Estimation of Kinetic Rate Constants from Multi-Channel Recordings by a Direct Fit of the Time Series

Arne Albertsen and Ulf-Peter Hansen
Institut für Angewandte Physik, Christian-Albrechts-Universität, D-24098 Kiel, Germany

ABSTRACT The maximum-likelihood technique for the direct estimation of rate constants from the measured patch clamp current is extended to the analysis of multi-channel recordings, including channels with subconductance levels. The algorithm utilizes a simplified approach for the calculation of the matrix exponentials of the probability matrix from the rate constants of the Markov model of the involved channel(s) by making use of the Kronecker sum and product. The extension to multi-channel analysis is tested by the application to simulated data. For these tests, three different channel models were selected: a two-state model, a three-state model with two open states of different conductance, and a three-state model with two closed states. For the simulations, time series of these models were calculated from the related first-order, finite-state, continuous-time Markov processes. Blue background noise was added, and the signals were filtered by a digital filter similar to the anti-aliasing low-pass. The tests showed that the fit algorithm revealed good estimates of the original rate constants from time series of simulated records with up to four independent and identical channels even in the case of signal-to-noise ratios being as low as 2. The number of channels in a record can be determined from the dependence of the likelihood on channel number. For large enough data sets, it takes on a maximum when the assumed channel number is equal to the "true" channel number.

INTRODUCTION

Biological transport molecules can switch between active (open) and inactive (closed) states. A powerful tool for the study of the kinetics of the transitions between different states of a transport molecule is patch clamp analysis. Usually, the analysis starts with the reconstruction of the noise-free record of channel switching by different kinds of detectors as compared by Schultze and Draber (1993). Dwelltime histograms for the open and closed states are generated from these records, and the time constants and amplitude factors are evaluated by fitting the histograms with exponential functions. On the basis of an adequate model, the rate constants of the transitions between states have to be calculated from the time constants. This is an indirect approach, because the time constants are intermediate quantities from which the rate constants have to be evaluated.

There is an alternative approach that is not based on dwelltime histograms and calculation of time constants. Fredkin and Rice (1992a) used the concepts introduced by Baum et al. (1970) for a direct fit of the stochastic time series of the measured patch current. The approach of Fredkin and Rice (1992a), however, was restricted to records of a single ion channel. Other authors (Fredkin and Rice, 1992b; Chung et al., 1990; Jackson, 1985; Horn and Lange, 1983) have addressed the problem of multi-channel analysis, but most of them do not present methods for the extraction of rate constants from the time series. Very commonly, these authors are interested in certain overall statistics of the observed current record like open probability, mean current levels, noise, and dwelltime distributions. Thus, these methods aim at the reconstruction of the underlying, noise-free current trace as it would arise from a Markov model (Fredkin and Rice, 1992b; Chung et al., 1990) or at the construction of dwelltime functions for a multi-channel system, where the individual channel possesses a small open probability (Jackson, 1985). Horn and Lange (1983) introduced a maximum likelihood approach to estimate rate constants from idealized, i.e., previously restored, time series. Their approach relies on the knowledge of the number of channels and is restricted to channels without subconductance states.

Here an extension of the approach outlined by Fredkin and Rice (1992a) is given. Our method extracts rate constants of the single channel directly from patch records containing several independent and identical channels given initial estimates for the rates and the channel number. Plots of the Log likelihood versus channel number provide an estimate for the actual channel number for which the likelihood is maximized. The application of the method is demonstrated for simulated time series of noisy and filtered multi-channel records, including channels with subconductance states.

THEORY AND SIMULATION

Incorporation of the method of Fredkin and Rice for the estimation of the likelihood of a measured time series for a given set of rate constants

The multi-channel analysis presented here is an extension of the approach of Fredkin and Rice (1992a), who employed techniques from the analysis of speech records (Baum et al., 1970). Briefly, the procedure of Fredkin and Rice (1992a) starts from a guess of a model of the configuration of open and closed states and of the rate constants of the transitions between them. Then, the so-called forward probability (Huang et al., 1990) is calculated. Starting from a guess of

Received for publication 26 October 1993 and in final form 1 April 1994.
Address reprint requests to Dr. Arne Albertsen, Inst. fur Angewandte Physik, Christian-Albrechts Univ., Olshausenstr. 40, 24098 Kiel, Germany. E-mail: arne@ang-physik.uni-kiel.d400.de.

0006-3495/94/10/1393/11 $2.00

the initial probabilities of being in a certain state s at time 0 ($a_0(s)$) the probability of being in state i at time $k\Delta t$ is computed recursively.

$$a_k(i) = \sum_{j=1}^{L} a_{k-1}(j)\, p_{ji} f_i(y_k) \tag{1}$$

The $a_{k-1}(j)$ give the probability of being in state j at the time $(k-1)\Delta t$. Here p_{ji} is the probability of a transition from state j to state i in the time interval between $(k-1)\Delta t$ and $k\Delta t$.

The comparison between measured and predicted data is introduced via $f_i(y_k)$. Here $f_i(y_k)$ is the probability that the measured current y_k can be observed when the channel is in state i. The product $a_{k-1}(j)p_{ji}f_i(y_k)$ obtains maximum values if predicted state and measured current match.

To obtain an estimate of the probability $f_i(y_k)$, the amplitude histogram of the measured current record has to be split into a sum of individual histograms (Gaussian if the influence of low-pass filtering on fast switching is ignored (Yellen, 1984; Draber and Schultze, 1994)), each related to the individual conductivity of the state i. Then, $f_i(y_k)$ is the ordinate at current y_k in the ith histogram.

