\, d\omega, \end{aligned} \tag{14}$$

where $Q_{a\bar{a}}$, $Q_{\bar{a}a}$ and $Q_{\bar{a}\bar{a}}$ are the submatrices of Q partitioned in an obvious manner.

Equation 14 provides the exact solution to the corrected probability density matrix of an apparent dwell time. However, it involves the integral of inversion of a complex matrix function, making computation intractable. A more useful solution is the so-called first-order approximation, obtained by ignoring the contribution of the durations of missed events to the total duration of an observed dwell time (Roux and Sauve, 1985), as follows (see Appendix):

$${}^eG_{ab}(t) = \exp({}^eQ_{aa}t)\,{}^eQ_{ab}, \tag{15}$$

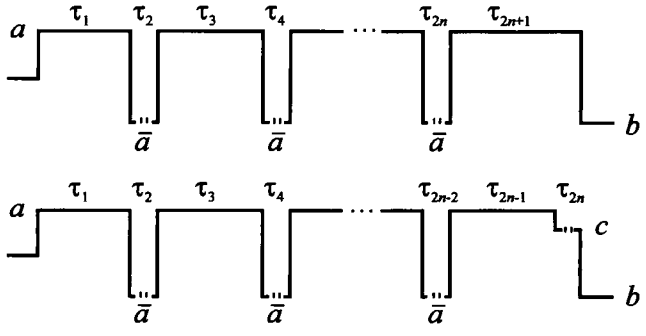


FIGURE 1 An apparent dwell time containing n undetected leavings may be implemented in two ways, one with all leavings occurring inside the dwell and the other with $n - 1$ leavings inside and one leaving at the end of the dwell. Here, $\bar{a}$ refers to the set of the kinetic states that do not belong to conductance a, and $c = \bar{a} \cap \bar{b}$. For binary channels, $\bar{a} = b$, and $c = \varnothing$; thus, only the first implementation is possible and the second one does not exist.

where ${}^eQ_{aa}$ and ${}^eQ_{ab}$ are defined as

$$ {}^eQ_{aa} = Q_{aa} - Q_{a\bar{a}}(I - e^{Q_{\bar{a}\bar{a}}t_d})Q_{\bar{a}\bar{a}}^{-1}Q_{\bar{a}a} \tag{16} $$

$$ {}^eQ_{ab} = \exp[t_dQ_{a\bar{a}}(I - e^{Q_{\bar{a}\bar{a}}t_d})Q_{\bar{a}\bar{a}}^{-1}Q_{\bar{a}a}] \times [Q_{ab} - Q_{ac}(I - e^{Q_{cc}t_d})Q_{cc}^{-1}Q_{cb}]. \tag{17} $$

Equation 15 has the same form as in the case of no missed events. This suggests that we can look at the observed dwell times as if generated from another aggregated Markov process with the transition rate matrix given by ${}^eQ = [{}^eQ_{ab}]$. In other words, the effects of missed events may be corrected by simply using the ${}^eQ$ matrix to calculate the likelihood. Below we consider some details on the calculations of ${}^eQ$ as well as its sensitivities to Q.

Computation of ${}^eQ$ and $\partial {}^eQ/\partial q_{ij}$

Inspection of Eqs. 16 and 17 shows that there are two major difficulties involved in the calculations of ${}^eQ$ and $\partial {}^eQ/\partial q_{ij}$. One is how to calculate a matrix exponential and its sensitivities with respect to the original matrix. The other is how to calculate the following matrix:

$$ R_{aa} = [I - \exp(Q_{aa}t)]Q_{aa}^{-1} \tag{18} $$

as well as its sensitivities to Q_{aa}. Once these two problems are solved, the overall calculations of ${}^eQ$ and $\partial {}^eQ/\partial q_{ij}$ become just plain matrix additions and multiplications.

The first problem can be solved by using an efficient recursive procedure based on the series expansion of a matrix exponential (Qin et al., 1994). Here we show that this procedure can also be extended to solve the second problem.

We expand the matrix exponential in Eq. 18 into its power series to eliminate the matrix inversion, yielding

$$ R_{aa} = \sum_{k=0}^{n} \frac{t^k}{(k+1)!} Q_{aa}^k, \tag{19} $$

where the value of n is determined from a prescribed truncation error. By differentiating the power series with respect to the (i, j)th element of Q_{aa}, we derive

$$ \frac{\partial R_{aa}}{\partial q_{a,i,j}} = \sum_{k=0}^{n} \frac{t^k}{(k+1)!} [Q_{aa}^{k-1}E_{ij} + Q_{aa}^{k-2}E_{ij}Q_{aa} + \cdots + E_{ij}Q_{aa}^{k-1}], \tag{20} $$

where E_{ij} is a constant matrix in which only the (i, j)th element is one and all other elements are zero.

Equations 19 and 20 provide the theoretical formulae for R_{aa} and its sensitivities to Q_{aa}. The practical calculations of these formulae may be carried out in a recursive manner. Let U_k and V_k be the kth terms of the series in Eqs. 19 and 20, respectively. Then, we can solve for U_k and V_k inductively by

$$ U_{k+1} = \frac{t}{k+1} U_kQ_{aa} \tag{21} $$

$$ V_{k+1} = \frac{t}{k+1} [Q_{aa}V_k + E_{ij}U_k], \tag{22} $$

where the initial conditions $U_0 = I$ and $V_0 = 0$. The required matrix R_{aa} and its sensitivities $\partial R_{aa}/\partial q_{a,i,j}$ are simply the sums of U_k and V_k, respectively.

After the sensitivities $\partial {}^eQ/\partial q_{ij}$ are obtained, we can then calculate the derivatives of the corrected likelihood function with respect to the original Q matrix, simply by combining $\partial L/\partial {}^eQ$ and $\partial {}^eQ/\partial q_{ij}$ using the chain rule.

GLOBAL FITTING, CONSTRAINTS, AND MULTIPLE CHANNELS

We have seen that the likelihood method may be corrected for the effects of missed events by appropriately modifying the Q matrix. In this section we further consider some practical issues often encountered in analysis of single-channel data and show how to extend the likelihood method to account for these issues. In particular, we show how to apply this method to global fitting, i.e., combining data sets obtained at different experimental conditions (Ball and Sansom, 1989; Vandenberg and Bezanilla, 1991; McManus and Magleby, 1991; Weiss and Magleby, 1992; Auerbach, 1993). We also show how to deal with typical constraints imposed on rate constants and how to analyze data containing the activity of multiple identical or nonidentical channels.

Global fitting

To allow global fitting, we need to identify the intrinsic parameters of a channel that do not vary with experimental conditions and then formulate the rate constants of the channel in terms of these intrinsic parameters. The rate constants of a biological channel are scaled by the experimental variables in two ways. For voltage-gated channels, the rates typically depend on voltage in an exponential manner. The same is true for force in treating mechanosensitive channels. For ligand-gated channels, the rates change linearly with ligand concentration. A unified representation of such dependencies may be written as

$$ q_{ij} = C_{ij}\exp(\mu_{ij} + \nu_{ij}V), \tag{23} $$

where V is the voltage and C_{ij} is the normalized concentration of the ligand to which the rate q_{ij} is sensitive. Obviously, we can simply set $C_{ij} = 1$ for the rates that are not sensitive to any ligand, and set $\nu_{ij} = 0$ for the rates that are independent of voltage. The parameters μ_{ij} and ν_{ij} usually

do not vary with the experimental conditions. They are intrinsic parameters of a channel.

In addition to allowing global fitting, the above representation also offers several computational advantages. First, the rate constants are automatically restricted to be nonnegative. Second, the constraint of detailed balance, which is highly nonlinear in the domain of q_{ij}, is now reduced to be linear in the domain of μ_{ij} and ν_{ij}. As described below, this property greatly facilitates the handling of detailed balance constraints. Third, as suggested by numerical examples, using μ_{ij} and ν_{ij} as free parameters often results in a more symmetrical likelihood surface, which helps optimization.

Constraints

Typical constraints imposed on rate constants include holding some rate constants at fixed values, linear scaling by other rate constants, and detailed balance conditions. It is easy to see that all of these constraints are essentially linear on the parameters μ_{ij} and ν_{ij}. Consequently, we can cast them into the form

$$\Gamma\theta = \xi, \tag{24}$$

where $\theta = (\cdots, \mu_{ij}, \nu_{ij}, \cdots)^\tau$, Γ is a coefficient matrix, and ξ is a constant vector. For simplicity, we assume that the imposed constraints are independent of each other; that is, the Γ matrix has a full rank.

