Biophysical Journal Volume 79 October 2000 1928–1944

Hidden Markov Modeling for Single Channel Kinetics with Filtering and Correlated Noise

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

ABSTRACT Hidden Markov modeling (HMM) can be applied to extract single channel kinetics at signal-to-noise ratios that are too low for conventional analysis. There are two general HMM approaches: traditional Baum's reestimation and direct optimization. The optimization approach has the advantage that it optimizes the rate constants directly. This allows setting constraints on the rate constants, fitting multiple data sets across different experimental conditions, and handling nonstationary channels where the starting probability of the channel depends on the unknown kinetics. We present here an extension of this approach that addresses the additional issues of low-pass filtering and correlated noise. The filtering is modeled using a finite impulse response (FIR) filter applied to the underlying signal, and the noise correlation is accounted for using an autoregressive (AR) process. In addition to correlated background noise, the algorithm allows for excess open channel noise that can be white or correlated. To maximize the efficiency of the algorithm, we derive the analytical derivatives of the likelihood function with respect to all unknown model parameters. The search of the likelihood space is performed using a variable metric method. Extension of the algorithm to data containing multiple channels is described. Examples are presented that demonstrate the applicability and effectiveness of the algorithm. Practical issues such as the selection of appropriate noise AR orders are also discussed through examples.

INTRODUCTION

Hidden Markov modeling (HMM) provides an efficient approach for analysis of single channel currents. It is particularly useful for records where the signal-to-noise ratio is low or the channel kinetics is rapid. The conventional dwell-time approach (Colquhoun and Sigworth, 1995; Magleby and Weiss, 1990a, 1990b; Horn and Lange, 1983; Chay, 1988; Ball and Sansom, 1989; Qin et al., 1996, 1997) fails in these cases because an appropriate idealization of such data is often difficult. The HMM approach analyzes the noisy data directly, thereby eliminating the necessity of idealization. As a result, it has an improved requirement on signal-to-noise ratio and has the potential to allow for more rapid channel kinetics. In this approach, each sample point is treated as a relevant interval and the probability of obtaining the sequence of the observed data samples is calculated according to a model. The model contains two terms: an amplitude term that is used to estimate the probability that a given data point belongs to a given current level, and a transition probability term that is used to estimate the probability of a state transition between data points. The amplitude term is usually described by Gaussian distributions centered about each current level. The transition probability is described by a matrix exponential composed of the model rate constants and describes the probability of switching states between adjacent data points. The product of the probabilities over all possible paths can be computed recursively, producing the likelihood of observing the data given the model. The likelihood is then optimized with respect to the parameters of the model to produce the best fit. Because the noise is taken into account explicitly, the HMM approach permits a relatively low signal-to-noise ratio. It is also less prone to errors due to missed events because the transition probability takes into account the undetected transitions between adjacent samples. The penalty for using the sample-based likelihood paradigm is computation time, as the probabilities have to be evaluated at every data point.

Baum's reestimation, a precursor to the general expectation-maximization (EM) approach for maximal likelihood estimation, is the standard approach for estimating the parameters in a hidden Markov model. When applied to single channel analysis, the method, however, has several drawbacks. First, it assumes a perfect Markov signal with white background noise. For experiments like patch-clamp recording, this assumption is usually too restrictive. The data are always low-pass filtered. Furthermore, the instrument noise has a power spectrum that increases quadratically with frequency, and when the channel is open there is often additional noise arising from a variety of sources, and usually has unknown spectral characteristics. For channels with slow kinetics, the effects of such distortions may be negligible, but when the channel activity is busy, they become significant and need to be addressed explicitly. Another problem is that Baum's algorithm estimates discrete transition probabilities. The channel kinetics, however, is a continuous process and the quantities of interest are the rate constants. Although it is possible to convert the transition probabilities into rate constants, the model topology generally cannot be retained. Finally, in Baum's algorithm there is no explicit control of the rate constants, thus it is difficult to impose constraints such as detailed balance,

Received for publication 18 January 2000 and in final form 14 June 2000.
Address reprint requests to Dr. Feng Qin, Dept. of Biophysical Sciences, SUNY at Buffalo, 124 Sherman Hall, Buffalo, NY 14214. Tel.: 716-829-3289; Fax: 716-829-2028; E-mail: qin@acsu.buffalo.edu.

0006-3495/00/10/1928/17 $2.00

and to fit multiple data sets from different experimental conditions simultaneously.

The issue of correlated noise has recently been addressed by extending Baum's reestimation formulae. Walsh and Sigworth (1992) took the approach of preprocessing the data to pre-whiten the noise, thereby reducing the problem to solving a high-order Markov process in white noise; this, in turn, was reformulated into a first-order Markov process with an enlarged state space. More recently, Venkataramanan et al. (1998a, 1998b) extended this approach to account for additional noise in the open states and to estimate the noise correlations directly. To enable the direct optimization of rate constants, Fredkin and Rice (1992b) have attempted to use standard optimizers to maximize the likelihood function. Their approach, however, is reported to be very computationally intensive.

We have recently developed a direct optimization approach to hidden Markov modeling of single-channel currents. The approach has two essential features, i.e., the direct optimization of rate constants and the use of analytical derivatives for optimization of the likelihood function. These in turn provide the algorithm with several desirable features, such as the capability for explicit control of model topology, the allowance for imposition of constraints on rate constants, and the flexibility of simultaneous fitting of multiple data sets on different experimental conditions. The availability of the analytical derivatives of the likelihood function makes it possible to use the efficient gradient-based optimizers such as the variable metric method to search the likelihood surface. Compared with Baum's reestimation, the method has a favorable convergence near the maximum, especially in the extreme cases of low signal-to-noise ratio and aggregated kinetics. Under these conditions, Baum's algorithm often exhibits a slow convergence in its approach to the maximum by taking very small steps.

In this paper we extend this approach to address the issues of band-limited signals and correlated noise. Following Venkataramanan et al. (1998a, 1998b), we model the noise by an autoregressive (AR) process, so that the data can be reduced to a higher-order Markov process in white noise. However, we exploit a different strategy to parametrize the noise. Instead of estimating the AR coefficients, we choose to optimize the autocorrelations of the noise. Such a choice was motivated by the fact that the autocorrelations of two additive noises are linear combinations of the individual ones. Thus the relationship for the noise at different conductance levels, which is often additive, can be taken into account explicitly through the use of linear constraints. This feature is particularly useful for modeling multiple channel activity where the noise at different conductance levels is highly related. Other issues, such as the correction for low-pass filtering, the optimization of initial probabilities, and the handling of multiple channels, are also addressed. Finally, several prototype examples are provided to illustrate the features and performance of the algorithm.

THE MODEL

We consider a channel with N conformation states partitioned into M conductance classes. Let I_i, $i = 1 \ldots M$ denote the current amplitude at each conductance. Transitions among the states are described by a time-homogeneous Markov process with an infinitesimal generator matrix $\mathbf{Q} = [q_{ij}]_{N\times N}$, where the (i, j)th off-diagonal element q_{ij} represents the transition rate from state i to state j, and the diagonal elements are defined so that each row sums to zero. The system can be either in equilibrium or transient following perturbation. Although not necessary, the model is usually irreducible with states reachable from each other.

In practice we observe a discrete sampling of the continuous process. A sampled Markov process can be considered a Markov chain whose transitions are described by a transition probability matrix, say $\mathbf{A} = [a_{ij}]$, where a_{ij} represents the probability of the channel being in state j at the next sampling time given that it is in state i at the current sampling time. For a given sampling interval Δt, $\mathbf{A}$ is related to $\mathbf{Q}$ by

$$\mathbf{A} = \exp(\mathbf{Q}\Delta t). \tag{1}$$

The equation mathematically defines a one-to-one mapping between $\mathbf{Q}$ and $\mathbf{A}$. However, it is worth noting that the inversion from a given $\mathbf{A}$ to $\mathbf{Q}$ may not always produce physically meaningful results because some Markov chains cannot be considered to be sampled from a continuous Markov process (Qin et al., 2000).

The noise in the data is modeled conductance-wise. That is, all states with the same mean conductance are assumed to have the same noise characteristics, but the noise for different conductances is allowed to be different. For each conductance, the associated noise is modeled using an autoregressive (AR) process to account for its color. Let $n^{(i)}(t)$ denote the noise at the ith conductance, and σ_i and $a_j^{(i)}$, $1 \le j \le m$ be the corresponding AR coefficients. For simplicity, we assume the same order for all AR models at different conductances. An AR process can be considered as the output of passing white noise through an all-pole filter, i.e.,

$$n_t^{(i)} + a_1^{(i)} n_{t-1}^{(i)} + \cdots a_m^{(i)} n_{t-m}^{(i)} = \sigma_i w_t \tag{2}$$

where w_t is white noise with unit variance. From the functional approximation point of view, an AR model is equivalent to using an all-pole rational function to approximate the power spectrum of the noise. Because there are no zeros, the AR model is more restrictive than the more general ARMA model (Kay and Marple, 1981), but given sufficiently high order, it can fit a wide range of functions with either smooth or wavy features.

The effect of filtering is accounted using a finite impulse response (FIR) filter whose coefficients are denoted by h_j, $j = 0 \ldots n - 1$. We assume that the filtering is applied only to the underlying signal. In practice it affects noise too, but we leave that effect to be taken care of by the noise model.

We also assume in this paper that the coefficients of the filter are known a priori. They can be determined from the transfer function of the recording system and those of the digital filters that may have been used. It should be emphasized that although the model takes filtering into account explicitly, one should always preprocess the data before analysis to correct for bandwidth. This can be done through appropriate inverse filtering and decimation. The filter introduced here is mainly intended for the leftover effect that cannot be corrected a priori. A filter with as few coefficients as possible is important because a long filter, as seen later, will cause the computation time to increase exponentially.

