[24] Signal Processing Techniques for Channel Current Analysis Based on Hidden Markov Models

By SHIN-HO CHUNG and PETER W. GAGE

Introduction

The patch-clamp technique can be used to record single-channel activity from a small patch of cell membrane. Although the noise from a small patch is much less than that from a whole-cell membrane, signals of interest are often obscured by the noise. Even if the signal frequently emerges from the noise, low-amplitude events such as small subconductance states can remain below the noise level and there may be little evidence of their presence. It is desirable, therefore, to have a method to measure and characterize not only relatively large ionic currents but also much smaller current fluctuations that are obscured by noise.

Extracting the real signal from a limited set of imperfect measurements is a problem that commonly occurs in science, and techniques have been developed to overcome this difficulty. Following digitization, a single-channel record consists of a sequence of data points. Any movement between successive data points can be due to the signal of interest, extraneous noise, or both. The challenge is to remove the noise leaving the biological signal untouched. Some of the methods that have been used to do this are linear filtering, nonlinear filtering,[1] and transition detectors.[2–6]

Hidden Markov Model

Although both linear filtering and nonlinear filtering suppress noise and nonlinear filtering produces little distortion of rapid transitions in underlying signals, neither method utilizes all of the knowledge available about the nature of the signal and interfering noise. Using such information improves the probability of reconstructing the original signal accurately. As an analogy, suppose that some printed text is difficult to read because it has been accidentally sprayed with black ink spots. To discern the true

[1] S. H. Chung and R. A. Kennedy, *J. Neurosci. Methods* **40**, 71 (1991).
[2] J. B. Patlak, *J. Gen. Physiol.* **92**, 413 (1988).
[3] S. D. Tyerman, B. R. Terry, and G. P. Findlay, *Biophys. J.* **61**, 736 (1992).
[4] A. Queyroy and J. Verdetti, *Biochim. Biophys. Acta* **1108**, 159 (1992).
[5] D. R. Laver, *J. Gen. Physiol.* **100**, 269 (1992).
[6] J. B. Patlak, *Biophys. J.* **65**, 29 (1993).

0076-6879/98 $25.00

text, it is sensible to make use of all the information available about the text and the spots: where the offending spots came from, whether the text is in English or French, the general theme of the message, etc. In attempting to reconstruct an obscured word, it is also useful to scan the text before and after the word. For example, if a word is obscured in the following sentence "I took ... bread out of the cupboard" one might be confident that the obscured four-letter word is "some." Broadly speaking, these kinds of strategies are used in the hidden Markov model (HMM) processing technique.

Signal Model

To apply a digital signal processing technique based on HMM to records of single-channel currents contaminated by noise, we first make a plausible guess about the origin of the observation sequence and then construct a signal model. It is assumed that the channel current signal can be represented as a Markov process with the following characteristics.

Discrete Time

Time is discrete, that is, it is broken up into discrete steps. It is more convenient to deal with a discrete-time rather than a continuous-time Markov process embedded in noise. Techniques for extracting continuous-time Markov processes from noise are presented elsewhere[7] but the mathematics associated with such techniques is relatively difficult and involves use of the properties of Wiener processes and Ito stochastic calculus. Because, in practice, the experimental record we deal with is digitized so that time is in discrete steps, there is no need to add unnecessary mathematical complexity by working with continuous-time processes, although it could be done.

Finite-State

For each discrete time k, the signal, s_k is at one of a finite number of states, $q_1, q_2, \ldots, q_N$. Each q_i, where $i = 1, 2, \ldots, N$, is called a *state* of the process and such a process is called an *N-state Markov chain.* In the context of channel currents, the Markov state, s_k, represents the true conductance level (or current amplitude) uncontaminated by noise at time k. The observed value, y_k, contains the signal s_k, random noise w_k, and possibly deterministic interferences p_k, such as sinusoidal interferences from electricity mains and baseline drift. Note that the meaning of the term *state* differs from that adopted in the Colquhoun–Hawkes model of channel

[7] O. Zeitouni and A. Dembo, *IEEE Trans. Inf. Theory* **34**, 890 (1988).

dynamics,[8-11] in which *state* refers to a hypothetical, not directly observable, conformation of the channel macromolecule.

First Order

The probability of the current being at a particular level (state) at time $k + 1$ depends solely on the state at time k. The transition probabilities of passing from state level q_i at time k to state level q_j at time $k + 1$, defined as

$$a_{ij} = P(s_{k+1} = q_j | s_k = q_i)$$

form a state transition probability matrix $\mathbf{A} = \{a_{ij}\}$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, N$. Note that $\mathbf{A}$ is an $N \times N$ matrix, with its diagonal elements denoting the probabilities of remaining in the same state at time $k + 1$ as at time k.

Homogeneous

We assume that the transition probabilities are invariant of time k. The process, in other words, is taken to be stationary. To characterize a finite-state Markov chain, we define the initial state probabilities $\pi_i = \{\pi_i\}$, where $\pi_i = P(s_1 = q_i)$. We also define the noise (known as the *probabilistic function* of the Markov chain or the *symbol probability*) as $\mathbf{B} = b_i(y_k)$.