For linearly constrained variables, it is possible to formulate them into linear combinations of a set of unconstrained variables. This is accomplished by the *QR* decomposition of the coefficient matrix (Golub and Van Loan, 1989). Let $\Gamma = UR$ be the *QR* factorization of Γ, where U is an orthogonal matrix and R is an upper triangular matrix. Partition R into

$$R = [R_1, R_2]$$

and θ into

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}.$$

Then Eq. 24 can be written as

$$\theta_1 = -R_1^{-1}R_2\theta_2 + R_1^{-1}U^{-1}\xi.$$

Let $x = \theta_2$; then we can restructure the above equation into the following form:

$$\theta = Ax + b \tag{25}$$

where

$$A = \begin{bmatrix} -R_1^{-1}R_2 \\ I \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} R_1^{-1}U^{-1}\xi \\ 0 \end{bmatrix}.$$

The significance of Eq. 25 is that it allows the constraints imposed on rate constants to be eliminated by choosing x as the variables to be optimized. The problem of maximizing the likelihood subject to constraints (Eq. 24) is thus reduced to an equivalent unconstrained maximization problem.

Conversion of the x variables to rate constants is straightforward from Eqs. 25 and 23. Furthermore, the derivatives of the likelihood with respect to x can be obtained by combining the derivatives $\partial f/\partial q_{ij}$ and $\partial q_{ij}/\partial x_k$ with the chain rule. The derivative $\partial q_{ij}/\partial x_k$ is given by

$$\begin{aligned} \frac{\partial q_{ij}}{\partial x_k} &= \frac{\partial q_{ij}}{\partial \theta_{2l-1}} \cdot \frac{\partial \theta_{2l-1}}{\partial x_k} + \frac{\partial q_{ij}}{\partial \theta_{2l}} \cdot \frac{\partial \theta_{2l}}{\partial x_k} \\ &= q_{ij}(a_{2l-1,k} + Va_{2l,k}), \end{aligned} \tag{26}$$

where we assume $\mu_{ij} = \theta_{2l-1}$ and $\nu_{ij} = \theta_{2l}$.

At the end of optimization, it is necessary to convert the error estimates on x variables back to the error estimates on rate constants. This can be accomplished by making use of the following relation (Kendall and Stuart, 1977):

$$\mathrm{Var}(q_{ij}) \approx \sum_k \left(\frac{\partial q_{ij}}{\partial x_k}\right)^2 \mathrm{Var}(x_k), \tag{27}$$

where Var denotes variance and $\mathrm{Var}(x_k)$ corresponds to the diagonal elements of the covariance matrix given in Eq. 10.

Multiple channels

Often, there will be more than one channel in a patch of membrane from which the recording is made. Suppose the patch contains n channels and the ith individual channel has k_i kinetic states. The patch at any time can be fully characterized by a vector

$$s_t = (s_t^{(1)}, s_t^{(2)}, \cdots, s_t^{(n)}), \tag{28}$$

where $s_t^{(i)}$ specifies which state the ith channel is in at time t. If the constituent channels are statistically independent, then the Markov models for $s_t^{(i)}$, $i = 1, 2, \cdots, n$ induces a Markov model for s_t. The state space of s_t has a dimension of $k_1\ k_2 \cdots k_n$. The transition rate matrix of s_t can be obtained from those of the constituent channels. Specifically, the rate constant for the transition from

$$(i_1, \cdots, i_l, \cdots, i_n) \qquad \text{to} \qquad (i_1, \cdots, j_l, \cdots, i_n)$$

is equal to the rate from i_l to j_l in the lth channel.

In other words, the data of multiple channels can be treated as if generated from a single channel, of which the states are defined as vectors and the transition rates are constrained to be those of constituent channels. Therefore, we can apply the likelihood method directly to these data by operating on the induced Markov model for s_t and by imposing appropriate constraints on its rate constants. Fortunately, all constraints are of the linear scaling type and are thus accommodated by Eq. 24.

The above approach is generally applicable to both identical and nonidentical channels. However, it should be emphasized that for identical channels there exists a more efficient way, as suggested by Horn and Lange (1983). In

this approach, we define the state vectors as

$$n_t = (n_t^{(1)}, n_t^{(2)}, \cdots, n_t^{(k)}), \quad (29)$$

where $n_t^{(i)}$ represents the number of channels in state i at time t, and clearly

$$\sum_{i=1}^{k} n_t^{(i)} = n.$$

The state space for n_t corresponds to all possible compositions of n into k parts. The dimension of the state space is thus given by the binomial coefficient $\binom{n+k-1}{n}$. The rate constant for the transition from

$$(m_1, \cdots, m_i, \cdots, m_j, \cdots, m_k)$$

to

$$(m_1, \cdots, m_i - 1, \cdots, m_j + 1, \cdots, m_k)$$

is equal to $m_i q_{ij}$, where the q_{ij} values are the rate constants of the constituent channel.

The composition vectors n_t for given n and k may be generated automatically by a standard combinatorial analysis algorithm (Nijenhuis and Wilf, 1978). Because $\binom{n+k-1}{n}$ is often much smaller than k^n, the state space of n_t is much smaller than that of s_t. Considering that the computation involved in the likelihood calculation is typically quadratic on the number of states, one may expect a considerable reduction in both processing time and memory space by using n_t for modeling multiple identical channels. This reduction is attributed to the fact that the state vector n_t fully exploits the indistinguishibility of constituent channels.

EXPERIMENTAL RESULTS

We present a number of examples illustrating various features of the likelihood method described in the previous sections. These examples are divided into five groups, the first illustrating the performance of the missed event correction, the second showing the applicability to nonstationary data, the third showing the capability of fitting data containing multiple channels, the fourth showing the capability of global fitting across different experimental conditions, and the last showing the applicability to real experimental data.

The program was written in ANSI C language. The subroutines for variable metric optimization, matrix singular value decomposition, and matrix eigenvalue calculation are taken from Press et al. (1992) with some modifications. The calculations were done on an IBM RS/6000 model 590, which is a 200 SPECfp computer, approximately four times faster than a Pentium 90.

For simulated data, the dwell times were generated by using exponentially distributed random numbers (Lauger, 1988). For real data, they were idealized by using a hidden Markov modeling procedure based on the segmental k-means method (Rabiner et al., 1986). This procedure, as compared with the traditional threshold detection, has greater accuracy, requires less filtering, and allows more short events to be detected. The dead time t_d is imposed retrospectively; that is, all events of duration less than a chosen t_d were deleted, the corresponding times being added to the preceding observed dwell.

Missed event correction

Binary conductance levels

We first investigate the performance of the method on missed event corrections in the case of binary conductance levels. The studies were carried out by imposing a sequence of dead times increasingly until the program failed to uncover the underlying rates. The model we employed in the simulation was based on a three-state linear gating scheme:

$$C_1 \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} O \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} C_2$$

Scheme 1

where $k_{12} = 100$, $k_{21} = 40$, $k_{23} = 60$, and $k_{32} = 5000$, all with units in s^{-1}. Because the mean lifetime of C_2 was much shorter than the other two states, brief closures occurred during opening. Thus, the imposition of a dead time would cause the apparent opening duration to be longer than what it should be.

Table 1 presents the parameter estimation results for different dead times. It can be seen that the estimates are very close to the underlying true values until a dead time of 0.7 ms. Notice that the mean lifetime of C_2 is 0.2 ms; thus $t_d = 0.7$ ms resulted in a considerable loss of the simulated data. Approximately 60% of the total events and nearly all of the second closures were omitted. In this case, the method seems to have been very successful in correcting for the effects of missed events. We also see from Table 1 that only the rates between C_2 and O are affected by missed events, whereas the rates between C_1 and O remain unchanged. This is because the majority of missed events came from the transitions between C_2 and O, as k_{32} is relatively large. If t_d is made longer than 0.7 ms, the likelihood space is so flat that the optimizer will not converge (see Fig. 2).