The advantage of using an AR model for noise is that it allows the noise to be pre-whitened. Let s_t denote the state sequence of the underlying signal and x_t be the corresponding conductance class. The observation can be written algebraically as

$$Y_t = H(z)I_{x_t} + \frac{\sigma_{x_t}}{A^{(x_t)}(z)} w_t \tag{3}$$

where $H(z)$ is the transfer function of the FIR filter and $A^{(i)}(z)$ is the denominator of the transfer function of the AR filter, i.e.,

$$A^{(i)}(z) \equiv \sum_{j=0}^{m} a_j^{(i)} z^{-j}.$$

The pre-whitening is done by multiplying $A^{(x_t)}(z)$ through Eq. 3, leading to

$$Y_t = \sum_{j=0}^{p} c_j^{(x_t)} I_{x_{t-j}} - \sum_{j=1}^{m} a_j^{(x_t)} Y_{t-j} + \sigma_{x_t} w_t \tag{4}$$

where $p \equiv m + n - 1$ and $c_j^{(i)}$, $j = 0 \ldots p$ are the coefficients of the product $H(z)A^{(x_t)}(z)$, or equivalently, the convolution of $\{h_j\}$ and $\{a_j^{(i)}\}$. The significance of such pre-whitening is that it reduces the noise in the data to be white, as implied by Eq. 4, thereby satisfying the HMM assumption. This is of course achieved at the expense of further distorting the underlying signal. When the AR model is known a priori, the pre-whitening can be done experimentally. But for the problem here, it can only be done in theory because not only is the noise model unknown, but the system is time-variant, i.e., the noise model varies with the channel's conductance.

A potential problem with the AR model is the difficulty of imposing constraints on the noise between different conductance levels. For example, one might need to constrain the open noise so that it is the superimposition of the closed noise with an extra white component. The ability to allow for such constraints becomes particularly useful for modeling multi-channel records, as will be shown later. The AR model, however, is not linearly additive in the sense that the sum of two AR processes is generally no longer an AR process. We circumvent this by reparametrizing the noise using autocorrelations instead of AR coefficients. The autocorrelation function has the desirable feature that the autocorrelations of two independent random processes are linear combinations of the autocorrelations of the individual ones.

The AR coefficients of an mth order AR process are fully determined by its first m autocorrelations. This is stated by the Yule-Walker equation (Kay and Marple, 1981)

$$\begin{pmatrix} r_0^{(i)} & r_1^{(i)} & \cdots & r_m^{(i)} \\ r_1^{(i)} & r_0^{(i)} & \cdots & r_{m-1}^{(i)} \\ \vdots & \vdots & & \vdots \\ r_m^{(i)} & r_{m-1}^{(i)} & \vdots & r_0^{(i)} \end{pmatrix} \begin{pmatrix} 1 \\ a_1^{(i)} \\ \vdots \\ a_m^{(i)} \end{pmatrix} = \begin{pmatrix} \sigma_i^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{5}$$

where $r_j^{(i)}$, $j = 0 \ldots m$ are the autocorrelations of the noise at the ith conductance level. The equation can be solved using either a general linear equation solver or, more efficiently, the Levinson-Durbin algorithm (Blahut, 1985). The latter has a complexity on the order of m^2 as opposed to m^3. For a process that is not strictly AR, the above equation can still be used to determine the AR parameters. In this case the resulting AR model fits the first m autocorrelations exactly, and the remaining ones are extrapolated based on the maximal entropy criterion.

With the above parametrization, the entire model, denoted by Θ, constitutes of the rate constants (q_{ij}), the current amplitudes (I_i), and the noise autocorrelations ($r_j^{(i)}$). The problem is then to estimate all these parameters from the given observations. In the paradigm of HMM, the maximum likelihood approach is used. Let $Y = Y_1 \cdots Y_T$ denote the observed samples. The likelihood function, denoted by $L(\Theta)$, is defined as the probability of observing Y given Θ, i.e.,

$$L(\Theta) = \Pr(Y_1 \cdots Y_T | \Theta). \tag{6}$$

The problem is equivalent to finding the maximum point of $L(\Theta)$. In the next section we describe how to evaluate $L(\Theta)$ and its derivatives for efficient optimization.

THE LIKELIHOOD FUNCTION AND ITS DERIVATIVES

Evaluation of the likelihood and its derivatives is a complicated process. It can be roughly divided into two major parts. The first is to reformulate the problem in a way that satisfies the conventional HMM assumption, i.e., a first-order Markov chain in white noise. The standard approach for doing this is to introduce a metastate Markov model that combines the current state of the channel with its history states into a tuple (Fredkin and Rice, 1992a; Venkataramanan et al., 1998a, 1998b). Once the HMM assumption is satisfied, the existing theory can be applied. The other part

is to determine the parameters of the metastate Markov model and their derivatives. The parameters that are optimized are the rate constants, current amplitudes, and noise autocorrelations. Thus we need to evaluate, for example, the initial probabilities of metastates and their derivatives with respect to the rate constants, the transition probabilities of metastates and their derivatives with respect to the rate constants, the AR coefficients and their derivatives with respect to the noise autocorrelations, and so on. In the following, we describe each step in detail. At the end, we also discuss an alternative definition for metastates along with other strategies that can be used to improve the computational efficiency.

Metastate Markov model

The standard HMM places a restrictive assumption on both the signal and noise. The hidden signal needs to be first-order Markovian, while the noise must be white. We have shown in the previous section that the correlation of noise can be removed by pre-whitening the data. The resulting data, however, still don't satisfy the HMM assumption due to the history of the underlying signal, as implied by Eq. 4. One solution to the problem is to consider the current state of the channel along with its history states in the memory as a group. If the system has a memory of p lags, we simply lump together the current state s_t with the previous $p - 1$ states $s_{t-1}, \ldots, s_{t-p}$. This group, when considered as a new process, is memoryless. Mathematically, this is equivalent to defining a vector process

$$\mathbf{s}_t = (s_t, s_{t-1} \cdots s_{t-p}) \tag{7}$$

where each component represents a state of the channel. Taking the vector process $\mathbf{s}_t$ as the underlying signal, the observation Y_t becomes dependent only on the most current $\mathbf{s}_t$ at any time and independent of its histories, thereby satisfying the assumption of standard HMM.

The vector process $\mathbf{s}_t$ is Markovian because its transition at each time involves only a single transition of its first component, s_t, which is Markovian. The states of $\mathbf{s}_t$, called *metastates*, are the $(p + 1)$-tuples

$$(i_0, \ldots, i_p) \tag{8}$$

where each component is a possible state of the channel. If the channel has N states, $\mathbf{s}_t$ will have N^{p+1} metastates. While the number of metastates may be large, many transitions among them are disallowed. For two metastates $I = (i_0 \ldots i_p)$ and $J = (j_0 \ldots j_p)$, the transition between them is allowed if and only if

$$i_k = j_{k+1}, \quad k = 0 \cdots p - 1 \tag{9}$$

i.e., J corresponds to a shift of I toward the right by one component.

Before proceeding, we introduce some standard notation that will be used throughout the paper. As already mentioned above, we use capital I and J to denote metastates. The capital I is also used for channel current amplitudes, but its exact meaning should be clear from the context. The component states of a metastate are represented by the small i's or j's, as in Eqs. 8 and 9, and the corresponding conductance class of these states is designated using the Greek letter μ.

Evaluation of the likelihood

Considering the vector process $\mathbf{s}_t$ as the hidden signal, we have a new first-order Markov process in white noise, as seen from Eq. 4. The standard forward-backward procedure can then be applied to evaluate the likelihood function. The forward and backward variables are defined as the partial likelihood of the observation samples with a given metastate for the hidden signal, i.e.,

$$\alpha_t(I) = \Pr[Y_1 \ldots Y_t, \mathbf{s}_t = I]$$

$$\beta_t(J) = \Pr[Y_{t+1} \ldots Y_T, \mathbf{s}_t = J]$$

where $t = 1 \ldots T$, and I and J span over all possible metastates. Let $b_I(t)$ be the probability distribution of the observation at time t given the underlying process $\mathbf{s}_t$ in metastate I. Strictly speaking, $b_I(t)$ also depends on the previous observations $Y_{t-1}, Y_{t-2} \ldots Y_{t-m}$ in addition to the metastate I, but for simplicity of notation we will not exploit the dependence explicitly. From Eq. 4 it can be formulated as

$$b_I(t) = \frac{1}{\sqrt{2\pi}\sigma_{\mu_0}} \exp\left[-\frac{\omega_I^2(t)}{2\sigma_{\mu_0}^2}\right] \tag{10}$$

where $\omega_I(t)$ can be considered as the noise residue given by

$$\omega_I(t) = \sum_{j=1}^{m} a_j^{(\mu_0)} Y_{t-j} - \sum_{j=0}^{p} c_j^{(\mu_0)} I_{\mu_j} \tag{11}$$

and μ_j is the conductance class of the jth component of metastate I. The forward and backward variables are calculated recursively by

$$\alpha_{t+1}(J) = \sum_I \alpha_t(I) a_{IJ} b_J(t+1) \tag{12}$$

$$\beta_t(I) = \sum_J a_{IJ} \beta_{t+1}(J) b_J(t+1) \tag{13}$$

where a_{IJ} is the transition probability between metastate I and metastate J. The forward recursion in Eq. 12 proceeds from the first sample at $t = 1$ to the last sample at $t = T$, and the backward recursion goes in the opposite direction. Initially, the forward recursion starts with the initial probabil-

ity of the metastates, i.e., $\alpha_0(I) = \pi_I$, and the backward recursion begins with $\beta_T(J) = 1$ for all J.

By definition, the likelihood function is equal to the forward variables summed over all metastates at time T, i.e.,

$$L(\Theta) = \sum_I \alpha_t(I). \tag{14}$$

Therefore, the likelihood can be calculated at the end of the forward recursion. The backward variables, although not required for the evaluation of the likelihood, will be needed for the calculation of its derivatives. In practice, the likelihood itself is usually out of the machine range and only its logarithm can be computed. This is done through appropriate scaling of the forward and backward variables. For details about scaling see, for example, Rabiner (1989).