In the special case when the noise is Gaussian,

$$b_i(y_k) = (2\pi)^{-1/2}\, \sigma^{-1} \exp[-(y_k - q_i)^2/2\sigma^2] \qquad (1)$$

Specification of a signal model involves choice of the number of states N, and their amplitudes or state levels (q). Then, transition probabilities from each of N states to each of the other possible states must be chosen, giving an $N \times N$ matrix ($\mathbf{A}$). In addition, the signal model requires a prior knowledge of the variance of the noise (σ^2) and the initial probability distribution (π). We start by making initial guesses of these unknown parameters, and use the notation $\lambda = (\mathbf{q}, \mathbf{A}, \mathbf{B}, \pi)$.

Example of Signal Model

Suppose we know that a record contains four current levels but we are not sure of the exact levels nor the exact signal sequence. (Later, we discuss the situation in which a channel can assume an unknown number of conductance sublevels.) We can set up an initial model with the following assumed characteristics. First, we make a reasonable guess, and say that the baseline level is 0 pA and that there are three open states at -1, -2, and -3 pA. Second, we assume that the noise is zero-mean Gaussian, with a standard deviation of, say, 0.25 pA. Third, we provide our initial guesses of transition probabilities from one state level at time k to another state level at time $k + 1$. For a four-state Markov chain, these transition probabilities form a 4×4 transition probability matrix. In the example given here, the first entry of the first row of the matrix represents the probability that the process remained in the closed state at time $k + 1$ given that it was closed at time k, whereas the second entry of the first row represents the probability of transiting to the first open level at time $k + 1$ given that it was closed at time k. Similarly, the last entry of the last row represents the probability that the process remained in the fourth level, or the -3-pA level, at time $k + 1$ given that it was at this level at time k. Finally, we stipulate that the probability of the signal being at each one of the four levels at time $k = 1$ is 0.25.

These assumptions can be represented as:

1. State levels ($\mathbf{q}$)		q_1 = 0 pA			
		q_2 = −1 pA			
		q_3 = −2 pA			
		q_4 = −3 pA			
2. Noise characteristics	Gaussian with	$\sigma = 0.25$ pA			
3. Transition matrix		a_{11}	a_{12}	a_{13}	a_{14}
		a_{21}	a_{22}	a_{23}	a_{24}
		a_{31}	a_{32}	a_{33}	a_{34}
		a_{41}	a_{42}	a_{43}	a_{44}
4. Initial probabilities		$\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.25$			

We can also stipulate, if needed, that there is AC hum (50 or 60 Hz and its odd harmonics) embedded in the data but, for simplicity, this is not included in this example. We compactly write all these initial guesses as $\lambda^1 = (\mathbf{q}, \mathbf{A}, \mathbf{B}, \pi)$, our first signal model.

Expectation–Maximization Algorithm

The signal model is compared with the data. Essentially, the probabilities of all possible "pathways" between adjacent data points are calculated, both forward and backward, and the true current levels derived from the highest probabilities. Because the initial parameters we have supplied (e.g., the transition probability matrix and conductance levels) are only guesses,

[8] D. Colquhoun and A. G. Hawkes, *Proc. R. Soc. Lond. B* **199**, 231 (1977).
[9] D. Colquhoun and A. G. Hawkes, *Proc. R. Soc. Lond. B* **211**, 205 (1981).
[10] D. Colquhoun and F. J. Sigworth, *in* "Single Channel Recordings" (B. Sakmann and E. Neher, eds.), pp. 191–263. Plenum Press, New York, 1983.
[11] D. Colquhoun and A. G. Hawkes, *Proc. R. Soc. Lond. B* **240**, 453 (1990).

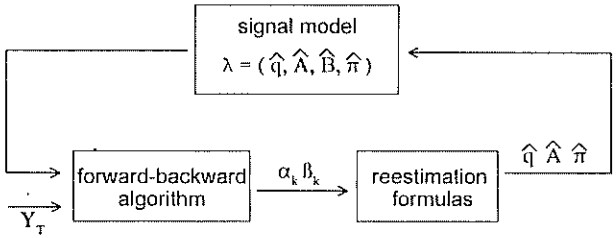


FIG. 1. A block diagram of the processing method. On the basis of the initial signal model λ, the observation sequence Y_T is processed, and the forward and backward variables, α_k and β_k, are computed for each discrete time k and each Markov state q_i. By using these variables, the parameters of the signal model are revised according to the reestimation formulas. The entire process is repeated many times.

there is going to be a mismatch between the model and data. The model is revised so that it will be more consistent with the data. Using the revised model, the observation sequence is compared again with the new model, and the model is again revised. This iterative process continues, as shown in Fig. 1.

The E Step

In this iterative process, the HMM processing technique utilizes two mathematical principles. The first is the *forward–backward procedure*, known also as the *E step* of the expectation–maximization algorithm.[12] For each data point k, the algorithm evaluates, using Bayes' rule, the forward and backward variables, α and β, defined as:

$$\alpha_k(i) = P(Y_k, s_k = q_i|\lambda), \qquad \beta_k(i) = P(\overline{Y}_k|s_k = q_i, \lambda) \tag{2}$$

where Y_k refers to the past observation sequence from $k = 1$ to k, and $\overline{Y}_k$ refers to the future observation sequence from $k + 1$ to T. In words, the forward variable $\alpha_k(i)$ is the joint probability of the past and present observation with the present signal in state q_i, given the model λ, and $\beta_k(i)$ is the probability of the future observation given that the present state is q_i and given the model λ.