In the case of Gaussian noise, the distribution $f_i(y)$ of y would only depend on the mean current μ_i and the SD σ_i.

The following normalizing conditions hold for the $f_i(y)$ with m_i denoting the multiplicity of states with the same conductance:

$$m_i \int_0^{\infty} f_i(y)\, dy = 1 \qquad i \in \{1, 2, \cdots, L\}. \tag{2}$$

In the literature on Hidden Markov Models, the term $f_i(y)$ is often referred to as symbol probability (Rabiner, 1989; Huang et al., 1990).

Because the output values are discrete, the integral in Eq. 2 becomes a finite sum with y_{max} depending on the resolution of the data acquisition system (in our case, $y_{max} = 4095$ corresponding to a 12 bit A/D converter):

$$m_i \sum_{y=0}^{y_{max}} f_i(y) = 1 \qquad i \in \{1, 2, \cdots, L\}. \tag{3}$$

The analysis of channel kinetics aims at the k_{ji}, the rate constants that describe the probability of a transition in an infinitesimally short time, in contrast to the p_{ji} of Eq. 1, which are related to a finite time interval Δt (sampling interval).

The matrix $\mathbf{P}$ (Baum et al., 1970) with the elements p_{ji} can be calculated from the matrix $\mathbf{K}$ with the elements k_{ji}

$$\mathbf{P}(\mathbf{K}, \Delta t) = \exp(\mathbf{K}\Delta t) \tag{4}$$

according to the solution of the Kolmogorov-Feller equations in the case of a finite state, stationary Markov model (Bharucha-Reid, 1960). The matrix exponential is defined via a Taylor series expansion (Golub and Van Loan, 1983). The symbol Δt is omitted from now on.

The $\mathbf{K}$ matrix is used because it enables the introduction of physically relevant constraints. The constraints for $\mathbf{K}$, which have to be incorporated in any case, are as follows:

$$\begin{aligned} k_{ij} &\geq 0 \qquad && i \neq j \in \{1, \cdots, L\} \\ k_{ii} &= -\sum_{j=1 \neq i}^{L} k_{ij} \qquad && i \in \{1, \cdots, L\}. \end{aligned} \tag{5}$$

In addition, certain model topologies may imply special conditions, e.g.: 1) usually some states are not connected; this is achieved by setting individual elements of the rate matrix $\mathbf{K}$ to zero; 2) microreversibility (detailed balance) may be required to hold in cyclic systems without energy input.

Equation 1 is the key for the restriction of computer time to a linear rise with the number N of samples, because the information about the precursors of the state at time $(k-1)\Delta t$ is ignored when the sum in Eq. 1 is built.

The recursion in Eq. 1 proceeds through the whole time series until it reaches the final sample at $T = N\Delta t$. Then using Eq. 3 the likelihood Lik($\mathbf{K}$) is calculated according to

$$\begin{aligned} \mathrm{Lik}(\mathbf{K}) &= \mathrm{Prob}(y_1, y_2, \cdots, y_T \mid \mathbf{P}(\mathbf{K}), f_i(y), a_0(i)) \\ &= \sum_{i=1}^{L} a_N(i). \end{aligned} \tag{6}$$

The $a_0(s)$ are the initial probabilities for each state in the model already mentioned above.

To prevent underflow, scaling is applied to the $a_k(i)$ depending on time k, but not on state i (Levinson et al., 1983; Fredkin and Rice, 1992a).

Extension of the single-channel approach to multiple channels

A direct approach to the extension to multi-channel analysis is the introduction of a "macro-channel" comprising all possible states of the involved ensemble of channels. If we assume M channels with the individual channel having L_i states, the number $L^{(\mathrm{Macro})}$ of states of the macro-channel is

$$L^{(\mathrm{Macro})} = \prod_{i=1}^{M} L_i. \tag{7}$$

This leads to $L^{(\mathrm{Macro})^2}$ elements in the macro-channel's $\mathbf{K}$ matrix and results in a large increase of computer time for the computation of the matrix exponential as compared with the single-channel analysis.

Here a simplification is presented that can be applied if the channels are independent as can be tested according to Dabrowski and McDonald (1992). Our method makes use of a specific property of the Kronecker product of matrix-exponentials (the Kronecker product is also often referred to as tensor product).

The following example illustrates how the transition probability matrix $\mathbf{P}_2$ of a system of two independent and identical channels is related to the single-channel transition rate matrix $\mathbf{K}$.

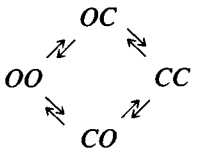
The transition probability matrix **P** for the single-channel model

$$C \rightleftarrows O \tag{8}$$

is given by

$$\mathbf{P} = \begin{pmatrix} p_{OO} & p_{OC} \\ p_{CO} & p_{CC} \end{pmatrix}. \tag{9}$$

The two-channel system has four distinct states.