Derivatives of the likelihood

The derivatives of the likelihood function are calculated in multiple steps. First, we derive the derivatives of the likelihood with respect to the starting probability and transition probability of the metastate Markov model, and the derivatives with respect to the channel current amplitudes and noise AR coefficients. The derivation for these derivatives is similar to that for the case of white noise (Qin et al., 2000), so we will not go through the detailed algebra. The results can be summarized as

$$\frac{\partial L}{\partial \pi_I} = \beta_1(I) b_I(1) \tag{15}$$

$$\frac{\partial L}{\partial a_{IJ}} = \sum_t \alpha_t(I) \beta_{t+1}(J) b_J(t+1) \tag{16}$$

$$\frac{\partial L}{\partial x} = \sum_t \sum_I \alpha_t(I) \beta_t(I) \frac{\partial \ln b_I(t)}{\partial x} \tag{17}$$

where x in the last equation could be any variable that appears in the distribution function $b_I(t)$. For the problem here, x is either the current amplitude or an AR coefficient, but it can be other parameters as well. For example, one may include deterministic perturbations such as baseline drift or harmonic interference into the model. In these cases, the distribution function $b_I(t)$, which is the only part of the algorithm that needs to be modified, contains the additional perturbation parameters. The same formula can then be applied to derive the derivatives of the likelihood function with respect to those perturbation parameters.

Substituting $b_I(t)$ by its definition and letting x be the current amplitude and AR coefficient, we can further derive from Eq. 17 the derivatives of the likelihood function with respect to these variables as the following:

$$\frac{\partial L}{\partial I_k} = \sum_t \sum_j \sum_{\substack{I \\ \mu_j = k}} \alpha_t(I) \beta_t(I) \omega_I(t) \sigma_{\mu_0}^{-2} c_j^{(\mu_0)} \tag{18}$$

$$\frac{\partial L}{\partial \sigma_k} = -\sum_t \sum_{\substack{I \\ \mu_0 = k}} \alpha_t(I) \beta_t(I) \sigma_{\mu_0}^{-3} [\sigma_{\mu_0}^2 - \omega_I^2(t)] \tag{19}$$

$$\frac{\partial L}{\partial a_i^{(k)}} = -\sum_t \sum_{\substack{I \\ \mu_0 = k}} \alpha_t(I) \beta_t(I) \omega_I(t) \sigma_{\mu_0}^{-2} \left[Y_{t-i} - \sum_j h_j I_{\mu_{i+j}} \right] \tag{20}$$

where the coefficient $c_j^{(i)}$, $j = 0 \ldots p$ in the first equation represents the convolution of the filter impulse response $\{h_j\}$ with the AR coefficient $\{a_j^{(i)}\}$, as defined earlier. Note that there are a set of these equations for each conductance level, since each conductance has its own noise model. From the equations we see that the derivatives for the current amplitude are summed not only over all metastates but also over the individual components of each metastate, while the derivatives for the noise are summed only through the metastates. Such a discrepancy is as expected because the current amplitude information is contained in all samples in the memory, while the noise model, by definition, is only dependent on the first component of the metastate. Also note that setting the AR coefficient $a_j^{(i)}$ to zero and the filter impulse response to a delta function reduces the above equations to be the same as those in the case of white noise.

The derivatives given above constitute only a part of the calculation for the final derivatives of the likelihood function. Except for the current amplitudes, they are not given with respect to the model parameters that are expected to be optimized. For example, both the starting probabilities and transition probabilities of the metastate Markov model are functions of the rate constants, while the AR coefficients are functions of the noise autocorrelations. The next step in the calculation is to evaluate these intermediate parameters and their derivatives with respect to the true model parameters, i.e., the rate constants and autocorrelations. The final derivatives of the likelihood function are then obtained by combining them using the differentiation chain rule.

Transition probability of metastates

We consider the calculation of the transition probability of the metastate Markov model and its derivatives with respect to the rate constants. To this end we need to first calculate the transition probability of the channel itself. This can be done using the same procedure that we have developed for the standard HMM (Qin et al., 2000). The calculation is based on the spectral expansion of the rate constant matrix $\mathbf{Q}$

$$\mathbf{Q} = \sum_{i=1}^{N} \lambda_i \mathbf{A}_i \tag{21}$$

where λ_i is the ith eigenvalue of $\mathbf{Q}$ and $\mathbf{A}_i$ is the product of the corresponding left and right eigenvectors. Making use of the spectral expansion, the transition probability of the channel, which is a matrix exponential of the rate constant matrix as given in Eq. 1, can be represented by

$$\mathbf{A} = \sum_{i=1}^{N} \mathbf{A}_i e^{\lambda_i \Delta t}. \tag{22}$$

The derivatives of the transition probability with respect to the rate constants can be formulated as

$$\frac{\partial \mathbf{A}}{\partial q_{kl}} = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{A}_i \frac{\partial \mathbf{Q}}{\partial q_{kl}} \mathbf{A}_j f(\lambda_i, \lambda_j, \Delta t) \tag{23}$$

where $f(\lambda_i, \lambda_j, \Delta t)$ is a scalar function defined by

$$f(\lambda_i, \lambda_j, \Delta t) = \begin{cases} \dfrac{\exp(\lambda_j \Delta t) - \exp(\lambda_i \Delta t)}{\lambda_j - \lambda_i} & \text{if } \lambda_j \neq \lambda_i \\ \exp(\lambda_i \Delta t)\Delta t & \text{otherwise.} \end{cases} \tag{24}$$

The derivative $\partial \mathbf{Q}/\partial q_{kl}$ in Eq. 23 is simply a constant matrix, in which all elements are equal to zero except the (k, l)th entry, which is plus one, and the kth diagonal element, which is minus one. Because most elements of the matrix are zero, the matrix product in the summation in Eq. 23 can be calculated efficiently.

Once we have the transition probability of the channel, the transition probability of the metastate Markov model along with its derivatives is readily obtainable. In particular, Eq. 9 specifies the condition for an allowable transition, and when the condition is satisfied, the corresponding transition probability a_{IJ} is equal to the transition probability of the channel from state i_0 to state j_0, where i_0 and j_0 are the leading components of the two metastates, respectively. The derivative of the metastate transition probability a_{IJ} with respect to any transition probability a_{ij} of the channel is either 0 or 1, depending on whether the transition is allowed and whether i and j are equal to i_0 and j_0, respectively.

Initial probability of metastates

By definition, a metastate defines which state the channel is in from the current sampling time through the previous histories in the memory, so the starting probability of a metastate $\mathbf{I} = [i_0, i_1 \ldots i_p]$ is essentially the probability for the channel being in state i_p initially and then in state i_k at the subsequent sampling times for $t = k\Delta t$, $k = 1 \ldots p$. As with the calculation of the transition probabilities of metastates, we need to first calculate the starting probability of the channel itself in order to determine starting probabilities of the metastates.

The starting probability of the channel is chosen as the equilibrium probability at the holding condition. It is given by

$$\boldsymbol{\pi}^{\tau} \mathbf{Q}_h = 0 \tag{25}$$

where $\mathbf{Q}_h$ is the rate constant matrix at the holding condition. In general, the rate constants at the holding condition are related to the rate constants at the activating condition, and they can be reformulated in a way to share a common set of independent parameters. This point is discussed further under the topic of likelihood maximization. Thus, when $\mathbf{Q}$ is optimized, $\mathbf{Q}_h$ will follow the change, and as a result, the starting probability of the channel is also optimized.

Equation 25 is homogeneous and needs to be solved by subjecting it to the probability totality constraint. The singular value decomposition technique can be applied for finding the solution (Press et al., 1992). It is possible that the model may become reducible at some holding conditions, in which case only the probabilities of the absorbing states are solved and others are fixed to zero. From Eq. 25 we can derive the derivatives of the initial probability as

$$\frac{\partial \boldsymbol{\pi}^{\tau}}{\partial x} \mathbf{Q}_h = \boldsymbol{\pi} \frac{\partial \mathbf{Q}_h}{\partial x} \tag{26}$$

where x could be any variable of interest. Notice that Eqs. 25 and 26 share a common coefficient matrix. Therefore, they can be solved together. One difference to note, however, is the constraint. For solving Eq. 26, namely, the sum of the unknowns is constrained to zero instead of one.

Given the starting probability of the channel and its subsequent transition probabilities, the starting probability of a metastate $\mathbf{I} = [i_0, i_1 \ldots i_p]$ can be determined by

$$\pi_I = \pi_{i_p} a_{i_p i_{p-1}} a_{i_{p-1} i_{p-2}} \cdots a_{i_1 i_0} \tag{27}$$

i.e., the probability of the channel entering state i_p at $t = 1$, multiplied the probability of a transition from state i_p to state i_{p-1} at $t = 2$, and then multiplied by the probabilities of the subsequent transitions until $t = p + 1$.

To derive the derivatives of the starting probabilities of metastates, we introduce two auxiliary quantities u_k, $0 \leq k \leq p$ and v_l, $0 \leq l \leq p$ where

$$u_k = u_{k-1} a_{i_k i_{k+1}} \tag{28}$$

$$v_l = a_{i_l i_{l+1}} v_{l-1} \tag{29}$$

with $u_0 = \pi_{i_p}$ and $v_p = 1$. That is, u_k represents the forward partial product in Eq. 27 and v_k the backward one. The derivative of π_I with respect to any variable x can then be written as

$$\frac{\partial \pi_I}{\partial x} = \frac{\partial \pi_{i_0}}{\partial x} u_1 + \sum_{k=1}^{p} u_{k-1} \frac{\partial a_{i_{k-1} i_k}}{\partial x} v_k \tag{30}$$

where the derivatives of the channel starting probability and transition probability are given by Eqs. 26 and 23, respectively. Depending on the size of a metastate, i.e., the length of the filter plus the AR order, the calculation of u_k and v_k above may require appropriate scaling to prevent underflow. The scaling can be done following the same strategy used in the calculation of the likelihood function.