The forward variable is calculated in a forward recursion and the backward variable in a backward recursion. Recursive formulas for Eq. (2) are

[12] A. P. Dempster, N. M. Laird, and D. B. Rubin, *J. R. Statist. Soc. B.* **39,** 1 (1977).

readily calculated[13] using Bayes' rule, as:

$$\alpha_k(i) = \sum_{i=1}^{N} \alpha_{k-1}(i)\, a_{ij} b_j(y_k), \qquad \alpha_1(j) = \pi_j b_j(y_1)$$

$$\beta_k(i) = \sum_{j=1}^{N} \alpha_{ij} b_j(y_{k+1}) \beta_{k+1}(j), \qquad \beta_T(i) = 1$$

The M-Step

Then, using the forward and backward variables, the model is reestimated, using the Baum–Welsh reestimation formulas, known also as the *M step* of the expectation–maximization (EM) algorithm. Estimation formulas stipulate how the model parameters should be revised, given the forward and backward variables. Loosely stated, the E step of the EM algorithm expresses the expectation of the likelihood function if the parameters of the old model are replaced by a new model. In our application, this is a lengthy equation involving many variables. The M step involves finding the parameters of the new model λ^2 to maximize this expectation of the likelihood function of the "fully categorized" data.[14] Just as we do when finding the maximum of a cubic equation, we differentiate this equation with respect to each of the particular variables (with certain constraints where applicable) and obtain a new set of equations by letting the derivative equal zero. The full derivations of these reestimation formulas are given elsewhere.[15] These new sets of equations, called the *reestimation formulas*, are used to replace the parameters of the first signal model with new parameters to create the second signal model. The same segment of data is now processed with the second signal model, and then the parameters of this model are replaced, again according to the reestimation formulas, in the third signal model. The process is iterated again and again, 15 to 1000 times.

Computational Procedures

For illustration, we use the example of the signal model given earlier and follow how the first few computational steps are carried out to evaluate the forward variables. Because the model stipulates that there are four signal states ($q = 0, -1, -2, -3$ pA), there are four forward variables

[13] S. H. Chung, J. B. Moore, L. G. Xia, L. S. Premkumar, and P. W. Gage, *Phil. Trans. R. Soc. Lond. B* **329,** 265 (1990).

[14] D. M. Titterington, A. F. M. Smith, and V. E. Makov, "Statistical Analysis of Finite Mixture Distributions." John Wiley & Sons, New York, 1985.

[15] S. H. Chung, V. Krishnamurthy, and J. B. Moore, *Phil. Trans. R. Soc. Lond. B* **334,** 357 (1991).

$[\alpha_1(1), \alpha_1(2), \alpha_1(3)$, and $\alpha_1(4)]$ to be calculated for time $k = 1$. Assume that the first data point happened to be -1.3 pA. The probability that the real signal at k was 0 pA but the additive Gaussian noise would have displaced this point to -1.3 pA is computed with Eq. (1), with the standard deviation of this noise taken to be 0.25 pA, as we specified. This probability is multiplied with the initial probability of the signal being at the 0-pA level, namely, 0.25, to obtain the first forward variable belonging to the first level, $\alpha_1(1)$. The forward variable at time $k = 1$ belonging to the second level, $\alpha_1(2)$, is computed by evaluating the probability that the noise would have displaced the signal from -1 to -1.3 pA and then multiplying this quantity with the initial probability of the signal being at the -1-pA level, 0.25. The remaining two forward variables, $\alpha_1(3)$ and $\alpha_1(4)$, are similarly tabulated.

When it comes to the second data point, again four forward variables, $\alpha_2(1), \alpha_2(2), \alpha_2(3)$, and $\alpha_2(4)$, need to be computed. Suppose that the second measured current value was -0.5 pA. If the signal at $k = 2$ was assumed to be at the 0-pA level, it would have made one of the four possible transitions from time $k = 1$; from the 0-pA level to the 0-pA level, from the -1-pA level to the 0-pA level, from the -2-pA level to the 0-pA level, and from the -3-pA level to the 0-pA level. In our signal model, the probabilities of making such transitions are stipulated in the first column of the 4×4 matrix. The probability that additive Gaussian noise would displace the current level from 0 to -0.5 pA is calculated using Eq. (1). Numerically, this probability is 0.216. In our example, $\alpha_2(1)$ is the sum of the four terms: $[\alpha_1(1) \times a_{11} \times 0.216] + [\alpha_1(2) \times a_{21} \times 0.216] + [\alpha_1(3) \times a_{31} \times 0.216] + [\alpha_1(4) \times a_{41} \times 0.216]$, as the recursive formula given in Eq. (2) states. The backward variable $\beta_2(1)$ is computed similarly, using the recursion formula given in Eq. (2).