(10)

Because the single channels are assumed to switch independently, the probabilities of transitions between the states of the macro-channel are computed by multiplying the transition probabilities of the individual channels. For example, the probability $p_{(OO)(OC)}$ for a transition from compound state (OO) to (OC) is computed according to Eq. 11.

$$p_{(OO)(OC)} = p_{OO}p_{OC} \tag{11}$$

The complete transition probability matrix $\mathbf{P}_2$ for a system of two independent and identical single channels of the type given by Eq. 8 is computed as the Kronecker product of the single-channel transition probability matrices **P** (Steeb, 1991)

$$\mathbf{P}_2 = \begin{pmatrix} p_{OO}p_{OO} & p_{OO}p_{OC} & p_{OC}p_{OO} & p_{OC}p_{OC} \\ p_{OO}p_{CO} & p_{OO}p_{CC} & p_{OC}p_{CO} & p_{OC}p_{CC} \\ p_{CO}p_{OO} & p_{CO}p_{OC} & p_{CC}p_{OO} & p_{CC}p_{OC} \\ p_{CO}p_{CO} & p_{CO}p_{CC} & p_{CC}p_{CO} & p_{CC}p_{CC} \end{pmatrix} = \begin{pmatrix} p_{OO}\mathbf{P} & p_{OC}\mathbf{P} \\ p_{CO}\mathbf{P} & p_{CC}\mathbf{P} \end{pmatrix} = \mathbf{P} \otimes \mathbf{P} \tag{12}$$

The Kronecker sum is defined as follows (Horn and Johnson, 1991):

$$\mathbf{A} \oplus \mathbf{B} \equiv (\mathbf{I}_B \otimes \mathbf{A}) + (\mathbf{B} \otimes \mathbf{I}_A), \tag{13}$$

with **A** and **B** being real *nxn* and *mxm* matrices and $\mathbf{I}_A$ and $\mathbf{I}_B$ the corresponding unit matrices. As shown by Steeb (1991), the following theorem holds:

$$\exp(\mathbf{A} \oplus \mathbf{B}) \equiv \exp(\mathbf{A}) \otimes \exp(\mathbf{B}). \tag{14}$$

In the case of several independent and identical channels, the rate matrix is expressed as the Kronecker sum of the constituent rate matrices of the single-channels:

$$\exp(\mathbf{A}_1 \oplus \cdots \oplus \mathbf{A}_M) \equiv \exp(\mathbf{A}_1) \otimes \cdots \otimes \exp(\mathbf{A}_M) \tag{15}$$

The conversion of the Kronecker product of the single-channel **P** matrices (Eq. 12 and top line of Eq. 16) to the Kronecker sum of the single-channel **K** matrices (last line of Eq. 16) is illustrated for $\mathbf{P}_2$ (Eq. 12).

$$\begin{aligned}
\mathbf{P}_2 &= \mathbf{P} \otimes \mathbf{P} = \exp(\mathbf{K}\Delta t) \otimes \exp(\mathbf{K}\Delta t) = (\mathbf{I}_1 + \mathbf{K}\Delta t + O((\Delta t)^2)) \otimes (\mathbf{I}_1 + \mathbf{K}\Delta t + O((\Delta t)^2)) \\
&= \begin{pmatrix} 1 - 2k_{OC}\Delta t & k_{OC}\Delta t & k_{OC}\Delta t & 0 \\ k_{CO}\Delta t & 1 - (k_{CO} + k_{OC})\Delta t & 0 & k_{OC}\Delta t \\ k_{CO}\Delta t & 0 & 1 - (k_{CO} + k_{OC})\Delta t & k_{OC}\Delta t \\ 0 & k_{CO}\Delta t & k_{CO}\Delta t & 1 - 2k_{CO}\Delta t \end{pmatrix} + 0((\Delta t)^2) \\
&= \mathbf{I}_2 + \begin{pmatrix} -2k_{OC} & k_{OC} & k_{OC} & 0 \\ k_{CO} & -(k_{CO} + k_{OC}) & 0 & k_{OC} \\ k_{CO} & 0 & -(k_{CO} + k_{OC}) & k_{OC} \\ 0 & k_{CO} & k_{CO} & -2k_{CO} \end{pmatrix} \Delta t + O((\Delta t)^2) \\
&= \exp\left\{\left[\left[\begin{pmatrix} -k_{OC} & k_{OC} & 0 & 0 \\ k_{CO} & -k_{CO} & 0 & 0 \\ 0 & 0 & -k_{OC} & k_{OC} \\ 0 & 0 & k_{CO} & -k_{CO} \end{pmatrix} + \begin{pmatrix} -k_{OC} & 0 & k_{OC} & 0 \\ 0 & -k_{OC} & 0 & k_{OC} \\ k_{CO} & 0 & -k_{CO} & 0 \\ 0 & k_{CO} & 0 & -k_{CO} \end{pmatrix}\right]\right]\Delta t\right\} \\
&= \exp\{[\mathbf{I}_1 \otimes \mathbf{K} + \mathbf{K} \otimes \mathbf{I}_1]\,\Delta t\} = \exp\{[\mathbf{K} \oplus \mathbf{K}]\Delta t\},
\end{aligned} \tag{16}$$

with $\mathbf{I}_1$ and $\mathbf{I}_2$ being the unity matrices of the one- and two-channel systems, respectively. The terms of quadratic or higher order are omitted.

Comparison of two methods for computing the multi-channel transition matrix P

Method 1 computes the Kronecker sum of the **K** matrices before calculating the matrix exponential by Padé approximation (last line of Eq. 16). Method 2 first computes the transition probability matrix for the single channel by Padé approximation and then computes the multi-channel transition matrix by use of the Kronecker product (top of Eq. 16).