AR coefficients

The AR coefficients are determined from the autocorrelation variables according to the Yule-Waker equation (5). The equation can be solved efficiently using the Levinson-Durbin algorithm (Blahut, 1985). We now give a brief description of the algorithm and then show how to modify the algorithm to also calculate the derivatives of the AR coefficients. For clarity, we suppress the sub or superscripts for the conductance class. Let $r_0, r_1 \ldots r_p$ be the autocorrelations of the noise at a certain conductance. The algorithm calculates the coefficients of all AR models with orders up to p, denoted by $\{a_{11}, \sigma_1^2\}, \{a_{21}, a_{22}, \sigma_2^2\} \ldots \{a_{p1}, a_{p2}, \ldots, a_{pp}, \sigma_p^2\}$. The final set at order p is the desired solution. Starting with $\sigma_0^2 = r_0$, the algorithm proceeds recursively for $k = 1, 2 \ldots p$ as below:

$$a_{kk} = -\frac{1}{\sigma_k^2} \sum_{j=0}^{k-1} a_{k-1,j} r_{k-j} \tag{31}$$

$$a_{kj} = a_{k-1,j} + a_{kk} a_{k-1,k-j}, \quad j = 1 \cdots k-1 \tag{32}$$

$$\sigma_k^2 = (1 - a_{kk}^2)\sigma_{k-1}^2 \tag{33}$$

that is, it first computes the leading coefficient a_{kk}, then uses it to compute the remaining a_{kj} values for $j = 1 \ldots k-1$, and finally the variance σ_k^2. The recursion terminates when the desired order is reached. The algorithm implicitly assumes that $|a_{kk}| \leq 1$ for all k in order to carry out the recursion for σ_k^2. This is true for a strict AR process or when the autocorrelation matrix is positive definite. For an arbitrary process, the assumption may fail, in which case the algorithm should be terminated when σ_k^2 becomes sufficiently small.

To calculate the derivatives of the AR coefficients, we differentiate the above recursive equations, leading to

$$\frac{\partial a_{kk}}{\partial x} = -\frac{1}{\sigma_{k-1}^2}\left[a_{k,k} \frac{\partial \sigma_{k-1}^2}{\partial x} + \sum_{j=0}^{k-1}\left(\frac{\partial a_{k-1,j}}{\partial x} r_{k-j} + a_{k-1,j} \frac{\partial r_{k-j}}{\partial x} \right)\right] \tag{34}$$

$$\frac{\partial a_{kj}}{\partial x} = \frac{\partial a_{k-1,j}}{\partial x} + \frac{\partial a_{kk}}{\partial x} a_{k-1,k-i} + a_{k,k} \frac{\partial a_{k-1,k-i}}{\partial x}, \quad j = 1 \ldots k-1 \tag{35}$$

$$\frac{\partial \sigma_k^2}{\partial x} = -2a_{k,k} \frac{\partial a_{kk}}{\partial x} \sigma_{k-1}^2 + (1 - a_{kk}^2) \frac{\partial \sigma_{k-1}^2}{\partial x} \tag{36}$$

where x represents an autocorrelation variable. The equations suggest that we can calculate the derivatives of the AR coefficients basically in the same way that the coefficients themselves are calculated. That is, the calculation is carried out recursively through the order, and the derivatives at the kth order are calculated from those at $(k-1)$th order. At each stage, the same recursions need to be repeated for each variable x. Because the original Levinson-Durbin recursion has a complexity on the order of p^2, the calculation of the derivatives for all variables will take on the order of p^3 operations, which is about the same as the complexity of a general linear equation solver. In practice, the time needed for the computation of the AR coefficients and their derivatives is negligible compared to the evaluation of the likelihood, because the AR order is usually small.

Computational complexity

Although the calculation of the likelihood function and its derivatives involves many steps, the most time-consuming part is the calculation of the forward and backward variables and the calculation of the derivatives of the likelihood with respect to the transition probabilities. Each of these steps takes on the order of $N^{2(p+1)}T$ operations, where N is the number of states of the channel and p is equal to the order of the noise AR model plus the filter length. The time for the rest of the calculations is negligible. Therefore, the overall computational complexity for the calculation of the likelihood and its derivatives is on the order of $N^{2(p+1)}T$. Such a complexity increases exponentially with the filter length or the AR order, which is the major limiting factor for the practical applicability of the algorithm.

There are several maneuvers that can be used to cut down the computation. One is to take advantage of the aggregation property of the channel in the definition of the metastate. Specifically, we can use the state only for the first component of a metastate while defining others to be conductance class. That is

$$\mathbf{s}_t = (s_t, x_{t-1} \ldots x_{t-p}) \tag{37}$$

where x_t represents the conductance class of the state sequence s_t. Such a new vector process remains Markovian because its transition is fully determined by the first component. The state space of the new process, however, is much smaller. If the channel has N states and M conductances, the number of metastates using this definition will be NM^p as opposed to N^{p+1} with the old definition. Because channels often have only two conductances, such a reduction is significant, especially considering that the computation is quadratic on the number of metastates. In the implementation of our algorithm we have exploited this

definition. The reason that we have used the other definition in the above is mainly for clarity of the presentation.

Another operation that may speed up the algorithm is to make use of the fact that many transitions between metastates are disallowed. As a consequence, the summation in the forward and backward recursions does not need to be extended through all possible metastates. Instead, we can restrict it only to the allowable transitions. As stated above, a transition between two metastates is allowed if and only if they satisfy Eq. 9. Therefore, there are only N allowable transitions for each metastate. The summation in the forward and backward recursions needs to be performed for only N times instead of NM^p. This is true for every metastate at each time t, thus leading to a total reduction in the computation by a factor of M^p, which is considerable as p becomes large.

MAXIMIZATION OF THE LIKELIHOOD

We have shown how to evaluate the likelihood and its derivatives for a given particular set of model parameters. The next step is to optimize the model to search for the maximum of the likelihood function. This is essentially an optimization problem and can be solved in a variety of ways. We have used the same approach that we have developed for the maximization of the likelihood with the standard HMM (Qin et al., 2000). The following is a brief outline of some major features of the approach.

The search of the likelihood space is performed using a quasi-Newton method (Fletcher, 1980, 1981). It is based on successive approximation of the likelihood surface with parabolas and uses an approximate inverse Hessian matrix built from the first-order partial derivatives. The main advantage of the method is the quadratic convergence near the maximum point because the likelihood surface there can be well-approximated by a parabola. It also avoids the need for exact line minimization. An approximate line minimization usually takes one to two function evaluations, but an accurate line minimization may require repetitive braking, and therefore many function calls. When the evaluation of the objective function is computationally costly, this may significantly degrade the overall efficiency of the algorithm. Another feature of the method is that it results in an estimate of the inverse Hessian matrix, which can be used to derive the standard errors of the parameter estimates.

Constraints on the model parameters are allowed. The constraints on rate constants include holding rates at fixed values, linear scaling between two rates, detailed balance, and so on. The current amplitudes can be linearly scaled or fixed. The noise can be constrained too, for example, to restrict the excess open noise to be white or the noise at different conductance to have the same amount of correlation. These constraints are either linear or can be converted to be linear. Therefore, they can be handled analytically by formulating the constrained variables into linear combinations of a set of unconstrained ones. As a result, the originally constrained problem is reduced to an equivalent unconstrained one, making the optimization more efficient.

The approach also allows multiple data sets obtained at different experimental conditions to be fit simultaneously. The feature not only allows more data to be used for modeling, but also becomes required for resolving models that degenerate at individual conditions. In our implementation, the rates in the model are represented explicitly as a function of the experimental condition such as ligand concentration, voltage, or force. The generic rate constants that are independent of experimental conditions are chosen as the variables to optimize. Such a parametrization of the rates also makes it possible to specify the initial probability of the channel by holding experimental conditions, which is particularly useful for channels not under equilibrium.

EXTENSION TO MULTIPLE CHANNELS

The expression of ion channels in membranes is often high, making the recording of currents from an individual channel difficult. In many cases one can only obtain recordings arising from multiple channels. The algorithm described above can be extended readily for the analysis of such data.

The kinetics of multi-channel activity remains Markovian provided that the individual channels act independently. Consider a patch containing K channels. Let $s_t^{(i)}$ denote the state sequence of the ith channel at time t. The configuration of the system at any time can be characterized by a K-tuple

$$\upsilon_t \equiv (s_t^{(1)}, s_t^{(2)}, \ldots, s_t^{(K)}) \tag{38}$$

which can be considered as a new Markov process. Because each component of the tuple has a finite number of states, the tuple itself also has a finite number of states. Specifically, if $s_t^{(i)}$ has N_i states, the tuple has $N_1 N_2 \ldots N_K$ states. The first-order transitions between two states occur if and only if there is one channel in the patch making a transition and the corresponding rate constant is the same as that of the underlying channel.

For identical channels, Horn and Lang (1983) suggest a more efficient definition, where the tuple has the form

$$\mathbf{n}_t = (n_t^{(1)}, n_t^{(2)}, \ldots, n_t^{(K)}) \tag{39}$$

where $n_t^{(i)}$ is the number of channels in state i at time t. Such a definition takes the advantage of the indistinguishability of the composition channels and therefore results in a smaller number of tuple states. For example, if the channel has N states, $\mathbf{n}_t$ has only $\binom{N+K-1}{N}$ instead of N^K states. The condition for an allowable transition between two tuple states stays the same, i.e., only one channel in the patch undergoes a transition at the same time. The transition rate, however, is different and needs to be multiplied by the number of the channels at the starting state of the transition.

For a detailed description of the two approaches, see Qin et al. (1996).

The observations for the tuple process can also be parametrized in terms of those for the single channels. For illustration, we consider a patch containing identical channels and each channel has three conductance classes, namely closed (C), partially open (S), and fully open (O). Suppose there is a tuple with n_C channels closed, n_S channels partially open, and n_O channels fully open. The current amplitude of the tuple state is given by

$$nI_C + n_S(I_S - I_C) + n_O(I_O - I_C) \tag{40}$$

where I_C, I_S, and I_O are the current amplitudes of the single channel. The first term is the current at the baseline, the second is contributed by the channels at the substates, and the third due to the channels at the fully open states. Similarly, the noise autocorrelations corresponding to the tuple state are given by

$$r_j^{(C)} + n_S(r_j^{(S)} - r_j^{(C)}) + n_O(r_j^{(O)} - r_j^{(C)}) \tag{41}$$

where $\{r_j^{(C)}\}$, $\{r_j^{(S)}\}$ and $\{r_j^{(O)}\}$ are the autocorrelations of the noise associated with single channels at the three conductance levels.