TABLE 1 Parameter estimates for Scheme 1 with different dead times

t_d	k_{12}	k_{21}	k_{23}	k_{32}
0	99 ± 4	43 ± 2	60 ± 2	5,285 ± 215
0.1	97 ± 3	42 ± 1	62 ± 3	5,329 ± 234
0.2	98 ± 3	41 ± 1	61 ± 4	5,182 ± 296
0.3	97 ± 3	41 ± 1	64 ± 8	5,279 ± 410
0.4	99 ± 2	39 ± 1	58 ± 8	5,005 ± 351
0.5	100 ± 1	39 ± 1	63 ± 12	5,254 ± 408
0.6	100 ± 1	40 ± 0	63 ± 18	5,299 ± 510
0.7	99 ± 1	39 ± 0	58 ± 18	5,006 ± 469

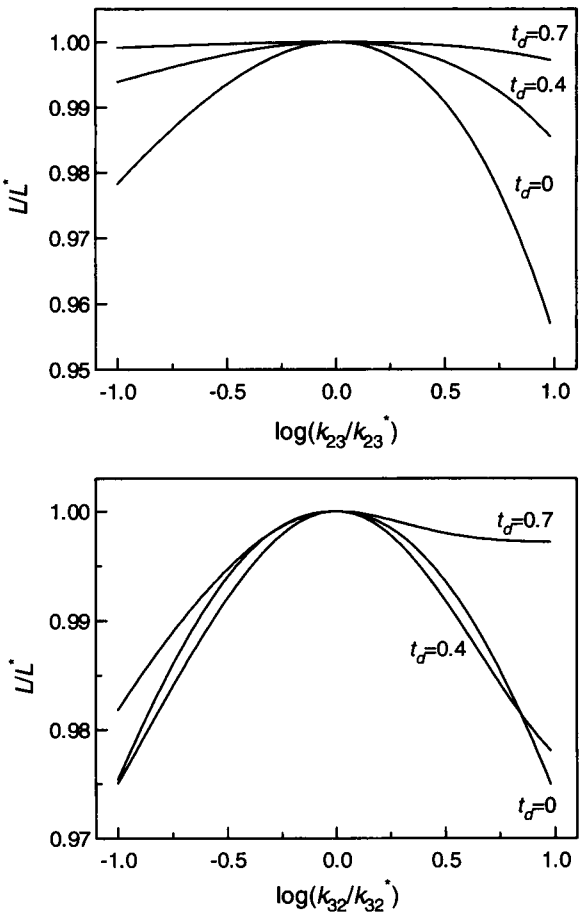


FIGURE 2 Sections of the likelihood surface for Scheme 1 at different dead times. The sections are cut parallel to the parameter axes and pass through the maximum.

The minimal number of events necessary to yield good estimates seems to vary with the imposed dead time; with larger dead times, more events are needed. However, the exact minimal number is difficult to define, as it depends on the model, the dead time, and the desired accuracy. To obtain the results in Table 1, we used approximately 3,000 events (after omission of short events) for t_d = 0–0.3, 6,000 for 0.4, 12,000 for 0.5, 20,000 for 0.6, and 40,000 for 0.7. The algorithm was fast. Even with 40,000 events, it took only ~26 s to converge to the optimal parameter values, each evaluation of the likelihood and its derivatives taking ~4 s. The initial guesses of the parameters were the values used to generate the data. Poor initial guesses may require a greater number of iterations, but not significantly. Typically, the program converged within 10~20 iterations, even with grossly inaccurate starting points.

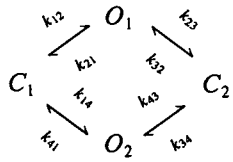
One might expect that there are local maxima points on the likelihood surface, as the likelihood function is highly nonlinear on model parameters. However, our experience showed that this was usually not the case. Different starting points were tried, and the same final results were obtained, suggesting that the likelihood surface was smooth within a reasonably large range. In this context, it is interesting to notice similar findings reported in the literature (Horn and Lange, 1983; Ball and Sansom, 1989).

The error limits on parameter estimates tend to increase as the dead time increases, as seen from Table 1. This agrees with our intuition that we should have less confidence in the fits when more events are missed. This observation is further confirmed by examining the likelihood surface. Fig. 2 shows the likelihood surface as sections passing through the maximum point and cut parallel to the parameter axes. Only the sections of k_{23} and k_{32} are presented as the other two rates are not affected by missed events. It is seen that the maximum is particularly distinct in the absence of missed events, less so for larger dead times, and becomes extremely flat when t_d = 0.7 ms.

We stress that even for this relatively simple model the correction for missed events is important. When no correction was applied, the estimates of k_{23} and k_{32} became highly biased, even with a small dead time. For example, when t_d = 0.1 ms, the estimation results without correction are k_{12} = 100, k_{21} = 42, k_{23} = 39, and k_{32} = 2895, of which k_{23} and k_{32} are nearly 50% biased from their true values.

Multiple conductance levels

We now explore the performance of the missed event correction in the case of multiple conductance levels. The model we tested is a cyclic doubling of that in Scheme 1, as follows.



Scheme 2

We assumed that the two branches $C_1O_1C_2$ and $C_1O_2C_2$ have the same kinetics but different conductances. The states are numbered as C_1 = 1, O_1 = 2, C_2 = 3, and O_2 = 4. The rate constants employed in our simulation were based upon those for $k_{12} = k_{14}$ = 100, $k_{21} = k_{41}$ = 40, $k_{23} = k_{43}$ = 60, and $k_{32} = k_{34}$ = 5000, all with units in s^{-1}. As with Scheme 1, the channel exhibited brief closures during openings. However, the behavior is more complicated because these brief closures occurred from two different conductance levels.

Parameter estimation was done by imposing the symmetric constraints, i.e., $k_{12} = k_{14}$, $k_{21} = k_{41}$, $k_{23} = k_{43}$ and $k_{32} = k_{34}$, so there were four independent variables. Table 2 presents the estimation results for different dead times. Maximization was repeated by using different initial guesses, with the same final result. The number of events used to obtain these estimates was ~16,000 for $t_d \leq 0.4$, 30,000 for 0.4, and 50,000 for 0.5. The processing time to find a maximum was ~12 s (six iterations) in the case of no missed events and 36 s (nine iterations) for t_d = 0.5. The

TABLE 2 Parameter estimates for Scheme 2 with different dead times: symmetric constraints

t_d	$k_{12} = k_{14}$	$k_{21} = k_{41}$	$k_{23} = k_{43}$	$k_{32} = k_{34}$
0	98 ± 2	39 ± 1	62 ± 1	5,074 ± 87
0.1	99 ± 2	39 ± 1	61 ± 1	5,079 ± 79
0.2	100 ± 1	40 ± 1	60 ± 1	5,100 ± 80
0.3	100 ± 1	40 ± 1	59 ± 1	5,198 ± 95
0.4	100 ± 1	40 ± 1	58 ± 1	5,229 ± 99
0.5	100 ± 1	40 ± 3	57 ± 3	5,774 ± 375

processing time appears to increase linearly with the number of events but does not seem to increase with the dead time.

From Table 2, we see that good estimates were obtained with a dead time up to 0.5 ms, which is as large as five times the minimal mean lifetime of the states in the model. This is quite remarkable in light of the complexity of the model as well as the considerable loss of the original data, approximately 50% of the total events and nearly all of the fast closures. The parameter error estimates are very small in the case of no missed events but increase with the dead time, especially for k_{32} (k_{34}). Examination of the likelihood surface confirms that the section of k_{32} (k_{34}) becomes flatter as the dead time gets larger, as illustrated in Fig. 3, and when $t_d = 0.5$ the maximum becomes poorly defined.

The success of the method does not seem to depend upon the imposition of the symmetric constraints. We also investigated fitting the model to the same data sets, but only with the detailed balance constraint; i.e., the clockwise and the anticlockwise products of the rate constants must be equal. The final parameter estimates were comparable to those obtained with the symmetric constraints, although the running time became a little longer as there were three more free parameters to be optimized. As an example, Table 3 lists the estimation results for $t_d = 0.5$ ms. The processing time taken for this example is approximately 1 min in 13 iterations.

Again, the correction for the effects of missed events was essential for parameter estimation with this model. For example, the estimates of those rates related to the fast

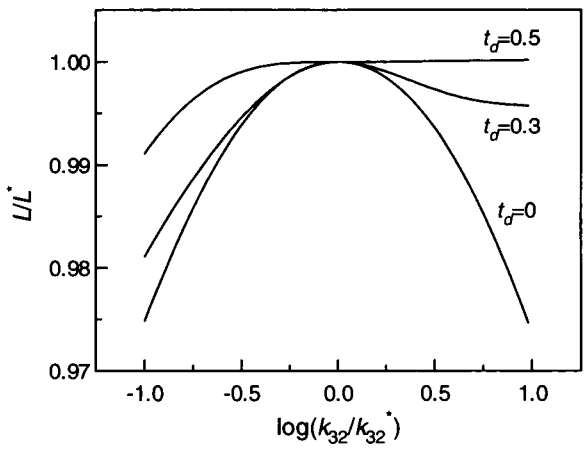


FIGURE 3 The k_{32} (k_{34}) section of the likelihood surface for Scheme 2 at different dead times.

TABLE 3 Parameter estimates for Scheme 2 with dead time 0.5 ms: detailed balance constraints

Rate	Estimate	SD
k_{12}	101	1
k_{21}	41	3
k_{23}	57	2
k_{32}	5,787	572
k_{14}	98	2
k_{41}	39	1
k_{43}	57	2
k_{34}	5,753	575

closures deviated by more than 30% from their true values after deleting events shorter than 0.1 ms. An increase to 0.2 ms without correction caused the algorithm not to converge because of numerical instability.