According to Golub and Van Loan (1983), the number of floating point operations (FLOPS) required to calculate the matrix exponential of a *nxn* matrix **P** using Padé approxi-

mation is of the order n^3. The number of FLOPS involved in computing the Kronecker product of M nxn matrices (Eq. 15) (let $M \geq 2$ be the number of channels with L states each) is

$$\mathrm{FLOPS}^{(\text{Kronecker Product})} = \sum_{i=2}^{M} n^{2i}. \qquad (17)$$

In the case of method 1, n equals L^M and the computation of a Kronecker sum of M matrices needs twice the number of FLOPS as that required for obtaining the Kronecker product of M matrices (Eq. 14). Thus, the number of FLOPS is

$$\mathrm{FLOPS}^{(\text{Method 1})} = L^{3M} + 2\sum_{i=2}^{M} L^{2i}, \qquad M \geq 2. \qquad (18)$$

In the case of method 2, n equals L and the Kronecker product is obtained directly:

$$\mathrm{FLOPS}^{(\text{Method 2})} = L^3 + \sum_{i=2}^{M} L^{2i}, \qquad M \geq 2. \qquad (19)$$

Equations 18 and 19 show that method 1 needs an exponentially increasing number of FLOPS in excess of method 2, as illustrated in Fig. 1 below. Even in the case of $M = 2$ channels with $L = 3$ states, method 2 consumes about 10 times less FLOPS than method 1.

There is no difference between the two methods with respect to accuracy. Both methods gave equal elements of the **P** matrix when calculations were done with double precision. Very rarely, there was a difference in the least significant digit.

Generation of the time series of patch clamp currents by computer simulation

For testing of the analysis described in this paper, time series of channel current have to be simulated and then fitted. The simulation starts from a Hidden Markov Model with a set of assumed rate constants for the transitions between the states of the single channel model. The analysis has to reveal these rate constants from the simulated time series.

The process of generating the randomly switching time series involves three steps, with the first two steps corresponding to the simulation method given by Läuger (1988). The system starts at state i, and the output provides the related current y_i. Then, the lifetime t_i of state i is calculated by using the fact that the distribution of life times for the individual state in Markov models is described by a single decaying exponential of the type

$$\chi_i = \exp\left[-\left(\sum_{j=1\neq i}^{L} k_{ij}\right) t\right]. \qquad (20)$$

An evenly distributed random number generator proposes a number χ_i from the interval [0,1). The related dwelltime t_i is obtained from an inversion of Eq. 20 and determines how long the model remains in state i.

After that time t_i, a transition to state ji has to occur. For this purpose, the states of the system are represented by fields. These fields do not only give the number of the state. Especially, the probability of the transition from state i to state j is introduced by the width W_j of the individual field.

$$W_j = \frac{k_{ij}}{\sum_{j=1\neq i}^{L} k_{ij}} \qquad i \neq j \in \{1, \cdots, L\} \qquad (21)$$

L represents the number of states of the model (Colquhoun and Hawkes, 1982). The widths of the individual fields W_j sum up to one.

Again, a random generator gives a number from the interval [0,1). This number is used to point into one of the fields W_j, thus selecting the state j. The probability W_j of selecting this state is given by the width of the related field.

Finally, the anti-aliasing filter is included. The cutoff frequency (f_{-3dB}) of this filter was set to a quarter of the sampling frequency. The filter was a fourth-order Bessel low-pass similar to that one used in the actual recordings. The unit step response of this filter is stored in the computer.

The step in current related to the jump from state i to state j (see above) is used as an input signal for the anti-aliasing filter. Every step in the time series resulting from the above Markov process causes a step response of the filter. Thus, the final time series is a superposition of delayed and weighted (by the appropriate current difference between consecutive states) filter responses.

However, because of the exponential decay of the envelope of these step responses, previous responses can be neglected when their envelope's magnitude drops below the resolution of the recording apparatus, which is 1/4096 in our case. This provides a limit for the number of responses that have to be taken into account. The value of this threshold results from the use of a 12 bit A/D converter in the actual recordings. The above routine is repeated until a time series of adequate length is created.

The spectrum of the added noise exhibits a shape similar to that observed in actual patch clamp recordings done in our laboratory. It is blue noise (resulting from the pipette and other sources as described by Sigworth, 1983), which shows a linear increase in spectral density starting at about 1 kHz to the cutoff frequency of the anti-aliasing filter (Schultze and Draber, 1993).

For the generation of the time series of the blue noise from white noise, an impulse-invariant transformed digital implementation of the adequate low- and high-pass filters was used, running at a clock rate of 10 times the sampling frequency.

The simulated channel current was sampled at 10 or 200 kHz and stored for the subsequent test of the analysis. The sampling rate depended on the rate constants of the model. It was chosen 3 to 5 times higher than the fastest expected rate to keep the deviations from the theoretical Gaussian amplitude histograms as low as possible.

Fitting to experimental data

The aim of many patch clamp experiments is the determination of the k_{ij} of **K**. They have to be evaluated from the measured time series. This is done by a fitting routine. In our

applications, we use the downhill-simplex algorithm of Caceci and Cacheris (1984). Fitting proceeds along the following routine.

0) Starting procedure: guesses are made for the initial probabilities $a_0(s)$ (from the steady-state distributions, Eq. 1) and the initial set of rate constants k_{ij} (e.g., from dwelltime histograms). The vertices of the starting simplex are computed from this initial guess of the rates by multiplying each of them with random numbers in the range from 0.5 to 1.5.