Once the numerical values of $\alpha_2(1)$ and $\beta_2(1)$ are tabulated, we multiply them and call this quantity $\gamma_2(1)$, after suitable normalization. This is the probability of the signal being in the 0-pA level at time $k = 2$, given the observation sequence y_T and the model λ. In symbols,

$$\gamma_k(i) = P(s_k = q_i | y_T, \lambda) \tag{3}$$

and this quantity can be computed from the forward and backward variables using the formula:

$$\gamma_k(i) = [\alpha_k(i)\, \beta_k(i)] \Big/ \sum_{i=1}^{N} \alpha_k(i)\, \beta_k(i)$$

The initial parameters are reestimated using the forward and backward variables. To revise the first current level, which we specified as 0 pA in our example, we multiply the first data point y_1 by $\gamma_1(1)$, the second data point y_2 by $\gamma_2(1)$, the kth data point y_k with $\gamma_k(1)$, and the last data point y_T with $\gamma_T(1)$, and then add all these values. This total sum, normalized by dividing by the sum of all γ's belonging to the first current level, namely, $\gamma_1(1) + \gamma_2(1) + \ldots + \gamma_k(1)$, gives the reestimated current level of the first Markov state. More generally,

$$^*q_i = \sum_{i=1}^{T} \gamma_k(i)\, y_k \Big/ \sum_{k=1}^{T} \gamma_k(i)$$

where $^*q_i$ is the estimate of the ith state level. The elements in the transition probability matrix are similarly revised using the reestimation formula. The old parameters are replaced with these revised parameters, and the new signal model, now λ^2, is used and all of the computational steps are repeated. This procedure is repeated many times.

Reestimation Theorem

After each iteration, we can compute from the forward variable a numerical value that we call the *likelihood function*—that is, how likely are the model parameters given the data sequence. The closer the model parameters are to the true parameters, the higher the likelihood function. If the likelihood functions were to increase and decrease erratically with successive iterations, this procedure would have been a waste of time. The rationale behind the iterative procedure rests on the elegant reestimation theorem formulated by Baum and colleagues,[16–18] which states:

$$P(Y_T|\lambda^{n+1}) \geq P(Y_T|\lambda^n)$$

In words, the probability of the observation sequence Y_T, given the reestimated signal model, is greater than or equal to the probability of Y_T, given the previous signal model. Thus, the signal sequence estimated using the revised model is more consistent with the data than that estimated using the previous signal model. When the iterative procedure converges, then $P(Y_T|\lambda^{n+1}) = P(Y_T|\lambda^n)$, and λ^n is termed the *maximum likelihood estimate* of the HMM. This important theorem, the proof of which is based on Jensen's inequality,[16] is the core of the HMM processing scheme.

There is a choice of numerical methods for calculating the maximum likelihood estimates. One approach is the Newton–Ralphson algorithm, which, when it converges, does so quadratically and thus rapidly. The EM algorithm, on the other hand, converges linearly, and so convergence can

[16] L. E. Baum and T. Petrie, *Ann. Math. Statist.* **37,** 1554 (1966).
[17] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, *Ann. Math. Statist.* **41,** 164 (1970).
[18] L. E. Baum, *Inequalities* **3,** 1 (1972).

be very slow. However, successive iterations with the Newton–Ralphson algorithm do not necessarily improve the likelihood function. In contrast, the EM algorithm is simple to implement and has the appealing property that the likelihood function is always improved after each iteration.

In the graph in Fig. 2, the likelihood function obtained from one of the data segments is plotted against successive iterations. This shows that the probability that the observed data sequence could have been generated by each successively reestimated model increases steadily and then converges by the time the process is iterated about 20 times.

Estimating the Number of States

Perhaps the most subjective part of the HMM processing method, like any parameter estimation scheme, is finding the state dimension—or the number of conductance states in our example—in hidden Markov chain processes. The error in fitting a model to a given set of data decreases with the number of free parameters in the model. Thus, it makes sense, in selecting a model from a set of models with different numbers of parameters, to penalize models having too many parameters. The question of how to

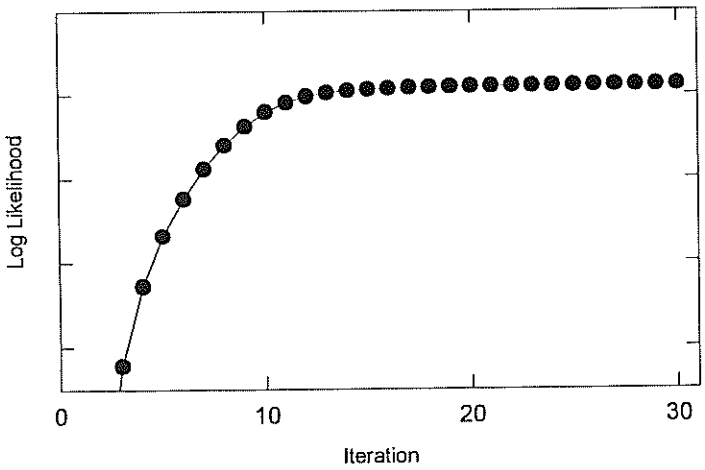


FIG. 2. Increase in the log likelihood function with successive iterations. The likelihood that the observation sequence could have been generated by each successively reestimated model increases rapidly at first and then slowly but steadily with each additional iteration. A segment of the record used to obtain the likelihood function is shown in the top panel of Fig. 3A, and the parameters of the first signal model are given in the text. In the first signal model, the diagonal and all off-diagonal elements of the transition matrix were assumed to be 0.91 and 0.03, respectively.

penalize HMM for having an excessive number of free parameters is an area of current research (see, for example, Ref. 19) and one proposed criterion for model-order selection is the compensated likelihood approach.[20]

In practice, however, it is relatively easy to identify the number of states present in the underlying Markov chains. One of several ways of doing this is by constructing the most likely amplitude histogram, as shown in Fig. 3 (also see Fig. 4B). Here we assume that the signal can be represented as a Markov chain with a large number of equally spaced states, say, 100 states, and assume that the actual state levels are fixed. We construct an amplitude histogram from the estimated signal sequence obtained under this assumption, as shown here. These three histograms were derived from the three sets of fictitious channel currents exhibited in the first panel. The maximum likelihood histograms, indicated as bars, clearly show four prominent peaks, the baseline at the right-hand side, and three open current levels. These peaks are evenly spaced at 0.38 pA. In contrast, no meaningful information can be gleaned from the amplitude histogram of the original record, indicated here as a superimposed continuous curve.