From this example it is seen that both the current amplitudes and noise autocorrelations for the tuple states are linear combinations of the singles. This is generally true and applies to both identical and nonidentical channels. It should be mentioned that such linear relationships are retained because of the use of autocorrelations to parametrize the noise. If the AR coefficients were used, the relation would be more complicated and generally cannot be specified analytically because the AR functions are not closed under the additive operations.

It is worth noting that the definition of a conductance class is sometimes ambiguous when multiple channels are present. For a single channel, the states are classified in terms of their observable current amplitudes, and the states with the same current amplitude are assumed to have the same noise properties. But for multiple channels, two tuple states may have the same current amplitude but different noise characteristics. Therefore, the current amplitude alone is not enough for classification. There are several solutions to the problem. One is to treat each state as a single class. The disadvantage of doing so is that it may result in an excessive number of metastates, because the redundancy between the states is not taken into account. A more efficient approach, which we have used in our implementation, is to use the number of channels at each conductance as a criterion. For the above example, the criterion would be the tuple (n_C, n_S, n_O) where n_C, n_S, and n_O are the number of channels in the patch at the three conductances, respectively. All tuple states with the same index are guaranteed to have the same current amplitude and noise property.

With the tuple process defined in Eq. 38 or 39 as the underlying Markov process, the multiple channel data can be fitted in the same way as the singles. The only difference is the constraints, since the parameters for the tuple process are not all independent. Instead, they are linear combinations of the singles, as shown above. To minimize the degrees of freedom, we have chosen to optimize the single channel variables directly. Fortunately, all constraints, including those for the noise, are linear, and therefore can be handled easily using the same approach described in the previous section.

EXPERIMENTAL RESULTS

We present a few examples to show some basic features of the HMM approach described in the previous sections. A set of four examples is given. The first example illustrates the general applicability of the algorithm to the typical correlated patch-clamp background noise. In the second example, we introduce excess open noise, both white and correlated, in addition to the correlated background noise. The third example tests the algorithm with filtered data. The last example demonstrates the applicability of the algorithm to data containing multiple channels.

The algorithm is implemented in a general way to allow for both single channels and multiple channels, with either white or correlated noise. The input to the program is a single channel model along with the number of channels. Internally, the program first builds a multi-channel Markov model for the tuple process, as defined above. In the case of a single channel, this multi-channel model reduces to the single one, but in the case of multiple channels, a set of linear constraints are set up automatically on the rate constants, current amplitudes, and noise autocorrelations. Following the multi-channel Markov model, the algorithm then builds a metastate Markov model based on the input filtering and noise correlation. The noise at different conductance levels can be constrained to be either the same or to differ by an excess variance that can be either white or correlated.

Colored background noise

We first consider an example with simulated patch-clamp noise. The power spectrum of the background noise in a typical patch-clamp experiment can be described by (Sigworth, 1995)

$$S(f) = a + bf^2 \tag{42}$$

over the experimentally accessible frequency range from dc to 100 kHz. The frequency-independent component is mainly due to the shot noise of resistors in the amplifier. The f^2-dependent component arises from current fluctuations resulting from the amplifier voltage noise imposed on the pipette and amplifier stray capacitance. In discrete time,

Venkataramanan et al. (1998a) show that the noise can be approximated by a first-order moving average (MA) process

$$n(t) = m_0 w(t) + m_1 w(t-1) \tag{43}$$

where $w(t)$ is white, Gaussian noise with zero mean and variance σ_w^2. Typical values for coefficients are $m_0 = 0.8$ and $m_2 = -0.6$, obtained by fitting a representative patch-clamp recording without channel activity.

Although the noise is a simple first-order MA process, it requires a more complicated AR model for a good approximation. A first-order MA process has a correlation extending over only one sample point, but a first-order AR process usually correlates over an infinite number of samples. Fig. 1 shows the power spectrum of the first-order MA model given in Eq. 43 in comparison with its AR approximations with various orders. From the figure we see that a first-order AR doesn't have a sufficient capacity to model the noise. But as the order increases, the approximation becomes more accurate. In theory, the approximation can be made into an arbitrary accuracy with a sufficiently large order. In practice, however, the exponentially increasing computation burden limits the use of very high-order models.

The data were simulated based on a two state model:

$$\mathrm{C} \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} \mathrm{O} \tag{I}$$

where the two rate constants were $k_{12} = 38310\ \mathrm{s}^{-1}$ and $k_{21} = 12770\ \mathrm{s}^{-1}$. A discrete version of this example was also considered by Venkataramanan et al. (1998a). The current amplitudes were chosen to be 0 pA for the closed state and 1 pA for the open state. The noise was generated according to Eq. 43 with a total variance $\sigma_w^2 = 0.64$. The data were sampled at 100 kHz, and a total number of 100,000 samples were simulated, giving a record of 1 s duration. Fig. 2 illustrates a stretch of 1000 samples of the signal before and after being superimposed with noise.

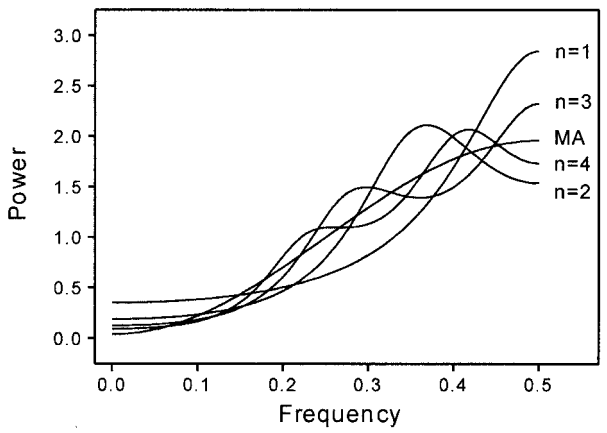

FIGURE 1 Power spectrum density of a first-order MA process in comparison with its AR approximations at different orders. The AR model is determined by solving the Yule-Walker equation. The approximation improves as the order increases.

For analysis, we used a third-order AR model to account for noise correlation. Thus the noise is parametrized by its first four autocorrelations. The program was tested in three different ways: 1) the open channel noise and the closed channel noise were assumed to be the same; 2) the open channel noise was assumed to have the same correlation as the closed noise, but a different variance; and 3) the two noises were assumed to be independent. The first case has the strongest constraint, i.e., $r_j^{(C)} = r_j^{(O)}$ for all $0 \le j \le 3$; the second imposes the constraint only for $1 \le j \le 3$; and the third has no constraint, in which case the noise model contains as many as eight unknown parameters. The algorithm performed well in all three cases. Table 1 illustrates the resulting estimates of the parameters obtained in the third case, which is most difficult. The standard deviations on the estimates were calculated from the approximate Hessian matrix generated by the optimizer. It is seen that all estimates are close to their true values within reasonable error. Also shown in the table are the starting values of the parameters, where the noise was initially set to be white. The corresponding maximum log likelihood value is $-123{,}243.85$.

Fig. 3 compares the power spectrum density of the noise with that predicted by the algorithm. The actual spectrum was calculated from a stretch of 2048 noise samples using Welch's averaged periodogram method (Oppenheim and Schafer, 1975). The predicted curves were calculated from the resultant AR models at the two conductances. The three power spectrums all fit reasonably well, as expected.

The program took about 21 iterations and 31 likelihood evaluations when the noise was most constrained, and 32 iterations and 48 likelihood evaluations when no constraints were imposed. With fewer constraints, the problem contains more independent variables, and therefore the convergence becomes more difficult. In general, the convergence appears to be more difficult with correlated noise than with white noise. The optimizer often tends to take a small step size even when the gradients of the likelihood function are large. The line search at each iteration also often requires more than one likelihood evaluation, suggesting that the likelihood surface is more complicated than a parabola. The convergence near the optimum, however, remains rapid.

In practice, the order of the noise model is unknown a priori. One possible solution is to select it retrospectively. We tested the algorithm with different AR orders. Table 2 shows the resulting estimates of the parameters. The same starting values specified in Table 1 were used. The algorithm failed when a white noise model was used. Even starting from the correct parameter values, the two current amplitudes merged with each other. Such a phenomenon was also observed with other examples when the noise model was widely wrong. The use of a first-order AR, however, seemed to be able to generate acceptable estimates

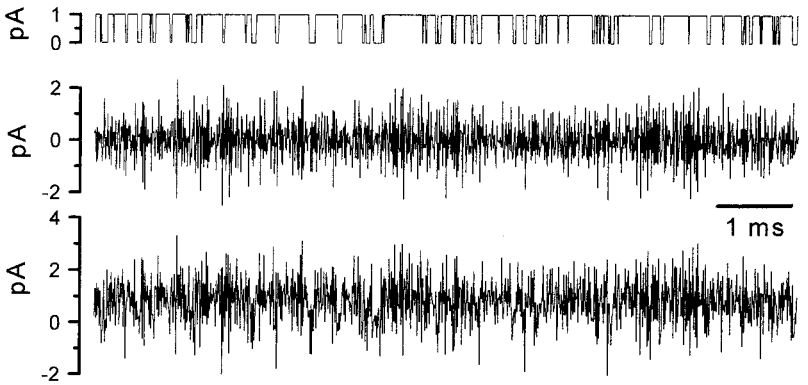

FIGURE 2 Single channel currents. The top trace shows the ideal current simulated from a two-state model with the current amplitudes at 0 and 1 pA, respectively. The middle trace shows the noise simulated from a first-order MA model with a total variance 0.64. The bottom trace is superimposition of the signal and the noise.

for both rate constants and current amplitudes. As the AR order further increases, the accuracy improves, but not dramatically. The fact that a first-order AR suffices to produce a reasonably good estimate suggests that the parameter estimation is not sensitive to the exact detail of the underlying noise model. This is particularly evident in that a first-order AR model can only poorly approximate the power spectrum of the noise, as shown in Fig. 1.