More examples on the missed event correction

In the previous examples we have deliberately chosen the underlying gating kinetics to satisfy the first-order approximation for the missed event correction (Roux and Sauve, 1985). However, we have also tested a number of examples in which this condition was less true, notoriously in buzz mode. For example, using the two-state model given by Magleby and Weiss (1990),

$$C \underset{1000}{\overset{2000}{\rightleftarrows}} O$$

we found that for an original data set of 200,000 events, the rate constants could be extracted with 10% accuracy using dead times up to 0.3 ms. With longer dead times, the rates oscillated, giving $k_{co} = 2264$ and $k_{oc} = 1367$ at $t_d = 3$ ms. In contrast, the simulation method (Magleby and Weiss, 1990) was able to extract accurate rates with dead times up to 3 ms.

For models that do not have such serious buzz mode kinetics, the likelihood method can yield good estimates with the imposed dead time at least as large as the minimal mean lifetime of the states in the model. Below we present several examples constructed on realistic gating schemes reported in the literature.

Scheme 3 is a widely used kinetic model for the nicotinic acetylcholine receptor. Table 4 presents the true values of the rate constants used in the simulation and their estimates obtained with $t_d = 0.058$ ms. This dead time was approximately five times the mean lifetime of C_3 and resulted in a

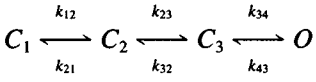
TABLE 4 Parameter estimates for Scheme 3 with dead time 0.058 ms

Rate	True value	Estimate	SD
k_{12}	200	201	7
k_{21}	500	527	31
k_{23}	400	417	91
k_{32}	25,000	25,443	9,147
k_{34}	60,000	59,054	22,555
k_{43}	240	240	125

TABLE 5 Parameter estimates for Scheme 4 with dead time 0.66 ms

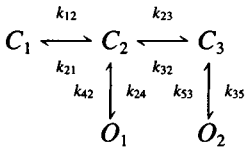
Rate	True value	Estimate	SD
k_{12}	34	32	1
k_{21}	180	176	11
k_{23}	285	292	25
k_{32}	600	585	257
k_{24}	120	112	43
k_{42}	2,860	3,036	654
k_{35}	3,950	4,632	1,464
k_{53}	322	390	270

loss of approximately 70% of the simulated events. The estimation was based on 15,000 events and took approximately 20 s, including 11 iterations and 24 evaluations of the likelihood function.



Scheme 3

Scheme 4 is a model for Ca^{2+}-activated K^+ channels in cultured rat muscle (Magleby and Pallotta, 1983). The two open states are of the same conductance but have different kinetics. The values of the rate constants in the model are given in Table 5. Also shown in Table 5 are the estimation results from a sequence of 6000 simulated events, obtained with a dead time of 0.66 ms. This dead time was approximately three times the mean lifetime of C_3 and led to a loss of 84% of the simulated events. The processing time for this example was approximately 21 s, including 17 iterations and 30 evaluations of the likelihood function.



Scheme 4

Scheme 5 is a simplification of a possible mechanism for fast Cl^- channels in rat skeletal muscle (Blatz and Magleby, 1986b). This model contains up to 10 indepen-

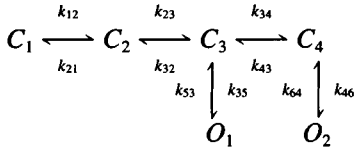
TABLE 6 Parameter estimates for Scheme 5 with dead time 0.0625 ms

Rate	True value	Estimate	SD
k_{12}	105	109	9
k_{21}	172	202	31
k_{23}	521	589	129
k_{32}	287	316	61
k_{34}	913	887	40
k_{43}	2,898	2,902	70
k_{35}	180	199	18
k_{53}	1,438	1,408	102
k_{46}	14,200	14,085	281
k_{64}	1,450	1,401	23

TABLE 7 Parameter estimates for Scheme 6 with dead time 0.125 ms

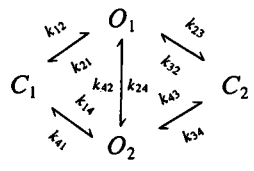
Rate	True value	Estimate	SD
k_{12}	10	10	0
k_{21}	100	101	3
k_{14}	10	10	0
k_{41}	100	97	2
k_{23}	100	79	2
k_{32}	4,000	3,328	88
k_{43}	100	79	2
k_{34}	4,000	3,311	90
k_{24}	400	409	6
k_{42}	400	413	6

dent parameters. The values of these parameters are listed in Table 6. Parameter estimation was done with a set of 35,000 events and with the imposition of $t_d = 0.065$ ms. This dead time was approximately equal to the minimal mean lifetime of the six states and resulted in the loss of ~55% of the simulated events. The estimation results are presented in Table 6. The running time to obtain these estimates was ~210 s, each evaluation of the likelihood function and its derivatives taking approximately 6 s.



Scheme 5

As a last example on missed event correction, we consider the following diamond model. The two open states have different conductances where O_1 is partly open and O_2 is fully open. The channel is assumed to be at thermodynamic equilibrium, and hence the rates on all loops must satisfy the detailed balance constraints. Table 7 presents the values of the rate constants employed in the simulation as well as the estimated values from a sequence of 40,000 simulated events. The estimation was done by imposing a dead time of 0.125 ms, which is more than the mean lifetime of C_2. The program took ~33 s to converge, each evaluation of the likelihood and its derivatives taking ~1 s.

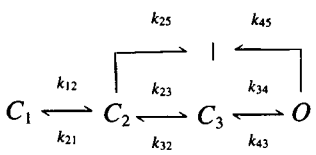


Scheme 6

Nonstationary data

To examine the applicability of the maximal likelihood algorithm to channels exhibiting nonstationarity, we em-

ployed the following model:



Scheme 7

where the values of the rate constants are listed in Table 8. This model was derived by Vandenberg and Bezanilla (1991) as a possible gating mechanism for the sodium channel in the squid giant axon. Assuming the channel began in state C_1, we simulated 3000 current responses to a 20-ms voltage step to 0 mV. As inactivation may occur directly from a closed state, there were a significant number of blank traces in which no channel openings occurred. Because the opening rate is very fast, the lifetime of the third closed state C_3 is short, leading to brief closures occurring before inactivation takes place. There were 835 blank traces, 4340 openings, and 7340 closings, an average of approximately 2 openings per trace.

Two sets of analyses were carried out: one with the simulated perfect data and the other with the imposition of a dead time $t_d = 0.034$ ms (approximately the mean lifetime of the brief closures). The processing time was approximately 3 min. Table 8 presents the estimation results. These estimates were obtained with the constraints $k_{12} = k_{23}$ and $k_{21} = k_{32}$. Such constraints appeared to be necessary as the program was unable to converge when they were allowed to be independent. The same behavior was found by Vandenberg and Bezanilla (1991) when they fitted real experimental data using the method of Horn and Lange (1983). From Table 8 we see that the estimates are reasonably good. The inactivation rates from the closed state are underestimated in both cases, suggesting the difficulty of extracting rates from blank records. The leftward rates between the closed states were overestimated in the presence of missed events, and the overestimation continued to increase with the dead time. The inability of the program to converge in the absence of constraints seems to be a fundamental problem of the reaction scheme as data sets as small as 30 traces gave comparable mean values (although having wider confidence limits). The problem may be solved, however, by global fitting across several voltages.

TABLE 8 Parameter estimates for nonstationary data

Rate	True value	Estimates (±SD) $t_d = 0$	Estimates (±SD) $t_d = 0.034$ ms
$k_{12} = k_{23}$	2,969	3,052 ± 72	3,381 ± 93
$k_{21} = k_{32}$	704	738 ± 130	1,145 ± 182
k_{34}	28,932	29,823 ± 665	30,043 ± 1,253
k_{43}	725	751 ± 16	747 ± 42
k_{25}	1,117	710 ± 30	730 ± 34
k_{45}	705	739 ± 16	722 ± 16

Multi-channel fitting

This is an example based on the three-state K^+ channel in *Chara corallina* (Albertsen and Hansen, 1994). The gating mechanism of the channel follows Scheme 1. The rate constants are $k_{12} = 70$, $k_{21} = 1{,}000$, $k_{23} = 15{,}000$, and $k_{32} = 300$, all with units in s^{-1}. We attempted to extract these rates from a record containing four such channels. A sequence of 30,000 events, totaling a duration of 15 s, were simulated and used for analysis.