Next, the probabilities $f_i(y)$ are estimated from the amplitude distribution of the current related to state i (Eq. 1). For this purpose, the amplitude histogram is fitted by a sum of Gaussian distributions centered at the mean values of the current related to state i. Gaussian distributions hold for unfiltered records superimposed by white noise. However, the analysis also allows for minor deviations from these assumptions as long as the related distributions can be fitted by (broadened) Gaussians.

We prefer the evaluation of the amplitude histogram, because it is done very quickly (within a few minutes) as compared with an automated procedure (e.g., a Baum-Welch iteration) taking up to several hours of computer time for long data sets (our data sets are in the order of 10^4–10^6 samples).

We are aware that automated analysis is useful, especially in the presence of a poor signal-to-noise ratio (snr). Chung et al. (1991) report successful automated estimation of μ_i and σ_i for a snr being as low as 0.5. In our experiments, we typically deal with signals having a snr of 2–20.

Then, the fitting routine starts.

1) The constraints are satisfied by setting

$$k_{ij} = |k_{ij}| \qquad i \neq j \in \{1, \cdots, L\}$$
$$k_{ii} = -\sum_{j=1 \neq i}^{L} k_{ij} \quad i \in \{1, \cdots, L\}. \tag{22}$$

2) Evaluation of $\mathbf{P(K)} = \exp(\mathbf{K}\Delta t)$ with the numerical computation based on the Padé approximation (Golub and Van Loan, 1983).

3) Computation of the likelihood as described by Eq. 6.

4) Modification of the rates $\mathbf{K}$ according to the downhill-simplex strategy and continue with step 1 or stop, if the simplex diameter shrinks below 10^{-7} (this threshold depends on the machine's accuracy).

The fit may fail when it gets caught in a local maximum of the likelihood. To circumvent this case, the fit should be carried out several times, each time starting with different initial values for the set of k_{ij} (and thus with a completely new simplex).

RESULTS

A numerical example illustrating the benefit of method 2 for the calculation of the matrix of transition probabilities

Fig. 1 shows the number of FLOPS required for methods 1 (corresponding to Eq. 16 bottom, *filled triangles*) and 2 (Eq.16 top, *open squares*) on a semilogarithmic scale for a model having three states and one to four channels.

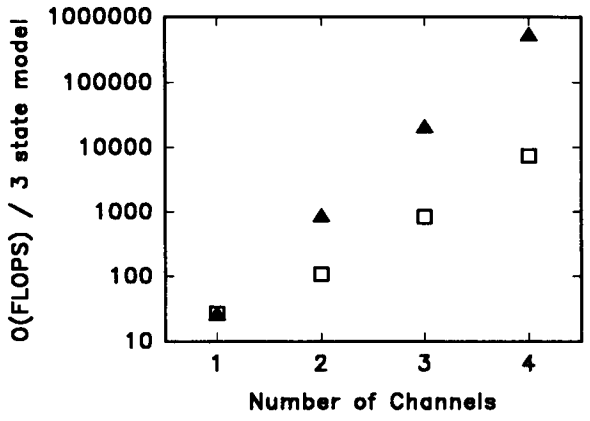


FIGURE 1 Semilogarithmic plots of the number of FLOPS (Eqs. 18 and 19) required for the computation of the transition probability matrix by the two methods given in the text. Method 1 (first Kronecker sum, then matrix exponential; ▲) and method 2 (compute exponential, then multi-channel transition-matrix by Kronecker product; □) are compared for M = 1–4 channels and L = 3 states.

Method 2 is clearly superior to method 1 because it needs two orders of magnitude less FLOPS than method 1 for a system of four channels with three states each. Even if there were only two channels present, method 2 computes about 10 times faster than method 1.

We implemented both methods and did not find any significant numerical differences between the resulting transition matrices using double precision arithmetic. The calculation of the multi-channel transition matrix occurs up to some hundred times during the parameter optimization and, thus, it is worthwhile to keep the number of FLOPS required for its computation as small as possible. The computation of the likelihood, however, consumes most of the CPU time, especially with long (i.e., more than 10^5 samples) time series. In the simulations presented here, we always used the second method.

Testing the multi-channel analysis by means of simulated data

The multi-channel analysis was tested by means of the three channel models described below. In each case, the time series of patch clamp current were generated as described above and blue noise was superimposed. Then the time series was subject to the multi-channel analysis, and it was tested whether the analysis could reveal the original rate constants used for the generation of the time series. Calculations were done on an IBM PC-compatible computer (486DX, 50 MHz, OS/2 2.1 operating system) as well as on UNIX workstations (Sun Sparc 10/30, DEC 3000-500X, and HP9000 Model 720). The programs were written in ANSI C and are available on request.

The power spectrum of the noise used in the first two examples (simulated at 200 kHz sampling rate) is shown in Fig. 2. The power spectrum of the noise for the third example

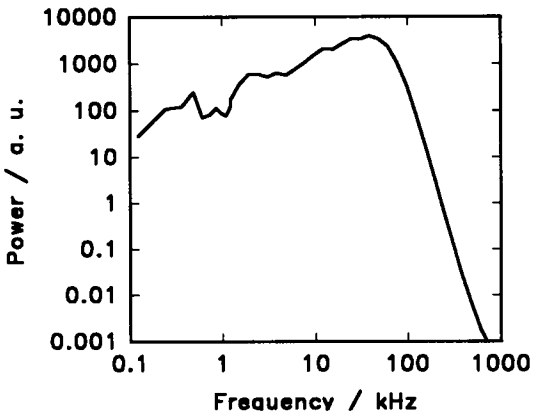


FIGURE 2 Power spectrum of the low-pass-filtered blue noise superimposed to the channel current. The time series of the noise comprised 10^6 samples and was generated at 2 MHz, highpass-, and then low-pass-filtered by a four-pole Besselfilter with a cutoff frequency (−3 dB) of 50 kHz.

is the same as in Fig. 2, but attenuated at high frequencies by an anti-aliasing filter of fourth order with a cut off frequency at 2.5 kHz.