The determination of an appropriate AR order for noise is essentially a model identification problem. Therefore, a conceivable approach is to use the likelihood as a testing criterion and choose the one with the best likelihood. In practice, this doesn't seem to be efficient. From Table 2, the log likelihood improved by as many as 5763, 2403, 914, and 400 natural log units, successively, as the order increased from 0 to 4. Because increasing the order by one introduces only one extra degree of freedom, these improvements would all be ranked to be statistically significant according to the common model identification criteria, such as the χ^2-distribution. This is true in the sense that the underlying noise is not an AR but a MA process, and therefore requires a large AR model to accurately identify it. However, there is little improvement in the estimates of current amplitudes and rate constants. Because the computation time increases exponentially with the AR order, the lowest order would be preferred in practice. Therefore, a more practical criterion to choose the AR order is to look at the change in the parameter estimates rather than the likelihood, unless the structure of the state noise is essential.

TABLE 1 Parameter estimates with correlated background noise

Parameter	True Value	Initial Value	Estimate	S.D.
k_{12}	38,310	4000	37,063	647
k_{21}	12,770	1000	11,951	209
I_C	0	0.2	0.0023	0.0043
I_O	1	0.8	0.9899	0.0017
$r_0^{(C)}$	0.64	0.5	0.6271	0.0077
$r_1^{(C)}$	−0.3072	0	−0.2981	0.0052
$r_2^{(C)}$	0	0	0.0120	0.0058
$r_3^{(C)}$	0	0	0.0114	0.0058
$r_0^{(O)}$	0.64	0.5	0.6485	0.0042
$r_1^{(O)}$	−0.3072	0	−0.3087	0.0027
$r_2^{(O)}$	0	0	0.0014	0.0031
$r_3^{(O)}$	0	0	0.0122	0.0030

Excess open noise

When a channel is open, there is an excess amount of current noise. This has been considered to arise from the translocation of ions hopping through the channel pore and the fluctuation of channel structures. In some experiments this excess open current noise is observed to be frequency-independent, presumably because the experimentally accessible frequency is much lower than the dynamics of underlying physical processes. In other experiments, however, the noise is observed to be frequency-dependent. For example, the spectral analysis of the single channel currents from the ACh receptor shows brief channel blockages produced by ACh (Sine et al., 1990). The NMDA receptor channel also shows substates with experimentally measurable Lorentzian

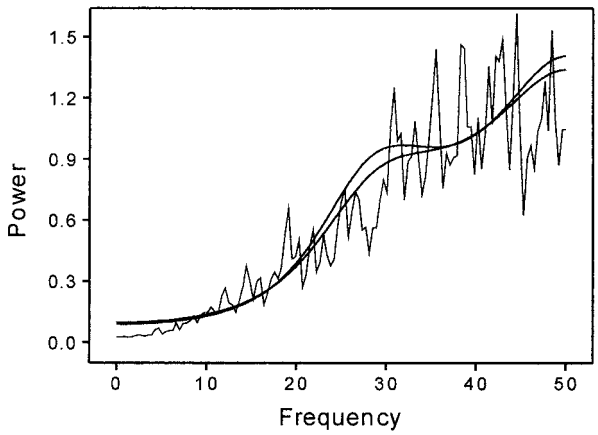

FIGURE 3 Power spectrum density of background noise. The noisy curve is the periodogram estimate from the noise sampled directly. The two smooth curves, one for the closed channel noise and the other for the open channel noise, are calculated from the AR models provided by the algorithm.

TABLE 2 Comparison of parameter estimates and maximum likelihood with different AR orders

n	k_{12}	k_{21}	I_C	I_O	r_0	r_1	r_2	r_3	r_4	LL
0	42,976	3512	0.75	0.75	0.826					−132,323
1	38,277	10,806	0.00	0.96	0.664	−0.284				−126,560
2	37,237	11,674	0.00	0.98	0.646	−0.302	0.011			−124,157
3	37,063	11,951	0.00	0.99	0.627	−0.298	0.012	0.011		−123,244
4	37,015	12,065	0.00	0.99	0.643	−0.308	0.000	0.007	0.003	−122,838

noise due to a high-affinity divalent cation-binding site (Premkumar and Auerbach, 1996).

The algorithm allows for both white and correlated excess open noise. We first consider an example where the excess noise is white. The simulation was the same as above except an extra white noise component was added when the channel was open. The standard deviation of this excess noise was 0.3 pA. Testing was performed with the constraint $r_j^{(C)} = r_j^{(O)}$ for $1 \le j \le 3$, i.e., the noise at the two conductances is assumed to have the same correlation. The same starting values were used as in the previous example. The program took about 38 iterations and 69 likelihood evaluations to converge. Table 3 lists the resultant estimates of the parameters along with their error estimates obtained from the approximate Hessian matrix. From the difference of the autocorrelations of the noise at the two conductance levels, we calculate the excess open noise variance $\sigma_E^2 = r_j^{(O)} - r_j^{(C)} = 0.0778$, or equivalently the standard deviation $\sigma_E = 0.28$ pA. It is seen that all parameter estimates agree with their true values to a good accuracy. A similar example but with smaller excess noise was also considered by Venkataramanan et al. (1998b) using an extended Baum's reestimation. Compared to their approach, the algorithm has a comparable convergence performance, but has the advantage of finding the true maximum of the likelihood while the extended Baum's algorithm usually converges only to a near-maximum.

The algorithm was also tested with correlated excess open noise. A first-order Lorentzian noise with a corner frequency f_c was used. In discrete time, the noise can be modeled by passing white, Gaussian noise with zero mean and variance σ^2 through a first-order AR filter:

$$n(t) + an(t-1) = w(t) \tag{44}$$

where $a \approx \exp(-f_c\Delta t)$. The noise was simulated with $\sigma = 0.1$ and $a = 0.90$, corresponding to a corner frequency $f_c = 10$ kHz at a sampling frequency 100 kHz. The total variance of the noise, given by $\sigma^2/(1 - a^2)$, was approximately $r_0^{(E)} = 0.0526$. The rest autocorrelations of the noise can be determined from the Yule-Walker equation (5), and the first three terms have values $r_1^{(E)} = -0.0474$, $r_2^{(E)} = 0.0426$, and $r_3^{(E)} = -0.0384$. The overall open noise then have autocorrelations $r_0^{(O)} = 0.6926$, $r_1^{(O)} = -0.3544$, $r_2^{(O)} = 0.0426$, and $r_3^{(O)} = -0.0384$, which are simply the superposition of the autocorrelations of the excess noise with those of the background noise. The estimation was done by treating the noise at the two conductances to be independent, i.e., no constraints were imposed on the autocorrelations. Table 4 shows the resultant estimates along with their standard deviations. The difference of the autocorrelations between the two conductances gives the estimates of the autocorrelations of the excess noise, which are $\hat{r}_0^{(E)} = 0.0734$, $\hat{r}_1^{(E)} = -0.0564$, $\hat{r}_2^{(E)} = 0.03625$, and $\hat{r}_3^{(E)} = -0.04387$. From these estimates we calculate the coefficients $a \approx 0.874$ and $\sigma \approx 0.155$, which are reasonably close to the true values. The algorithm took about 40 iterations with 55 likelihood evaluations to converge in this case.

Low-pass filtering

In general, the HMM can accommodate a large background noise. It is therefore preferable to record the data with minimal filtering. For data already low-pass filtered, the strategy is to undo the filtering by passing it through an inverse filter. This reduces the number of samples on the rising or falling edges of channel openings and thereby minimizes distortion. For illustration, we consider the same example as the first one, where the simulated patch-clamp noise had a total variance 0.64. The data were low-pass filtered at $f_c = 10$ kHz using a Gaussian filter. Given the sampling frequency 100 kHz, this filtering resulted in about eight samples on the rising phase of the step response.

For analysis, we first unfiltered the data to extend the bandwidth up to 30 kHz. The inverse filter was designed so that its combination with the low-pass Gaussian filter has a net effect similar to that of a 30 kHz Gaussian filter. The actual design was carried out using linear programming, so that both frequency- and time-domain constraints were allowed. The frequency-domain constraints were the devia-

TABLE 3 Parameter estimates with white excess open channel noise

Parameter	True Value	Initial Value	Estimate	S.D.
k_{12}	38,310	4000	37,668	713
k_{21}	12,770	1000	12,691	303
I_C	0	0.2	0.011	0.0040
I_O	1	0.8	0.997	0.0026
$r_0^{(C)}$	0.64	0.5	0.652	0.0051
$r_1^{(C)} = r_1^{(O)}$	−0.307	0	−0.306	0.0033
$r_2^{(C)} = r_2^{(O)}$	0	0	0.003	0.0029
$r_3^{(C)} = r_2^{(O)}$	0	0	0.009	0.0029
$r_0^{(O)}$	0.73	0.5	0.730	0.0042

TABLE 4 Parameter estimates with correlated excess open channel noise

Parameter	True Value	Initial Value	Estimate	S.D.
k_{12}	38,310	4000	38,242	708
k_{21}	12,770	1000	12,428	243
I_C	0	0.2	−0.0069	0.00407
I_O	1	0.8	0.9939	0.00158
$r_0^{(C)}$	0.64	0.5	0.6239	0.00769
$r_1^{(C)}$	−0.307	0	−0.2969	0.00507
$r_2^{(C)}$	0	0	0.00525	0.00551
$r_3^{(C)}$	0	0	0.01897	0.00538
$r_0^{(O)}$	0.6926	0.5	0.6973	0.00499
$r_1^{(O)}$	−0.3544	0	−0.3533	0.00328
$r_2^{(O)}$	0.0426	0	0.0415	0.00360
$r_3^{(O)}$	−0.0384	0	−0.0249	0.00357

tions of the frequency response, and the time-domain constraints were the overshoot of the step response. Fig. 4 shows the overall step response of the inverse filter in combination with the original Gaussian filter. The small amount of ringing arises because the inversion was done only up to the frequency where the attenuation of the original Gaussian filter was no lower than 10^{-3} (−60 dB). Beyond that, the frequency response of the 10 kHz Gaussian filter was attained.