Table 9 presents the estimates of the rate constants of a single channel. A dead time of $t_d = 0.0625$ ms was imposed. This dead time was equal to the mean lifetime of the open state and caused ~70% of the simulated events to be lost. Thus there were only 10,000 events that were actually used in estimation. From Table 9 it can be seen that the parameter estimates are close to their true values. The processing time to obtain these estimates was approximately 11 min, including 10 iterations and 17 evaluations of the likelihood function.

It is worth comparing our method with the one recently proposed for multi-channel data analysis (Albertsen and Hansen, 1994). For the same model and the same number of channels, but only one-third of our data length, their method took as much as 24 h on a similar workstation (HP 9000). We notice that they employed the Kronecker product technique for computing the transition probability matrix of multiple channels from the individual matrices. As discussed previously, this technique is computationally much more expensive than the composition technique. For n channels and k states, the composition technique yields a reduction by a factor of $(N_c/N_k)^2$ where $N_c = k^n$ and $N_k = \binom{n+k-1}{n}$. For $n = 4$ and $k = 3$, this factor is ~29. In other words, if the composition technique had been adapted, their method would have been 29 times faster than it was.

Fig. 4 shows the increase of the computation time per evaluation of the likelihood and its derivatives with the number of channels. It is seen that the time increases dramatically as the number of channels increases.

Global fitting

Scheme 8 is taken from a model of a batrachotoxin-modified sodium channel in neuroblastoma cells (Huang, 1984). All of the four rate constants in the model are voltage dependent: $k_{ij} = \kappa_{ij}\exp(\nu_{ij}V)$. Thus there are actually eight independent parameters. The true values of these parameters are listed in Table 10. Simulations based on these parameter

TABLE 9 Parameter estimates for multiple channels

Rate	True value	Estimate	SD
k_{12}	70	61	20
k_{21}	1,000	1,008	665
k_{23}	15,000	13,430	687
k_{32}	300	255	20

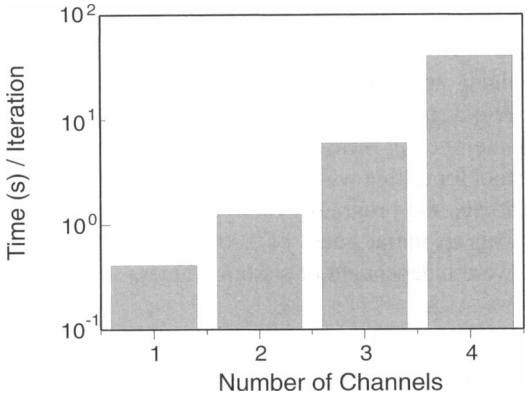


FIGURE 4 Increase of the time per evaluation of the likelihood and its derivatives with the number of channels.

values were done at five different voltages: −90, −70, −50, −30, and −10 mV, and approximately 3500 channel openings were generated at each voltage.

$$C_1 \underset{k_{21}}{\overset{k_{12}}{\longleftrightarrow}} C_2 \underset{k_{32}}{\overset{k_{23}}{\longleftrightarrow}} O$$

Scheme 8

Parameter estimation was done by fitting the records of the five voltages simultaneously. A dead time of 0.1 ms was imposed. The estimation results are presented in Table 10. It can be seen that the estimates are close to the true values employed in the simulation, and the error estimates are small. The processing time to obtain these estimates was approximately 24 s, including 14 iterations and 37 evaluations of the likelihood. The maximization was repeated by using different initial guesses and was shown to be robust.

The advantages of global fitting are apparent; the likelihood surface is well defined, thus yielding more reliable estimates. For this example, global fitting is not just helpful but also necessary. Fredkin and Rice (1992a) have tried the individual fitting of the model at −70 mV. They concluded that good estimates on k_{12} and k_{21} were impossible no matter how many events were analyzed because the closed duration density can be well approximated by a single exponential and a fit with two exponentials is inherently unstable. Here we see that this problem can be successfully resolved by fitting the model across a voltage range.

Application to experimental data

A nicotinic acetylcholine receptor channel

Our first experimental example is that of a mutant, recombinant mouse acetylcholine (ACh) receptor expressed via transient transfection of human embryonic kidney cells (Sine et al., 1995). The recordings were made from a cell-attached patch with the agonist, 1 μM ACh, in the patch pipette only. Although the data were recorded continuously for extended periods, the currents occurred in clusters that were well separated and could thus be unambiguously defined. The sampling rate was 94 kHz, and the data were low-pass filtered at 15 kHz using a digital linear phase Gaussian filter. The idealization process was performed with the segmental k-means method. A total of 5176 events (~12 s) from 15 clusters was included in the fit. The model fitted to the data was the standard kinetic model for the activation of these receptors as given in Scheme 3.

The results of estimation are presented in Table 11. The analysis imposed a dead time of 11 μs. Longer dead times were also tried, with similar results. As before, the maximization was repeated by using different starting points, and the same final estimates were obtained. Examination of likelihood surface sections shows that the maximum was well defined. The processing time for this example was approximately 18 s, starting from an initial guess of 100 for all rates except k_{34}, which was 10,000.

As a verification of the estimation results, we calculated the open and closed interval duration distributions from the fitted rates and compared them with the experimentally obtained histograms. From Fig. 5, it can be seen that there was good agreement. There was an excess of short-lived open intervals (probably from mono-liganded receptors); to account for these, a second open state would need to be added to the model.

An N-methyl-D-aspartate (NMDA) receptor channel

Our second experimental example is that of a mutant, recombinant mouse NMDA-activated receptor expressed in *Xenopus* oocytes. The recordings were made from outside-out patches (Fig. 6, top trace). In these patches, only a single channel was active. A total of ~100 s of data was

TABLE 10 Global fitting across different voltages

Parameters	True values	Estimates	SD
κ_{12}	3.39	3.15	0.22
κ_{21}	6,357	6,177	437
κ_{23}	12,345	12,428	458
κ_{32}	4.97	5.10	0.15
ν_{12}	−0.047	−0.048	0.0024
ν_{21}	0.105	0.103	0.0027
ν_{23}	0.079	0.079	0.0005
ν_{32}	−0.042	−0.042	0.0005

TABLE 11 Parameter estimates for an ACh receptor channel

Rate (s^{-1})	Estimate	SD
k_{12}	239	35
k_{21}	1,316	305
k_{23}	526	105
k_{32}	3,885	215
k_{34}	36,896	859
k_{43}	580	12

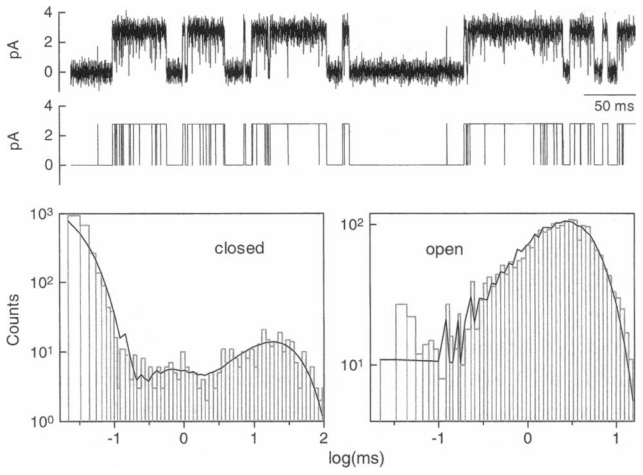
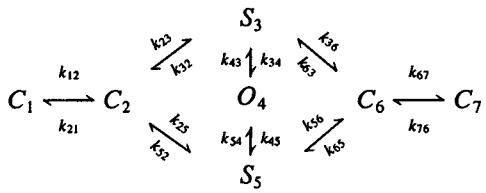


FIGURE 5 Analysis of a mutant ACh receptor kinetics. The theoretical curve (*solid line*) was calculated as the integral over each bin of the pdf derived from the fitted rates with allowance for missed events. It is not smooth because the bin widths were rounded to multiples of the sampling duration.

divided into 21 segments of equal duration. We selected the data to analyze, choosing stable baselines and no obvious noise spikes. The data were digitized at a sampling rate of 20 kHz and low-pass filtered to 10 kHz. The segmental k-means program was used to estimate the amplitudes and to idealize the currents creating an idealized dwell time list that contained 14,727 intervals. The currents appeared to reside at three amplitude levels: closed (0 pA), subconductance (−5 pA), and open (−7 pA).