Two-state model

The model and the rate constants are given by the following scheme

$$O \underset{1000\,s^{-1}}{\overset{10000\,s^{-1}}{\rightleftharpoons}} C \qquad (23)$$

with the current of the open state being 10 pA. Simulations were done at snr varying among 2, 5, and 10 and for channel numbers from 1 to 4.

Fig. 3 exhibits a 3-ms section of the simulated current from a four-channel patch. With a sampling rate of 200 kHz and 10^6 data points, the whole time series lasted 5 s. The corner frequency (−3 dB) of the four-pole Besselfilter was set to 50 kHz. The snr was 5.

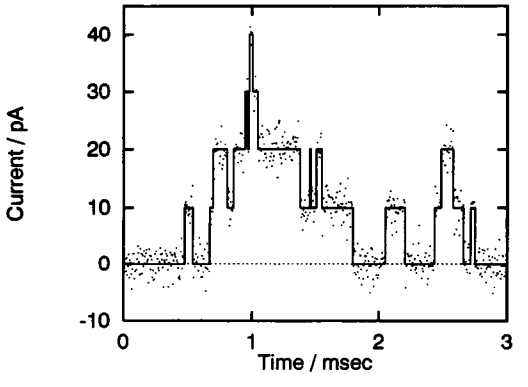


FIGURE 3 3-ms section from a time series of simulated patch current. The data were generated by four identical channels given by the two-state model of Eq. 23. The current of the single open channel was 10 pA. Sampling rate was 200 kHz. The 4-pole Besselfilter was set to 50 kHz. The noise with the spectrum of Fig. 2 was superimposed with an snr of 5. The dots indicate the noisy and filtered data, whereas the solid line gives the output of the Markov model.

The analysis started with the generation of the amplitude histogram from the whole time series. The amplitude histogram (Fig. 4, for a snr of 5) was split into the sum of four Gaussians, which were used as $f_i(y)$ Eq. 1) with $i = 1$ to 4.

Fig. 5 *A–C* show the ability of the multi-channel analysis to reveal the rate constants k_{OC} and k_{CO}, which were used for simulating the time series. Fitting of the time series with the simplex algorithm started from three different sets of starting values for the rates. They were "correct," "wrong by a factor of 3," and "wrong by a factor of 10." Guesses for the initial values of the rates k_{ij} can be obtained from the inverse time constants of dwelltime histograms. Thus, the case "wrong by a factor of three" covers the situation of real applications. As fits were done for time series with one to four channels at three different snrs, there were $3 * 4 * 3 = 36$ runs of the fitting routine for this model.

The data points depicted in Fig. 5 *A–C* show the deviations of the estimated k_{ij} in percent of the "real" ones, which have been used for the simulation of the time series. Different starting values resulted in nearly identical fits. The crosses (k_{OC}) and the open squares (k_{CO}) give mean values of the estimated rate constants as obtained from fitting with three different sets of starting values. The error bars indicate maximum and minimum values obtained with these different sets of starting values. In Fig. 5, *A* and *C*, the error bars are not visible because all starting values resulted in identical estimates.

In all cases, the slow rate constant (k_{CO}, *open squares*) could be revealed with an error less than 15%. This also holds for k_{OC} (*crosses*) for channel numbers up to three for all snrs and even for four channels at a snr of 5. A surprising feature, which was consistent in different variations of the fitting conditions, was the high error of k_{OC} for four channels, even at snr = 10. A 50% error in the case four channels at a snr of 2 may still be regarded as a reasonable result, depending on the expectations of the experimenter. A possible reason for the less reliable fits of k_{OC} is its magnitude, because it is faster than k_{CO}. Probably, it may become a problem to distinguish fast closures from blue noise.

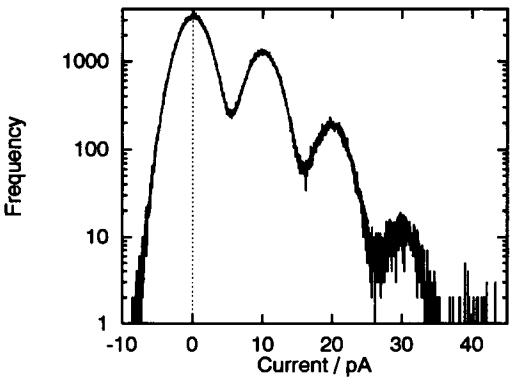


FIGURE 4 Amplitude histogram of the patch current generated by the two-state model of Eq. 23 comprising 10^6 data points. Four identical channels are involved, and the snr was 5. Note the small peak at 40 pA (corresponding to 4 open channels) caused by the low open probability (~9%) of the single channel.