Alternatively, we can appeal to the principle of parsimony in deciding the number of conductance states. We measure the goodness of fit by evaluating the likelihood of the model, and weigh this against what is to be gained by increasing the number of parameters, which generally increases the likelihood. Thus, we process the same data segment under the assumption that the underlying signal has a different number of conductance states. If a plot of log likelihood versus model order (number of states) shows a "knee" for a certain model order (Fig. 4A, asterisk), we would prefer this model to one of higher order. This approach has been used to determine the number of conductance substates in channel currents activated by γ-aminobutyric acid (GABA).[21]

Extension of Signal Model

Under the simple assumption that the signal sequence can be represented by a discrete-time, first-order, finite-state, homogeneous Markov chain, the HMM processing technique provides the following information. First, it identifies the number of conductance states in the true signal that is hidden in the noise. Second, it identifies the state levels of the signal.

[19] D. S. Poskitt and S. H. Chung, *Adv. Appl. Probab.* **28,** 405 (1996).
[20] L. Finesso, "Consistent estimation of the order of Markov and hidden Markov chains." PhD Dissertation, University of Maryland (1990).
[21] P. W. Gage and S. H. Chung, *Proc. R. Soc. Lond. B* **255,** 167 (1994).

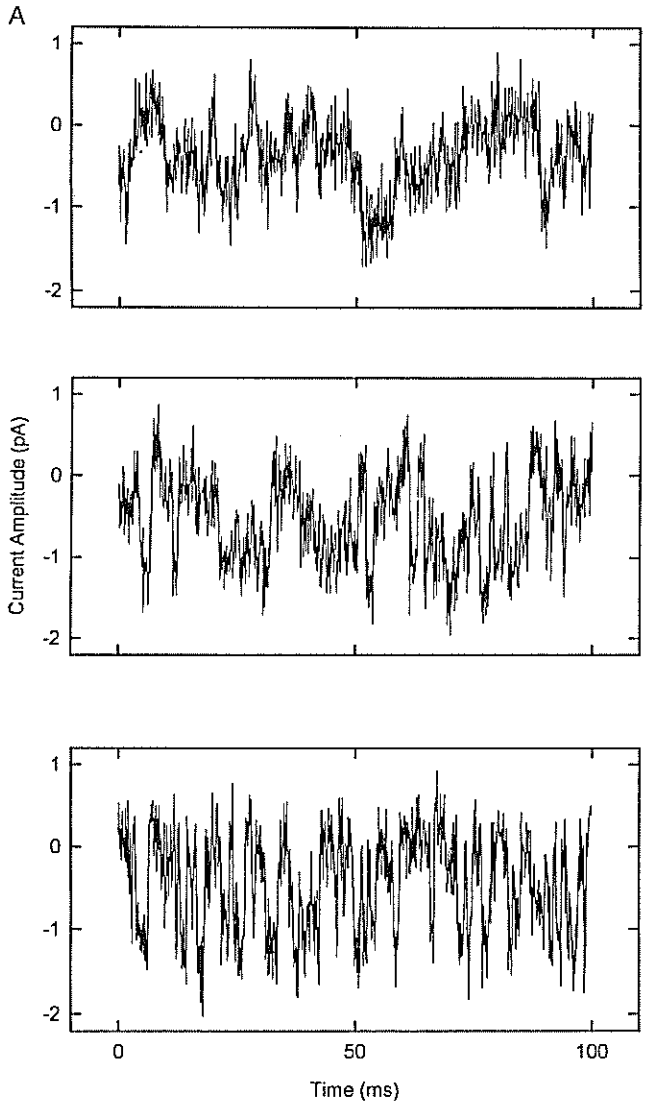
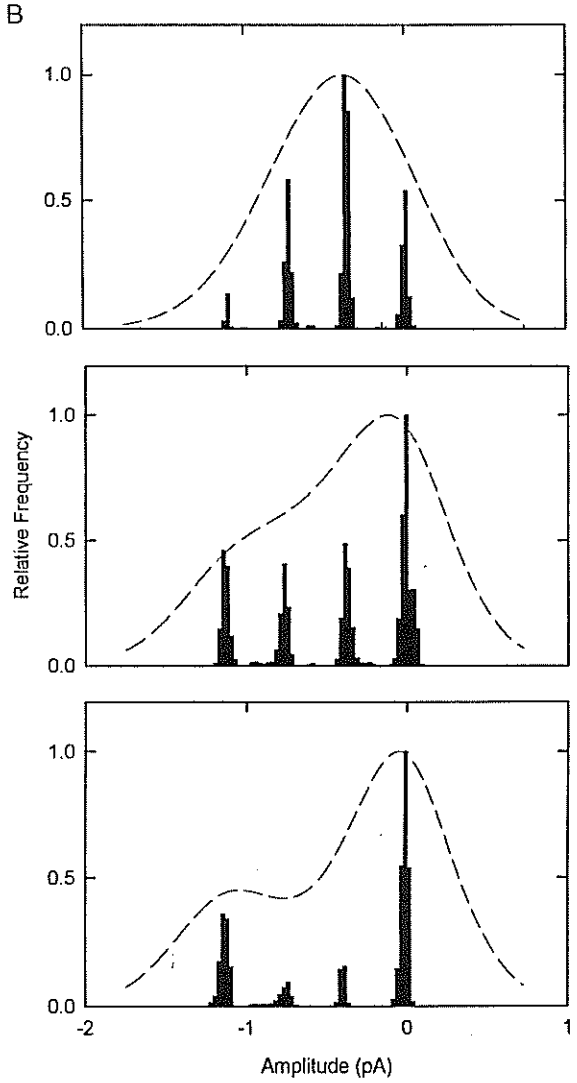
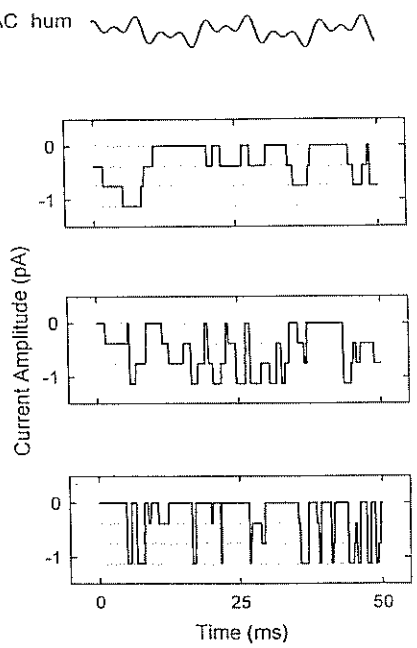