Scheme 9

Exploration of a variety of models suggested that there are at least four closed states, two sublevel states, and one open state. After some experimentation, we chose the seven-state symmetric model shown in Scheme 9. Because the model has several loops that need not be in equilibrium, we performed two analyses, one with and one without detailed balance constraints. In both cases, a dead time of 0.1 ms, i.e., two sample durations, was imposed. The optimal rate constants are shown in Table 12. The corresponding duration histograms for each conductance level are shown in Fig. 6. Although the calculated duration distributions fit the experimental ones in both cases, the log likelihoods are very different: 70,001 without constraints and 69,987 with constraints. Thus, the nonequilibrium model was $e^{14} \approx 10^6$ times more likely. The outer loop of the model (2–3-6–5) preferentially cycles in the clockwise direction, gaining approximately $0.7k_bT$ per cycle, where k_b is the Boltzmann constant and T is the temperature. By readily solving for the rates of different models, the program provides a simple tool for testing whether channels are in equilibrium. For this example, the program took approximately 3 min to converge with an initial guess of 100 s^{-1} for all 16 rates. The results were independent of starting values.

Discussion

We have described a full maximal likelihood method for estimating kinetic parameters of a single channel from idealized dwell times. The method applies a correction for missed events and works on stationary or nonstationary records with multiple conductance levels and multiple channels. It also allows data sets obtained at different independent variables (concentration, voltage, or force) to be fit as a single data set. The method is very fast, typically taking approximately 1 min to extract the rate constants for a typical model.

There are two major improvements made by the present method over the existing ones. One is the capability of correcting for the effects of missed events. A general solution that provides a corrected transition rate matrix and works on records with multiple conducting states was derived and implemented. The other improvement made by the method is the use of analytical derivatives of likelihood function for optimization. This offers several computational advantages. First, it avoids numerical calculation of likelihood derivatives, thus yielding a significant reduction in computation. Second, it greatly helps the search of the likelihood surface, especially when the parameter space is large or the likelihood surface is nearly flat. Finally, it enables one to use the efficient variable metric optimizer. This optimizer not only converges rapidly but also generates good estimates of the curvature of the likelihood surface at the maximum, thereby allowing the parameter error estimates to be calculated without additional computation.

The program incorporates a unified representation of rate constants by which global fitting across different experimental conditions is made possible. This representation may be interpreted physically as formulating transition rates in terms of free energy. In addition to allowing global fitting, it also has the advantage of converting nonlinear detailed balanced constraints into linear ones that are computationally easier to deal with. Another advantage is that it results in a more symmetric likelihood surface, which may be better approximated parabolically over a larger range instead of just the vicinity of the maximum point. As a consequence, optimization is made faster and error estimates are more accurate.

The method has been tested extensively based on a variety of models with both simulated and experimental data. The results have been encouraging. Examples with missed events correction showed that good parameter estimates

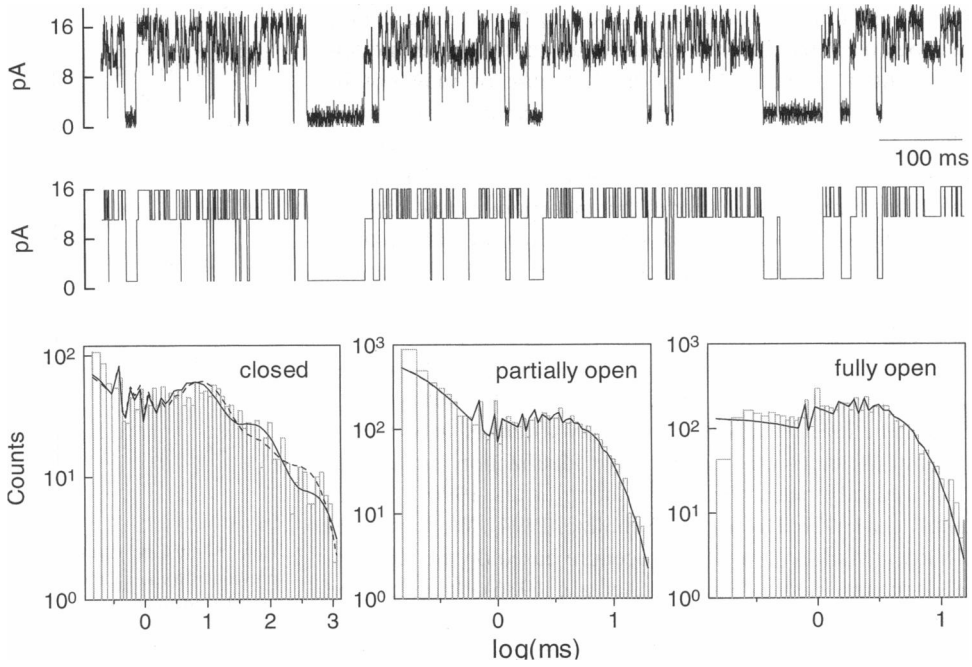


FIGURE 6 Analysis of a mutant recombinant NMDA receptor kinetics. A portion of the actual and idealized currents are shown at the top, and the duration histograms are shown at the bottom. Calculated histograms were made as described in Fig. 5 and are displayed as lines. The calculated curves obtained with the detailed balance constraint are plotted as dashed lines and those obtained without constraints are shown as solid lines. The curves can only be distinguished for the closed state.

could be revealed even after the majority of the data were omitted. For models that met the first-order approximation condition, a dead time as long as several times the minimal mean lifetime of the states in a model could be imposed. For models that did not satisfy this condition, a dead time longer than the minimal mean lifetime could still be imposed without changing estimation results significantly. The method was also shown to be effective for data containing multiple channels and data scaled by different experimental variables, e.g., voltage, force, and concentration. Finally, analysis of two examples, one with an ACh receptor and the other with an NMDA receptor, indicated that the current method performed well with real patch-clamp data, suggesting its applicability to more general and complicated problems.

TABLE 12 Parameter estimates for an NMDA receptor channel

Rate (s^{-1})	Estimates (±SD) unconstrained	Estimates (±SD) constrained
k_{12}	32 ± 5	37 ± 6
k_{21}	47 ± 8	655 ± 111
k_{23}	25 ± 3	835 ± 110
k_{32}	14 ± 2	37 ± 4
k_{34}	342 ± 6	342 ± 19
k_{43}	367 ± 7	366 ± 21
k_{45}	181 ± 6	181 ± 11
k_{54}	2,700 ± 84	2,688 ± 167
k_{36}	35 ± 3	10 ± 1
k_{63}	1,140 ± 165	15 ± 1
k_{56}	477 ± 61	1,387 ± 81
k_{65}	1,360 ± 142	127 ± 11
k_{67}	446 ± 82	26 ± 4
k_{76}	3 ± 0	5 ± 1

Although global fitting should be better than fitting individual data records, we have found that in practice there may be enough variability in the channel kinetics from patch to patch to reduce the potential advantages. The assumption in global fitting is that the same reaction scheme is present in all of the data sets and when that is not the case, global fitting can be misleading. In our experience, fitting of individual data sets is often adequate and simpler than global fitting.

For most nonlinear optimization problems, initial guesses of parameters are important as poor starting values may lead to local maxima. Most surprising to us, local maxima do not seem to occur in our case. We have tried different starting points and obtained the same results. This suggests that the likelihood surface for this kind of problem is smooth. Similar results have been reported by other authors (Horn and Lange, 1983; Ball and Sansom, 1989). Nevertheless, caution must be taken and different initial guesses should be tried, especially for large and complicated models.

One problem encountered when applying this method in practice is that of defining a dead time. Our suggestion is to start with a value calculated from the analog system response time as presented by Colquhoun and Sigworth (1983); $t_d = 0.179/f_c$ where f_c is the 3-db cutoff frequency of the low-pass filter. We then search a range of dead times for a region in which the rates are stable. This at least provides a self-consistent answer.

The method assumes the availability of an explicit kinetic model. Sometimes, the model can be derived from the data. The minimal number of states for a given amplitude class can be ascertained by determining the number of exponentials that is required to fit the interval duration histogram of that class. The connectivity between states can be tested by comparing their best log likelihood fits, for example, using

the Akaike information criteria (Akaike, 1974). Models may also be constructed from a priori ideas. In this sense, models can contain global information that may or may not be available within a single data set. For example, structural, pharmacological, and biochemical results may tell us that a protein has exactly two sites for binding a drug. Then, we may start with a kinetic model that incorporates two binding sites before we even examine the characteristics of the data.

Here is not the place to discuss the details of idealization; however, the success of the method does depend on the availability of an accurately idealized dwell time list. Traditionally, the idealization process has been performed with threshold detection (Sachs et al., 1982; Colquhoun and Sigworth, 1983) that relies on heavy filtering, leaving fast transitions undetected. Alternative idealization procedures have recently been proposed (Chung et al., 1990; Fredkin and Rice, 1992b; Draber and Schultze, 1994). We have used a hidden Markov modeling procedure based on the segmental k-means method (Rabiner, et al., 1986). It should be emphasized that, although the present method has the ability to correct for missed events, more perfect data, whenever possible, are still desirable.