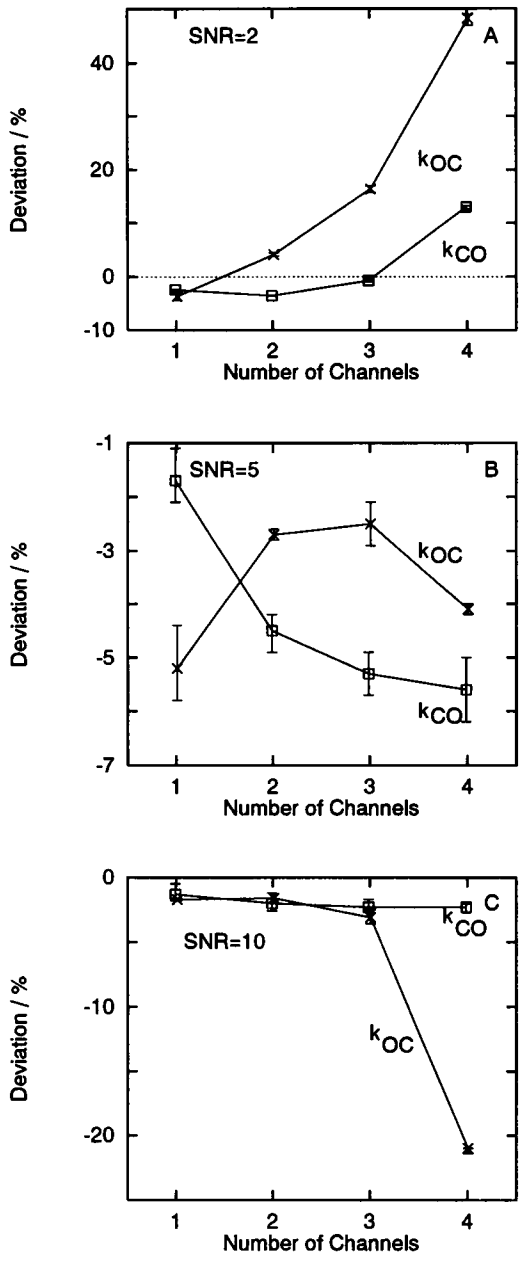


FIGURE 5 Relative fitting errors of the estimated rate constants $\Delta k_{ij}/k_{ij}$ versus channel number for different snr settings in the case of the two-state model (Eq. 23). The open squares and the crosses represent the mean values obtained from fits with three different sets of starting values for k_{CO} and k_{OC}, respectively. The error bars give upper and lower limits, if different starting values resulted in different rate constants. Fitting was done for three different signal-to-noise ratios. (*A*) snr = 2. (*B*) snr = 5. (*C*) snr = 10.

The computer time (using a PC) required by the fitting routine to converge increased dramatically with the number of channels from about 40 min in the case of one channel to about 14 h for four channels. The rise in computer time was the reason that the analysis was restricted to four channels.

Three-state model with sublevels

The second example is a three-state channel having two open states with different conductivities (i.e., sublevels), resulting in currents of 8 pA (O_1) and 11 pA (O_2) according to the following scheme:

$$O_1 \underset{1000\,s^{-1}}{\overset{1000\,s^{-1}}{\rightleftharpoons}} O_2 \underset{20\,s^{-1}}{\overset{400\,s^{-1}}{\rightleftharpoons}} C \qquad (24)$$

This channel is similar to the nutrient-permeable channel found in the intraerythrocytic form of the malaria parasite (Desai et al., 1993). The amplitude histogram for a simulated data set with 10^6 samples containing two channels is given in Fig. 6. The snr with respect to the difference between the currents of the two open levels in this amplitude histogram is 5.

According to the slower rate constants, a sampling rate of 10 kHz was selected. The four-pole Besselfilter was set to 25% of the sampling frequency as in the examples above. Fig. 7 *A–C* show the results of the multi-channel analyzes.

In the case of this model, fits of reasonable quality were obtained only for up to three channels. In the case of four channels, the fitted values of k_{O1O2} and k_{O2O1} were too small by 40–55% at all signal-to-noise ratios. The rate-constants between O_2 and C were too large by 500–800% for all signal-to-noise ratios. Because of this, the four-channel fits are not shown in Fig. 7.

The dependence on the starting values is higher than in the case of the two-state model above. Independence on starting values was found for two rate constants at snr of 2 and 5. In all other cases, the influence seems to be greater. However, it has to be taken into account that the errors imply only a few percent in the error of the absolute values. Because of this, the dependence of the starting values can be neglected for channel numbers up to three. The influence of the snr becomes important in Fig. 7 *A* at snr = 2, even though a deviation of 25% may not be considered to be too serious.

Three-state model

The third example is a model of the three-state K^+ channel in *Chara corallina* similar to that one used by Schultze and Draber (1993). It comprised two closed and one open state

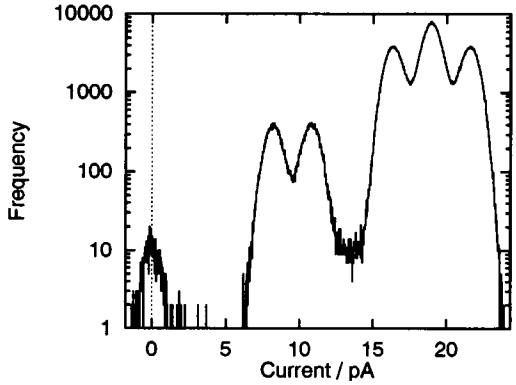


FIGURE 6 Amplitude histogram of the patch current resulting from the three-state model with sublevels (Eq. 24) generated from 10^6 data points. Two channels with two open conductivities each (8 and 11 pA) were simulated, and the snr with regard to the difference of the open levels was 5.