FIG. 3. Segments of synthesized channel currents and estimation of their current levels. (A) The traces show three first-order, two-state, Markov chains with $a_{11} = 0.99$ and $a_{22} = 0.98$ opening and closing independently of each other (top) or in partial synchrony (middle and bottom). The coupling coefficients of the signals contained in the three segments from top down were 0, 0.05 and 0.15, respectively. The open current level of each binary chain was 0.375 pA. To the signal sequences, white Gaussian noise (standard deviation 0.25 pA) and AC hum (50, 150, and 250 Hz) were added: the amplitude of each component was 0.25 pA. Once buried in the random and deterministic interferences, the original signal sequences and their current levels became obscured. (B) The maximum likelihood estimates of the signal sequences embedded in the record and AC hum contaminating the signal were obtained using the iteration scheme shown in Fig. 1. The amplitude histograms obtained from the estimated signal sequences, corresponding to the three segments of the current traces displayed in part (A), are shown as dark bars. The peaks of the histograms are evenly spaced at about 0.38 pA. In contrast, no meaningful information can be gleaned from the all-points amplitude histograms of the original records, indicated here as continuous lines superimposed on the bar graphs. (C) Using the fact that there are four Markov states separated by 0.38





pA, the most likely signal was extracted from the currents in Fig. 3A together with the underlying hum (top trace). The three signal sequences differ in the degree of coupling between three channels. In the first example, the three binary channels were uncoupled; they opened and closed independently of each other. In the second example, they were weakly coupled in that they opened and closed not totally independent of each other but in partial synchrony. In the third example, they were strongly coupled. When one channel opened (or closed), there was a tendency for the other two channels to open (or close) in synchrony.

Third, it gives the maximum likely estimate of the signal sequence—what the signal must have been like before it became obscured by noise. Finally, it gives the stochastic matrix of the signal from which all the necessary kinetic parameters can be obtained. These include the mean open and closed times and interval histograms. If there is more than one channel (more than one Markov chain), the structure of this stochastic matrix enables us to deduce whether these chains are totally independent, or whether they show some form of dependency.[22,23]

What if the real signal deviates from the underlying assumptions? Typical biological signals may not be best approximated by a first-order Markov process, especially if the number of events sampled is small. Moreover, if

[22] S. H. Chung and R. A. Kennedy, *Math. Biosci.* **133,** 111 (1996).
[23] R. A. Kennedy and S. H. Chung, *Intl. J. Adap. Control Signal Proc.* **10,** 623 (1996).