The applicability of the program is not likely to be limited by the size of the model. The program is usually fast and requires only a small memory space. For example, the NMDA receptor example we tested involved as many as 16 independent variables, and the program converged in a few minutes. One limitation is the adequacy of the missed events correction. If a higher order correction can be developed this would extend the applicability of the current method. In the absence of that, point likelihood methods may help (Walsh and Sigworth, 1992; Fredkin and Rice, 1992a; Qin et al., 1994; Albertsen and Hansen, 1994).

The method suffers from a problem common to all kinetic analyses, that of identifiability, as illustrated above for the two sodium channel models. The problem can be reduced by using nonstationary data (in time or across different values of the rate constants) as in Scheme 8. Imposing constraints on rates, such as detailed balance, can also improve identifiability. The problem of identifiability arises primarily from the aggregation of states. If mutant channels are constructed so that more states have distinguishable conductances, the problems of identifiability are reduced.

APPENDIX

We present here a derivation for the corrected probability density matrix ${}^{e}G_{ab}(t)$ given by Eq. 14 as well as its first-order approximation given by Eq. 15. The method employed here is a generalization of the one introduced by Roux and Sauve (1985) in the case of binary conductance levels. A minor error occurred in their derivation, where the constraint that the first stay τ_1 (Fig. 1) must be greater than t_d was not taken into account. We have made the correction in the following derivation. In addition, we present the derivation with more commonly used notation.

As shown in Fig. 1, there are two exclusive ways to implement an apparent dwell time of a channel with multiple conductance levels, one without missed events on the falling phase and the other with missed events on the falling phase. (Rather arbitrarily, missed events on the rising phase are lumped with the previous dwells.) Let $\Gamma_1(n)$ be the probability density matrix of the first case, containing n undetected leavings. Similarly, let $\Gamma_2(n)$ be the probability density matrix of the second case. Then the probability density matrix of the observed dwell time after correcting for missed events is given by

$$ {}^{e}G_{ab}(t) = \sum_{n=0}^{\infty} [\Gamma_1(n) + \Gamma_2(n)]. \tag{A1} $$

Thus the derivation of ${}^{e}G_{ab}(t)$ reduces to the derivations of $\Gamma_1(n)$ and $\Gamma_2(n)$.

We first consider $\Gamma_1(n)$. The probability density matrix for a given set of missed events (see the upper panel of Fig. 1) is given by

$$ \left[\prod_{i=1}^{n} G_{a\bar{a}}(\tau_{2i-1})G_{\bar{a}a}(\tau_{2i})\right] G_{ab}(\tau_{2n+1}) \equiv F_n(\tau_1, \cdots, \tau_{2n+1}). \tag{A2} $$

Here τ_1 must be greater than t_d; otherwise it would be grouped into the previous dwell. τ_{2i} must be less than t_d so that all the leavings go undetected, and the sum of all τ_i values must be equal to the duration of the observed dwell. Then $\Gamma_1(n)$ can be obtained by integrating Eq. A2 over all possible values of τ_i; i.e.,

$$ \Gamma_1(n) = \int\int\cdots\int_{\substack{\sum \tau_i = t \\ \tau_1 > t_d \\ \tau_{2i} \le t_d}} F_n(\tau_1, \cdots, \tau_{2n+1})\, d\tau_1\, d\tau_2 \cdots d\tau_{2n+1}. $$

This is an integral of a matrix function over a restricted region of a $(2n + 1)$-dimensional hyperplane. By subtracting t_d from τ_1 and introducing a Dirac function into the integrand, we can eliminate the constraints $\tau_1 \le t_d$ and $\Sigma\tau_i = t$ from the integration limits, yielding

$$ \Gamma_1(n) = e^{Q_{aa}t_d} \int\int\cdots\int_{\substack{0 \le \tau_{2i-1} < \infty \\ 0 \le \tau_{2i} \le t_d}} \delta(\textstyle\sum \tau_i - t') \times F_n(\tau_1, \cdots, \tau_{2n+1})\, d\tau_1\, d\tau_2 \cdots d\tau_{2n+1} \tag{A3} $$

where $t' = t - t_d$. Representing the Dirac function in the Fourier domain as

$$ \delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega x}\, d\omega $$

and substituting for $F_n(\tau_1, \tau_2, \cdots, \tau_{2n+1})$, we may rewrite the integrand in Eq. A3 as

$$ \begin{aligned} &\delta(\textstyle\sum \tau_i - t')F_n(\tau_1, \cdots, \tau_{2n+1}) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\omega t'}\left[\prod_{i=1}^{n} e^{j\omega\tau_{2i-1}}G_{a\bar{a}}(\tau_{2i-1})e^{j\omega\tau_{2i}}G_{\bar{a}a}(\tau_{2i})\right] \\ &\qquad \times e^{j\omega\tau_{2n+1}}G_{ab}(\tau_{2n+1})\, d\omega \qquad \text{(A4)} \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\omega t'}\left[\prod_{i=1}^{n} e^{(j\omega I + Q_{aa})\tau_{2i-1}}Q_{a\bar{a}}e^{(j\omega I + Q_{\bar{a}\bar{a}})\tau_{2i}}Q_{\bar{a}a}\right] \\ &\qquad \times e^{(j\omega I + Q_{aa})\tau_{2n+1}}Q_{ab}\, d\omega. \end{aligned} $$

Substituting Eq. A4 into Eq. A3 and then making use of the following equalities:

$$\int_0^{\infty} e^{(j\omega I+Q_{aa})\tau}\,d\tau = -(j\omega I + Q_{aa})^{-1} \equiv \Phi_a(\omega)$$

$$\int_0^{t_d} e^{(j\omega I+Q_{aa})\tau}\,d\tau = -(I - e^{(j\omega I+Q_{aa})t_d})(j\omega I + Q_{aa})^{-1} \equiv \Psi_a(\omega)$$

it follows that

$$\Gamma_1(n) \tag{A5}$$
$$= \frac{1}{2\pi} e^{Q_{aa}t_d} \int_{-\infty}^{\infty} e^{-j\omega t'}[\Phi_a(\omega)Q_{a\bar{a}}\Psi_{\bar{a}}(\omega)Q_{\bar{a}a}]^n\Phi_a(\omega)Q_{ab}\,d\omega.$$

By similar arguments, we can show that

$$\Gamma_2(n) = \frac{1}{2\pi} e^{Q_{aa}t_d} \int_{-\infty}^{\infty} e^{-j\omega t'}[\Phi_a(\omega)Q_{a\bar{a}}\Psi_{\bar{a}}(\omega)Q_{\bar{a}a}]^{n-1} \times \Phi_a(\omega)Q_{ac}\Psi_c(\omega)Q_{cb}\,d\omega. \tag{A6}$$

So far, we have derived $\Gamma_1(n)$ and $\Gamma_2(n)$. Now, substituting Eqs. A5 and A6 into Eq. A1 and noting that

$$\sum_{n=0}^{\infty} A^n = (I - A)^{-1},$$

we obtain

$${}^eG_{ab}(t) = \frac{1}{2\pi} e^{Q_{aa}t_d} \int_{-\infty}^{\infty} e^{-j\omega t'}[I - \Phi_a(\omega)Q_{a\bar{a}}\Psi_{\bar{a}}(\omega)Q_{\bar{a}a}]^{-1} \times \Phi_a(\omega)[Q_{ab} + Q_{ac}\Psi_c(\omega)Q_{cb}]\,d\omega. \tag{A7}$$

If we substitute $\Phi_a(s)$ and $\Psi_a(s)$ by their definitions, then Eq. A7 becomes Eq. 14.

First-order approximation

Under the first-order assumption, i.e., that the total duration of missed events is much less than the duration of the observed dwell time, we have

$$\delta(\sum \tau_i - t) \approx \delta\left(\sum_{i\ \text{odd}} \tau_i - t\right).$$

Replacing the Dirac function in Eq. A3 with the above one and then repeating the subsequent derivation, it follows that

$${}^eG_{ab}(t) = \frac{1}{2\pi} e^{Q_{aa}t_d} \int_{-\infty}^{\infty} e^{-j\omega t'}[I - \Phi_a(\omega)Q_{a\bar{a}}\Psi_{\bar{a}}(0)Q_{\bar{a}a}]^{-1} \Phi_a(\omega)[Q_{ab} + Q_{ac}\Psi_c(0)Q_{cb}]\,d\omega. \tag{A8}$$

In other words, the first-order approximation of Eq. A7 is obtained by simply replacing the matrix function $\Psi_a(s)$ with its value at $s = 0$.