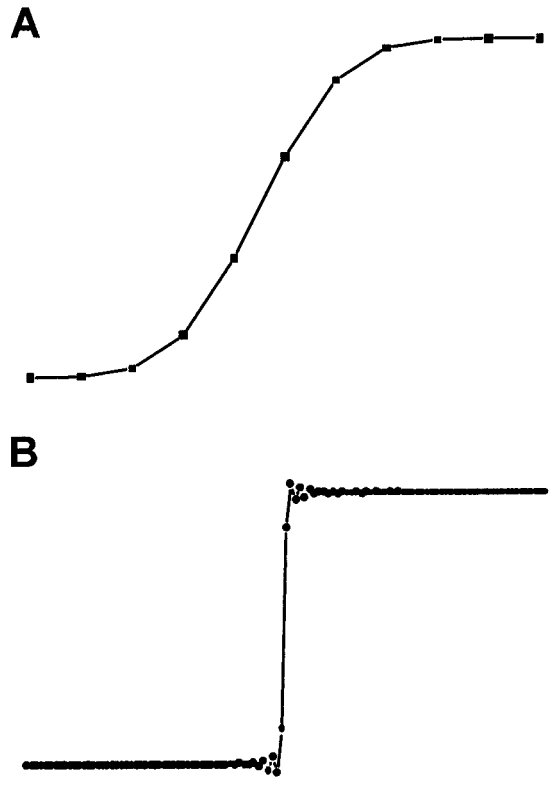

FIGURE 4 (*A*) Step response of the 10 kHz Gaussian filter applied to the data. (*B*) Step response after inverse filtering. The inverse filter has a length 100 and a cutoff frequency at 30 kHz. The overshoot of the step response was limited to 3%.

The unfiltered data still have a limited rise time, but extend for only about three samples. The data were analyzed in two different ways. In the first approach we modeled the finite rising time explicitly. That is, we introduced a second-order FIR on the underlying signal. The coefficients of the filter were chosen so that its step response gave the same three points on the rising phase as shown in Fig. 4. This leads to $h_0 = 0.13$, $h_1 = 0.74$, and $h_2 = 0.13$. The noise in the inverted data was modeled using a second-order AR, which was constrained to have the same amount of correlation but an independent variance at the two conductances. Table 5 lists the estimation results, obtained in 27 iterations and 38 likelihood evaluations. Both rate constants and current amplitudes of the channel were recovered successfully. The noise estimates do not match their original values used in simulation; this is as expected because the noise has been low-pass filtered. Fig. 5 shows the spectrum of the noise after inverse filtering and the spectrum represented by the resultant AR model. The AR spectrum is not a great fit, but suffices to provide successful recovery of all channel parameters. A higher-order AR may give rise to a better fit, but the computation time would increase exponentially. The use of the three-coefficient FIR was also found to be necessary. Without it, the results became significantly biased. This is as expected because the mean lifetime of the closed state is less than three samples, which is approximately the length of the rise time of the filter. Under these conditions, most closures were corrupted.

The other way to analyze the unfiltered data is simply resample it without explicitly incorporating the low-pass filter into the model. The resampling rate must be large enough so that the rise time of the residual filtering spans only about one sampling interval. By so doing, the distortion on the underlying transitions of channel currents can become negligible. We tested this approach by decimating the data with a factor of 2. The noise was still modeled by a second-order AR. The resulting estimates were also shown in Table 5, which are comparable to those obtained with explicit modeling of the finite rising time of the filter.

For this particular example the resampling approach has a comparable performance to the explicit modeling of the low-pass filtering. However, we don't know whether this is true in general. Intuitively, over-sampling data has two advantages. It can reduce the so-called random sampling error, which occurs due to the asynchronization of the sampling clock with the transitions of the channel. As a result, the current has a larger standard deviation around the transition points. Resampling data may also limit the resolution with regard to channel kinetics because only the rates that are slower than the sampling rate can be reliably estimated (Qin et al., 2000). In the above example, resampling the data with a factor of 3 led to large errors in the estimation of both kinetic rates and current amplitudes. The explicit modeling approach, however, is computationally more

TABLE 5 Parameter estimates with low-pass filtering

Parameter	True Value	Initial Value	Estimate*	Estimate†
k_{12}	38,310	4000	43847 ± 769	43833 ± 1304
k_{21}	12,770	1000	14123 ± 262	13248 ± 151
I_C	0	0.2	−0.01 ± 0.004	0.04 ± 0.011
I_O	1	0.8	0.99 ± 0.002	0.96 ± 0.004
$r_0^{(C)}$	0.64‡	0.5	0.283 ± 0.003	0.279 ± 0.006
$r_0^{(O)}$	0‡	0.5	0.279 ± 0.002	0.278 ± 0.003
$r_1^{(C)} = r_1^{(O)}$	−0.307‡	0	−0.056 ± 0.001	−0.078 ± 0.003
$r_2^{(C)} = r_2^{(O)}$	0‡	0	−0.084 ± 0.002	−0.021 ± 0.002

*A second-order FIR was explicitly used to account for the filtering.
†The data were resampled by a factor of 2, and no FIR was used.
‡These are the values before filtering is applied.

intensive because the time increases exponentially with the filter length.

Multiple channels

We consider an example of two identical channels with a five-state scheme:

$$\begin{array}{ccccc} C_1 & \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} & C_2 & \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} & C_3 \\ & & k_{42}\uparrow\downarrow k_{24} & & k_{53}\uparrow\downarrow k_{35} \\ & & O_4 & & O_5 \end{array} \qquad \text{(II)}$$

The scheme was considered by Magleby and Pallotta (1983) as a model for calcium-activated potassium channels. The data were generated by first simulating the single channel currents and then superimposing background noise. The background noise followed a correlation scheme given in Eq. 43 with a standard deviation of 0.5 pA. The excess open channel noise was white with a standard deviation of 0.2 pA. This resulted in an overall noise with a standard deviation of 0.5 pA when both channels are closed, 0.53 pA when one channel is open, and 0.58 pA when both are open. The correlation of the noise at all three conductance levels was the same because the excess noise was white. A total of 200,000 samples was simulated. The data were sampled at 50 μs without any filtering.

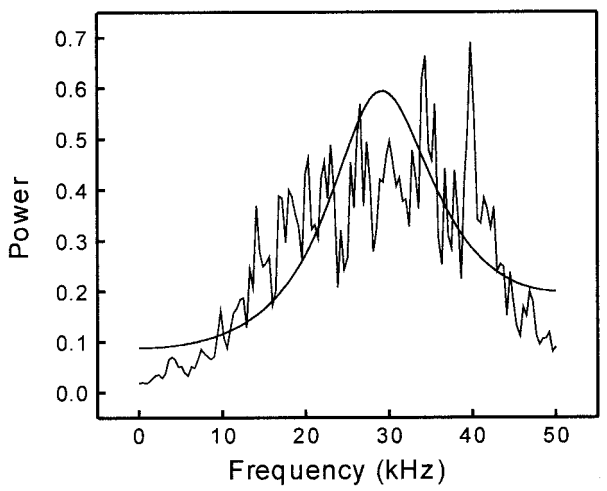

FIGURE 5 Power spectrum of noise. The original noise has a spectrum increasing quadratically with frequency, similar to the one shown in Fig. 3. The noise is then low-pass filtered at 10 kHz and corrected to 30 kHz by passing the data through an inverse filter. The spectrum of the resulting noise is shown as the ragged curve, which peaks at ~30 kHz. The smooth curve corresponds to the spectrum of the second-order AR model given by the algorithm.

A second-order AR was used to approximate the noise correlation. Note that the algorithm models the noise associated with the single channels, as if recordings from a single channel could be obtained. The noise in the multichannel data is derived from the singles. For a second-order AR, the noise is completely parametrized by the first three autocorrelations, say $r_0^{(C)}$, $r_1^{(C)}$, $r_2^{(C)}$ for the closed channel noise, and $r_0^{(O)}$, $r_1^{(O)}$, $r_2^{(O)}$ for the open channel noise. The difference of the autocorrelations between open and closed sojourns gives the autocorrelations of the excess open noise, denoted by $r_j^{(E)} \equiv r_j^{(O)} - r_j^{(C)}$, $0 \le j \le 2$. Because this excess noise is white, the constraints $r_1^{(C)} = r_1^{(O)}$ and $r_2^{(C)} = r_2^{(O)}$ hold. The multi-channel current has three conductance levels, corresponding to no channels open, one channel open, and both channels open. The autocorrelations of the noise at these three levels are given by $\{r_0^{(C)}, r_1^{(C)}, r_2^{(C)}\}$, $\{r_0^{(C)} + r_0^{(E)}, r_1^{(C)} + r_1^{(E)}), r_2^{(C)} + r_2^{(E)}\}$, and $\{r_0^{(C)} + 2r_0^{(E)}, r_1^{(C)} + 2r_1^{(E)}), r_2^{(C)} + 2r_2^{(E)}\}$, respectively.

Table 6 shows the maximum likelihood estimates of the parameters along with their starting values. All parameters, including rate constants, current amplitudes, and noise coefficients, were successfully recovered. Different starting values were also tried. Although no local solution was found, the algorithm did fail sometimes with some rate constants becoming either too small or too large. The problem is mainly with regard to the kinetics. The estimates of the current amplitudes and noise parameters were usually correct even when the algorithm failed to converge. With white background noise, the convergence appeared to be more robust, suggesting that noise correlations may cause

TABLE 6 Parameter estimates with multiple channels

Parameter	True Value	Initial Value	Estimate	S.D.
k_{12}	34	100	36	4
k_{21}	180	100	170	25
k_{23}	285	100	285	33
k_{32}	600	100	588	48
k_{24}	120	100	109	17
k_{42}	2860	1000	2796	406
k_{35}	3950	1000	3795	126
k_{53}	322	100	311	7
I_C	0	0.2	0.00	0.001
I_O	1	0.8	1.00	0.001
$r_0^{(C)}$	0.25	0.1	0.247	0.001
$r_0^{(O)}$	0.29	0.1	0.289	0.001
$r_1^{(C)} = r_1^{(O)}$	−0.12	0	−0.118	0.001
$r_2^{(C)} = r_2^{(O)}$	0	0	−0.000	0.001

ambiguities in the determination of channel kinetic parameter estimates.