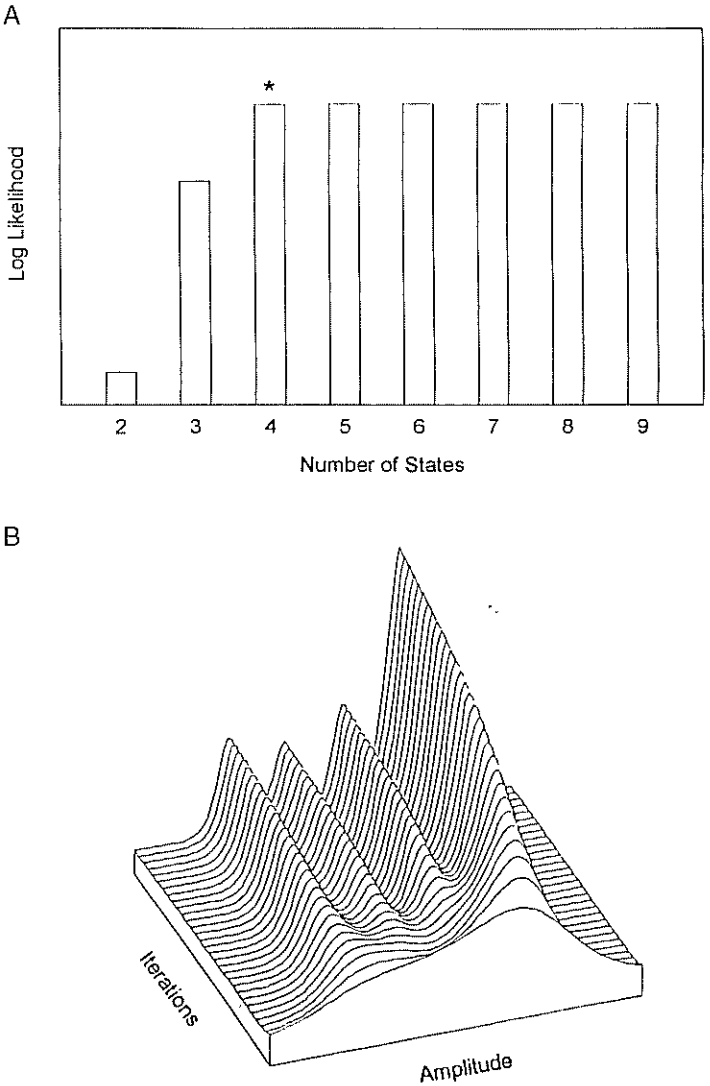


FIG. 4. Estimation of the state dimensions. (A) The likelihood values (bars) of the observation sequence, illustrated at the top of Fig. 3A, were computed under the assumption that the signal was a two-state, three-state, and up to a nine-state Markov chain. The likelihood value first increased as the number of states was increased from two to four, and then remained unchanged with increasing model order. Appealing to the principle of parsimony in determining the model order, we conclude that a four-state model (marked with an asterisk) fits the data. (B) The maximum likelihood amplitude histogram was constructed to determine the model order. We assumed that the signal can be represented as a Markov chain with 100 states. With successive iterations of the data under this assumption, the location of the peaks become more and more distinct. The large peak on the right-hand side is the baseline and three additional peaks on the left correspond to the open current levels.

the underlying conformational changes of a channel macromolecule can be represented by the kinetic scheme: $C_1 \rightleftharpoons C_2 \rightleftharpoons O$, where C_1 and C_2 are two distinct closed conformational states and O is the open conformational state and the transitions between states are governed by a first-order Markov process, the observed two-state signal sequence would not be a Markov process. (Technically, such a process can be construed as a switched Markov chain.) It has been demonstrated previously that departures from the underlying assumptions do not seriously degrade the performance of the processing method.[13,15] Even when the signal embedded in noise deviates drastically from the first-order Markov assumption and the additive noise is correlated, the technique extracts the signal and estimates its statistics with an acceptable degree of accuracy.

In the same framework of HMM techniques and the EM algorithm, the signal model can be further modified and extended to make the signal processing schemes far more versatile than those described here. For example, a Markov process, after entering one of its states, can be allowed to decay back in time exponentially (or otherwise) to the original state. Such a signal process can be formulated as an augmented homogeneous HMM problem, and a scheme for estimating this stochastic process can be devised. Also, it is not difficult to extend the signal model to a higher order Markov or a semi-Markov process and devise a scheme for estimating this process. A mathematical description of this extension is described elsewhere.[24] In the algorithm that is currently available, it is assumed that the amplitude and fundamental frequency of AC hum that is added to the recordings remain unchanged within the short data segment being analyzed. In reality, the amplitude of the harmonic wave may wax and wane slowly. It is again easy to incorporate these variables into the processing scheme at the expense of added computational cost.

Since the HMM processing techniques were introduced for analyzing single-channel currents,[13,15,25] several groups have refined and adapted the algorithms so that the method can be applied to practical patch-clamp recordings when the assumptions underlying the idealized signal model no longer hold. For example, one of the assumptions underlying the processing method is that noise is white and uncorrelated with the signal. If this assumption is violated, the algorithm performs poorly, as pointed out originally.[13] In reality, noise superimposed on single-channel currents is not white but colored. Its spectral power, instead of being flat, rises steeply at high frequencies. Moreover, the standard deviation of noise increases when

[24] V. Krishnamurthy, J. B. Moore, and S. H. Chung, *Signal Proc.* **24,** 177 (1991).
[25] D. R. Fredkin and J. A. Rice, *Proc. R. Soc. Lond. B* **249,** 125 (1992).

the channel is in the open state.[26] This issue was addressed by use of an autoregressive noise model to represent temporal correlation in the background noise contained in patch-clamp recordings and then the EM algorithms for such correlated and additive state-dependent excess noise were reformulated.[27,28] It was demonstrated with simulations that the performance of the algorithm was markedly improved when the background noise was modeled realistically.