Noticing that

$$[I - \Phi_a(\omega)Q_{a\bar{a}}\Psi_{\bar{a}}(0)Q_{\bar{a}a}]^{-1}\Phi_a(\omega) = [\Phi_a^{-1}(\omega) - Q_{a\bar{a}}\Psi_{\bar{a}}(0)Q_{\bar{a}a}]^{-1}$$

and making use of the equality

$$-\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jst}(j\omega I + A)^{-1}\,d\omega = \exp(At),$$

we can further simplify Eq. A8 into Eq. 15; i.e.,

$${}^eG_{ab}(t) = \exp({}^eQ_{aa}t){}^eQ_{ab}, \tag{A9}$$

where ${}^eQ_{aa}$ and ${}^eQ_{ab}$ are defined as

$$\begin{aligned} {}^eQ_{aa} &= Q_{aa} - Q_{a\bar{a}}(I - e^{Q_{\bar{a}\bar{a}}t_d})Q_{\bar{a}\bar{a}}^{-1}Q_{\bar{a}a} \\ {}^eQ_{ab} &= \exp[t_dQ_{a\bar{a}}(I - e^{Q_{\bar{a}\bar{a}}t_d})Q_{\bar{a}\bar{a}}^{-1}Q_{\bar{a}a}] \\ &\quad \times [Q_{ab} - Q_{ac}(I - e^{Q_{cc}t_d})Q_{cc}^{-1}Q_{cb}]. \end{aligned} \tag{A10}$$

Note that Eq. A9 has the same form as in the case of no missed events. This suggests that we may view the observed process after incorporating missed events as generated from another channel, the transition rate matrix of which is determined by Eq. A10.

We thank Drs. S. Sine and C. Bouzat for providing the experimental ACh receptor data and Dr. L. Premkumar for providing the experimental NMDA receptor data. We also thank Drs. J. Rice, D. Fredkin, and K. Magleby for their helpful comments on the manuscript.

This work was supported by the National Science Foundation (9015986 and 9102233) and the Muscular Dystrophy Association (to F. S.).

REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* AC-19:716–723.

Albertsen, A., and U. Hansen. 1994. Estimation of kinetic rate constants from multi-channel recordings by a direct fit of the time series. *Biophys. J.* 67:1393–1403.

Auerbach, A. 1993. A statistical analysis of acetylcholine receptor activation in *Xenopus* myocytes: stepwise versus concerted models of gating. *J. Physiol.* 461:339–378.

Ball, F. G., and M. S. P. Sansom. 1988. Aggregated Markov processes incorporating time interval omission. *Adv. Appl. Prob.* 20:546–572.

Ball, F. G., and M. S. P. Sansom. 1989. Ion-channel gating mechanisms: model identification and parameter estimation from single channel recordings. *Proc. R. Soc. Lond. B.* 236:385–416.

Blatz, A. L., and K. L. Magleby. 1986a. Correcting single channel data for missed events. *Biophys. J.* 49:967–980.

Blatz, A. L., and K. L. Magleby. 1986b. Quantitative description of three modes of activity of fast chloride channels from rat skeletal muscle. *J. Physiol.* 378:141–174.

Chay, T. R. 1988. Kinetic modeling for the channel gating process from single channel patch clamp data. *J. Theor. Biol.* 132:449–468.

Chung, S. H., J. B. Moore, L. G. Xia, L. S. Premkumar, and P. W. Gage. 1990. Characterization of single channel currents using digital signal processing techniques based on hidden Markov models. *Proc. R. Soc. Lond. B.* 329:265–285.

Colquhoun, D., and A. G. Hawkes. 1977. Relaxation and fluctions of membrane currents that flow through drug-operated channels. *Phil. Trans. R. Soc. Lond. B.* 199:231–262.

Colquhoun, D., and A. G. Hawkes. 1981. On the stochastic properties of single ion channels. *Proc. R. Soc. Lond. B.* 211:205–235.

Colquhoun, D., and A. G. Hawkes. 1983. The principles of the stochastic interpretation of ion-channel mechanisms. *In* Single-Channel Recording. B. Sakmann and E. Neher, editors. Plenum Press, New York. 135–174.

Colquhoun, D., and F. J. Sigworth. 1983. Fitting and statistical analysis of single channel records. *In* Single-Channel Recording. B. Sakmann and E. Neher, editors. Plenum Press, New York. 191–264.

Cox, D. R., and H. D. Miller. 1965. The Theory of Stochastic Processes. Methuen, London.

Crouzy, S. C., and F. J. Sigworth. 1990. Yet another approach to the dwell-time omission problem of single-channel analysis. *Biophys. J.* 58:731–743.

Draber, S., and R. Schultze. 1994. Detection of jumps in single-channel data containing subconductance levels. *Biophys. J.* 67:1404–1413.

Fletcher, R. 1981. Practical Methods of Optimization. John Wiley & Sons, Chichester, UK.

Fredkin, D. R., M. Montal, and J. A. Rice. 1985. Identification of aggregated Markovian models: application to the nicotinic acetylcholine receptor. *In* Proceedings of the Berkeley Conference in Honor of Jerzy Neymann and Jack Kiefer. L. M. LeCam and R. A. Olshen, editors. Wadsworth, Belmont, CA. 269–289.

Fredkin, D. R., and J. A. Rice. 1992a. Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. R. Soc. Lond. B.* 239:125–132.

Fredkin, D. R., and J. A. Rice. 1992b. Bayesian restoration of single-channel patch clamp recordings. *Biometrics.* 48:427–448.

Golub, G. H., and C. F. Van Loan. 1989. Matrix Computations. Johns Hopkins University Press, Baltimore.

Hawkes, A. G., A. Jalali, and D. Colquhoun. 1992. Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Phil. Trans. R. Soc. Lond. B.* 337:383–404.

Horn, R., and K. Lange. 1983. Estimating kinetic constants from single channel data. *Biophys. J.* 43:207–223.

Huang, L. M. 1984. Gating kinetics of batrachotoxin-modified sodium channels in neuroblastoma cells determined from single-channel measurements. *Biophys. J.* 45:313–322.

Kendall, M. G., and A. Stuart. 1977. The Advanced Theory of Statistics. Griffin, London.

Lauger, P. 1988. Internal motions in proteins and gating kinetics of ionic channels. *Biophys. J.* 53:877–884.

Magleby, K. L., and B. S. Pallotta. 1983. Calcium dependence of open and shut interval distributions from calcium-activated potassium channels in cultured rat muscle. *J. Physiol.* 344:585–604.

Magleby, K. L., and D. S. Weiss. 1990. Estimating kinetic parameters for single channels with simulation: a general method that resolves the missed event problem and accounts for noise. *Biophys. J.* 58:1411–1426.

McManus, O., and K. Magleby. 1991. Accounting for the Ca^{2+}-dependent kinetics of single large-conductance Ca^{2+}-activated K^+ channels in rat skeletal muscle. *J. Physiol.* 443:739–777.

Milne, R. K., G. F. Yeo, R. O. Edeson, and B. W. Madsen. 1988. Stochastic modeling of a single ion channel: an alternating renewal approach with application to limited time resolution. *Proc. R. Soc. Lond. B.* 233:247–292.

Neher, E., and B. Sakmann. 1992. The patch clamp technique. *Sci. Am.* 266:28–35.

Nijenhuis, A., and H. S. Wilf. 1978. Combinatorial Algorithms. Academic Press, New York.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes in C. Cambridge University Press, Cambridge, UK.

Qin, F., J. L. Chen, A. Auerbach, and F. Sachs. 1994. Extracting kinetic parameters using hidden Markov techniques. *Biophys. J.* 66:A392

Rabiner, L. R., J. G. Wilpon, and B. H. Juang. 1986. A segmental k-means training procedure for connected word recognition. *AT & T Tech. J.* 65:21–31.

Roux, B., and R. Sauve. 1985. A general solution to the time interval omission problem applied to single channel analysis. *Biophys. J.* 48: 149–158.

Sachs, F., J. Neil, and N. Barkakati. 1982. The automated analysis of data from single ionic channels. *Pflugers Arch.* 395:331–340.

Sine, S. M., K. Ohno, C. Bouzat, A. Auerbach, M. Milone, J. N. Pruitt, and A. G. Engel. 1995. Mutation of the acetylcholine receptor a-subunit causes a slow-channel myasthenic syndrome by enhancing agonist binding affinity. *Neuron.* In press.

Vandenberg, C. A., and F. Bezanilla. 1991. A sodium channel gating model based on single channel, macroscopic ionic, and gating currents in the squid giant axon. *Biophys. J.* 60:1511–1533.

Walsh, J. L., and F. J. Sigworth. 1992. Extraction of single channel current from correlated noise via a hidden Markov model. Ph.D. dissertation. Yale University, New Haven, CT.

Weiss, D., and K. Magleby. 1992. Voltage-dependent gating mechanism for single fast chloride channels from rat skeletal muscle. *J. Physiol.* 453:279–306.