The above example is relatively complicated, having both correlated noise and multiple channels. As a consequence, the program took about 18 hours on a PC, involving 79 iterations and 150 likelihood evaluations. Each evaluation of the likelihood function and its derivatives took about 7 minutes. Such a heavy computation is mainly due to the large number of metastates. With two channels, the multichannel Markov model has 15 tuple states, and then adding a second-order AR, it gives rise to 135 metastates, even with using the improved definition (Eq. 37) for metastates. If the conventional definition (Eq. 7) was used, there would be 3375 metastates, and the computation time would increase accordingly by a factor of 625.

DISCUSSION

We have described a hidden Markov modeling approach for the analysis of single channel currents. The algorithm allows for both filtering and noise correlation. The filtering is modeled using an FIR filter applied to the underlying signal to account for the finite rising time of the system. The noise correlation is accounted for using the AR model. The typical patch-clamp noise can be sufficiently approximated with a second-order AR model. The algorithm also allows for conductance-dependent noise that can be either white or correlated. A number of examples have been given to demonstrate the effectiveness and the limitations of the algorithm.

The algorithm differs from the standard HMM approach in that it uses a general optimization method to search the likelihood space. Such a choice was made in order to optimize the rate constants directly. To maximize the efficiency of optimization, we have derived the analytical derivatives of the likelihood function with respect to all model parameters. The availability of the derivatives also allows one to use more efficient optimizers, such as the variable metric method. The method has the advantage of quadratic convergence near the optimum, and provide estimates of the curvature of the likelihood surface at the maximum from which standard errors of the parameter estimates can be calculated.

The ability to optimize rate constants directly is a major advantage over the standard HMM implementation. It provides the algorithm with many desirable features such as the ability to impose physical constraints on rate constants, the ability to fit multiple data sets across different experimental conditions, the ability to allow for multiple channels, and the applicability to nonstationary channels where the starting probability depends on unknown kinetics. The standard HMM approach based on Baum's reestimation usually doesn't have such capabilities.

The algorithm exploits a similar strategy to handle noise correlation as the extended Baum's algorithm presented by Venkataramanan et al., 1998a, 1998b). There are, however, some improvements in the current algorithm. One is that although the noise is modeled with AR, it is parametrized by autocorrelations. The noise at different conductance levels is modeled separately, and the relationship is attained by imposing linear constraints on the autocorrelations. This provides the algorithm with the ability to model both white and correlated excess open channel noise. Another difference is that an improved metastate definition is used, where only the first component of the metastate is defined as a state of the channel and others as conductance classes. Such a definition gives rise to a significant reduction of the number of metastates because a channel usually has fewer conductances than states.

Compared to the dwell-time-based approach, the algorithm has several advantages. First, it avoids the necessity of idealizing the data. Second, because the noise is taken into account explicitly, it allows relatively lower signal-to-noise ratios. Third, it doesn't suffer from missed events, a problem that often limits the applicability of the dwell-time approach. All together, these features make the algorithm ideal for the problems where the channel current is too small or the kinetics too busy for accurate idealization.

Application of the algorithm in practice requires a priori knowledge of the filter coefficients, the AR order, and the topology of the kinetic model. The source of this information is largely by trial and error. The filter that is incorporated in the algorithm does not have to be exactly the same as the one actually used in filtering. Instead, it only needs to capture the shape of the rising edge of the step response of the filter, as shown in the examples. The determination of the AR order for the noise can be done retrospectively. For example, one can start with the white noise and then gradually increment the order until there is no significant change in the parameter estimates. With simulated data, we find that if the order is chosen to be higher than it should be, the additional autocorrelations become negligible. The determination of the model topology is more complicated. In theory, it could be done by calculating the likelihood of all

possible models and then choosing the most likely one. However, given the large number of candidates and the time complexity of the algorithm, this is nearly impossible in practice. A more realistic solution is to limit the search to a small number of candidates. Good candidates can be found based on the idealized data using the dwell-time maximum likelihood approach, which is computationally much faster. Sometimes, models may also be constructed from a priori knowledge. For example, structural, pharmacological, and biochemical results may suggest a number of binding sites of the channel that must be incorporated into the model even without examining the data.

The major limitation of the algorithm, or the HMM in general, is the need for intensive computation. This arises because the probability of all states at every data point needs to be evaluated. For single channels with white noise, the computation time is not a serious issue and the algorithm usually takes on the order of minutes for a reasonably large model. However, the burden increases exponentially with filtering and noise correlation. As a consequence, the algorithm, although general in theory, is limited in practice only to the problems where the data contains little filtering and the noise is gently correlated. In these cases, the algorithm takes on the order of hours.

There are several ways to extend the computational limit of the algorithm. One is to use a parallel implementation. For example, a long data set can be subdivided into multiple segments, which can then be processed simultaneously over a cluster of machines. The parallelization can also be done over states. For example, a large noise model may result in hundreds or even thousands of metastates, which may be distributed to multiple processors for the calculation of their probabilities. In general, such parallelizations can only achieve a limited improvement because the reduction in time is only linear, while the increase with filtering and noise is exponential. Another way to improve the algorithm is to seek efficient approximations of the likelihood function. Although there are a large number of metastates, many transitions may have a negligible probability. The likelihood function can therefore be approximated by discarding those unlikely transitions. This will give rise to a substantial reduction in the computational cost without significantly biasing the likelihood value, as recently shown by Fredkin and Rice (1997).

There are several problems that need to be addressed in the future. One is related to low-pass filtering. There are two kinds of filtering in the data: one is analog and applied before sampling and the other is digital and applied after sampling. The correction incorporated in the present algorithm is mainly for the second type. The first one is more difficult to handle because the transition of the underlying signal occurs randomly with respect to the sample clock. In other words, the sampling point may occur anywhere in the transition phase. This gives rise to a spread amplitude distribution instead of two delta functions, as expected in the noise-free case. It seems that such distortions cannot be modeled deterministically. For first-order filtering, it has been shown that the distribution follows a beta function (Fitzhugh, 1983). One possible solution is to extend the observation distribution $b_i(t)$ in the current algorithm to incorporate the probabilistic distribution of the samples due to filtering. With first-order filtering, it would be the convolution of a beta distribution of the filtered signal with a Gaussian distribution of the noise. Other problems that we haven't addressed include, for example, baseline drift, harmonic interference, and so on. Chung et al. (1990) have shown that the HMM can be extended to take account of these deterministic interferences explicitly. For example, baseline drift may be modeled by a polynomial function of time. The same approaches can be incorporated into the current algorithm.

This work was supported by National Institutes of Health Grant RR-11114 and W. M. Keck Foundation.

REFERENCES

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.

Blahut, R. E. 1985. Fast algorithms for digital signal processing. Addison-Wesley.

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 F. J. Sigworth. 1995. Fitting and statistical analysis of single channel records. *In* Single-Channel Recording. B. Sakmann and E. Neher, editors. Plenum Publishing Corp., New York. 483–585.

Fitzhugh, R. 1983. Statistical properties of the asymmetric random telegraph signal. *Math Biosci.* 64:75–89.

Fletcher, R. 1980. Practical Methods of Optimization. John Wiley & Sons, Chichester.

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

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

Fredkin, D. R., and J. A. Rice. 1992b. 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. 1997. Fast evaluation of the likelihood of an HMM: ion channel currents with filtering and colored noise. *Manuscript.*

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

Kay, S., and S. Marple. 1981. Spectrum analysis—a modern perspective. *Proc. IEEE.* 69:1380–1419.

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. 1990a. 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.

Magleby, K. L., and D. S. Weiss. 1990b. Identifying kinetic gating mechanisms for ion channels by using two-dimensional distributions of simulated dwell times. *Proc. R. Soc. Lond. B. Biol. Sci.* 241:220–228.

Oppenheim, A. V., and R. W. Schafer. 1975. Digital Signal Processing. Prentice-Hall, Englewood Cliffs, New Jersey.

Premkumar, L. S., and A. Auerbach. 1996. Identification of a high affinity divalent cation binding site near the entrance of the NMDA receptor channel. *Neuron.* 16:869–880.

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

Qin, F., A. Auerbach, and F. Sachs. 1996. Estimating single channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* 70:264–280.

Qin, F., A. Auerbach, and F. Sachs. 1997. Maximum likelihood estimation of aggregated Markov processes. *Proc. R. Soc. Lond. B.* 264:375–383.

Qin, F., A. Auerbach, and F. Sachs. 2000. A direct optimization approach to hidden Markov modeling for signal channel kinetics. *Biophys. J.* 79:1915–1927.

Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE.* 77:257–286.

Sigworth, F. 1995. Electronic design of the patch clamp. *In* Single-Channel Recording. B. Sakmann and E. Neher, editors. Plenum Press, New York and London. 95–126.

Sine, S. M., T. Claudio, and F. Sigworth. 1990. Activation of *Torpedo* acetylcholine receptors expressed in mouse fibroblasts: single channel current kinetics reveal distinct agonist. *J. Gen. Physiol.* 96:395–437.

Venkataramanan, L., R. Kuc, and F. J. Sigworth. 1998b. Identification of hidden Markov models for ion channel currents. II. State-dependent excess noise. *IEEE Trans. Signal Processing.* 46:1916–1929.

Venkataramanan, L., J. L. Walsh, R. Kuc, and F. J. Sigworth. 1998a. Identification of hidden Markov models for ion channel currents. I. Colored background noise. *IEEE Trans. Signal Processing.* 46: 1901–1915.

Walsh, J. L., and F. J. Sigworth. 1992. Extraction of single channel current from correlated noise via a hidden Markov model. Thesis. Yale University.