On many occasions, recordings obtained from a membrane patch contain more than one channel and currents from several channels are superimposed. It is possible to extract, again using a modified HMM algorithm, the transition probability matrices and current amplitudes for each individual channel from such multichannel patches.[15,29] Finally, a method for handling the situation in which a Markovian signal embedded in noise is passed through a low-pass filter has also been addressed.[30]

Model Evaluation

One additional application of the HMM processing method may be very useful. The technique can be used to evaluate different types of models. Many schemes have been proposed to explain ion channel kinetics, including aggregated Markov models,[8] fractal models,[31–34] semi-Markov models encompassing fractal models,[35,36] hidden Markov models,[25,37] and coupled Markov models.[22,23] With so many feasible models, it would be useful to have an objective analytical tool to rank them and to reject any that are implausible. The HMM technique can be used for this. Not only will it estimate (in an optimal sense) the model parameters based on the observation sequence, but it will also discriminate between a set of alternative

[26] J. D. Becker, J. Hohnerkamp, J. Hirsch, U. Frobe, E. Schlatter, and R. Greger, *Eur. J. Physiol.* **426,** 328 (1997).
[27] L. Venkataramanan, J. L. Walsh, R. Kuc, and F. J. Sigworth, *IEEE Sig. Processing* (1998). In press.
[28] L. Venkataramanan, R. Kuc, and F. J. Sigworth, *IEEE Sig. Processing* (1998). In press.
[29] S. Klein, J. Timmer, and J. Honerkamp, *Biometrics* **53,** 43 (1997).
[30] V. Krishnamurthy and L. B. White, *Proc. IFAC Intl. Conf. Adapt. Syst. Cont. Signal Proc.* 633 (1992).
[31] L. S. Liebovitch, J. Fischbarg, and J. P. Koniarek, *Math. Biosci.* **78,** 203 (1986).
[32] S. J. Korn and R. Horn, *Biophys. J.* **54,** 871 (1988).
[33] L. S. Liebovitch, J. Fischbarg, J. P. Koniarek, I. Todorova, and M. Wang, *Biochim. Biophys. Acta* **896,** 173 (1987).
[34] L. S. Liebovitch and J. M. Sullivan, *Biophys. J.* **52,** 979 (1987).
[35] F. Ball and M. S. P. Sansom, *Proc. R. Soc. Lond. B* **236,** 385 (1989).
[36] F. Ball, G. F. Yeo, R. K. Milne, R. O. Edeson, R. W. Madsen, and M. S. P. Sansom, *Biophys. J.* **64,** 357 (1993).
[37] D. R. Fredkin and J. A. Rice, *Biometrics* **48,** 427 (1992).

models by assessing their plausibility.[38] Traditionally, identification of membrane channel models has focused on techniques involving the numerical fitting of the interval histograms or the power spectrum of the noisy channel data. These approaches are often not robust when the data are noisy and are limited in the amount of information they can yield about the process parameters. Such approaches only work when the relevant time constants (i.e., mean durations) differ sufficiently. Some other approaches that have been tried include maximum likelihood estimation,[35] which suffers an exponential complexity problem, and correlation functions,[33] which have only been applied to identify binary Markov chains.

Hidden Markov modeling is a computationally efficient maximum likelihood estimator that is robust in the presence of white noise and applicable to processes with many states. It may well prove useful for evaluating the likelihood of models proposed for ion channels. As an illustration, let us suppose that we observed in a coin-tossing experiment a sequence (HHHTHH). We construct two alternative signal models, the first being that the coin is unbiased and the second being that it is bent in such a way that a head would turn up 90% of time, $P(\mathrm{H}) = 0.9$. By comparing the likelihood values, we can state that this observation sequence is more consistent with the second signal model than with the first. As we try out various pairs of signal models, we discover that there is a unique $P(\mathrm{H})$ that gives the largest likelihood value, L_{T}, and we call this the *maximum likelihood estimate* of $P(\mathrm{H})$, 0.067 in this example. The principle of selecting the most likely signal model as illustrated with the simple coin-tossing example can readily be generalized to channel currents. A more complete discussion of the model selection problem is available.[38] In a similar vein, a scheme for testing whether a recorded sequence of channel currents can be described adequately as Markov chain has been proposed.[39] This statistical test relies on the fact that the asymptotic distributions of the log-likelihood will be distributed normally if the signal embedded in noise is a homogeneous, first-order Markov chain. Using simulated signal sequences modified from a homogeneous, first-order Markov chain systematically to render them non-Markovian, they demonstrated that their test procedure can detect when the signal deviates appreciably from the assumed model.

Computational Cost

Modern fast desktop computers have made the HMM technique completely accessible to all researchers. The number of computational steps

[38] G. W. Pulford, J. C. Gallant, R. A. Kennedy, and S. H. Chung, *Biometrics* **37,** 39 (1995).
[39] J. Timmer and S. Klein, *Phys. Rev. E* **55,** 3306 (1977).

involved in one forward–backward process is N^2T, where N is the number of Markov states and T is the number of data points. Typically, we analyze a 100,000-point record using about 10 allowed states, and the process is iterated about 100 times, resulting in computational steps of approximately 10^9. A modern, inexpensive, desktop computer can perform about 200 million computations per second, and mainframe supercomputers that can carry out 5×10^9 floating-point operations per second (5 gigaflops) are now readily available to researchers. Thus, the real time involved in processing such a record, once the codes are optimized, can be of the order of minutes.