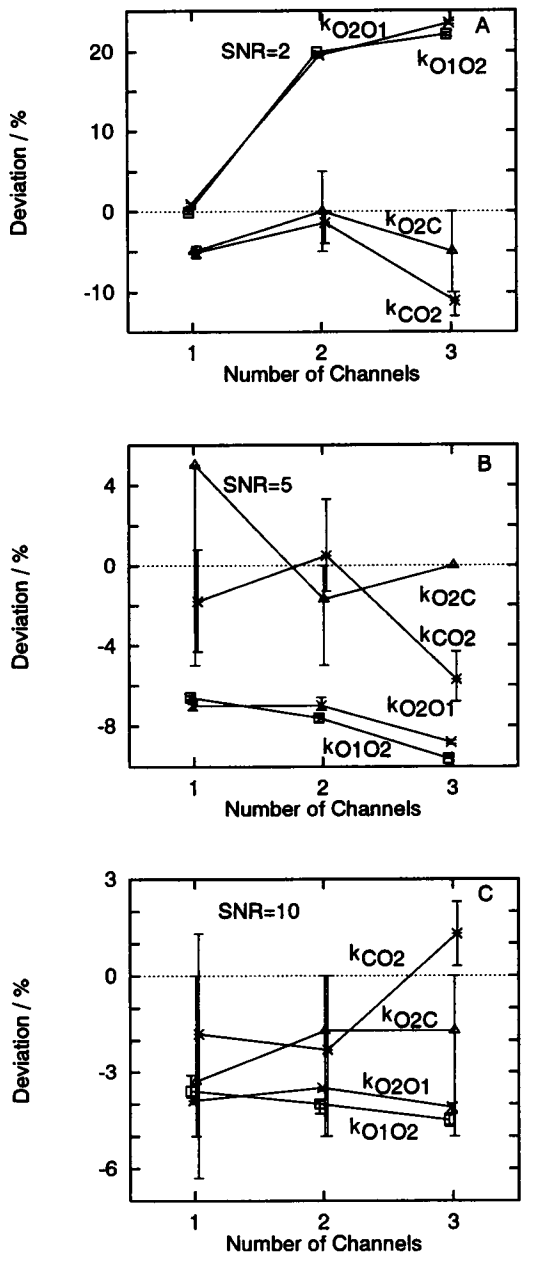


FIGURE 7 Relative fitting errors of the estimated rate constants $\Delta k_{ij}/k_{ij}$ versus channel number for different snr settings in the case of the three-state model with sublevels (Eq. 24). The open squares (k_{O1O2}), crosses (k_{O2O1}), asterisks (k_{CO2}), and triangles (k_{O2C}) represent the mean values obtained from fits with three different sets of starting values. The error bars give upper and lower limits, not standard mean errors. Fitting was done for three different signal-to-noise ratios. (*A*) snr = 2. (*B*) snr = 5. (*C*) snr = 10.

with the rate constants given in the following scheme:

$$C_1 \underset{1000\,s^{-1}}{\overset{70\,s^{-1}}{\rightleftharpoons}} O \underset{300\,s^{-1}}{\overset{15000\,s^{-1}}{\rightleftharpoons}} C_2 \qquad (25)$$

Because the rates are quite fast, a simulated sample rate of 200 kHz was used. Fig. 8 shows the related amplitude histogram for a data set with 10^6 samples from a three-channel patch.

From Fig. 8 (snr = 10) it can be seen that rarely are all channels closed, but yet there are four distinct peaks in the

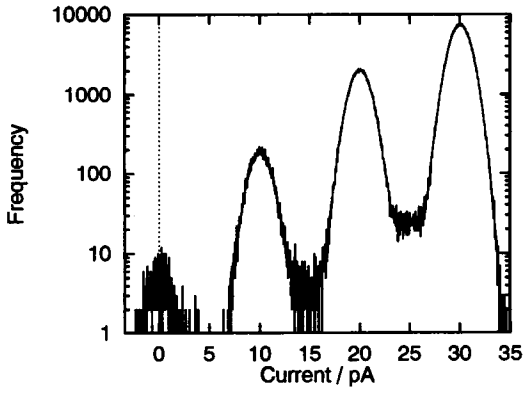


FIGURE 8 Amplitude histogram of the patch current resulting from the three-state model of Eq. 25 generated from 10^6 data points. Three identical channels are involved, and the snr was 10.

amplitude histogram suggesting that there were three channels in the patch. The single-channel current was determined to be 10 pA. Fig. 9 *A–C* give the results of the fits performed for channel numbers from one to four and signal-to-noise ratios ranging from 2 to 10.

In the case of this model, the signal-to-noise ratio plays a more important role than in the models above. For snr = 2, the fits of data sets containing more than two channels did not converge. Because of this, no results can be shown.

In contrast, in the simulations with better snr (5 and 10), all fits succeeded up to channel numbers of four and yielded good estimates of the rate constants (Fig. 9, *B* and *C*). This shows that no general rules about the convergence of the algorithm can be given. In the above example of Eq. 24, it was the channel number that imposed a limit. Here, noise seems to be the more serious problem. Fits for five channels were not tested because of the rise in computer time. Already the four-channel calculations required 24 h on a workstation (HP 9000).

There seems to be a stronger dependence on starting values than in the above models as indicated by the wider error bars in Fig. 9, *B* and *C*. However, also here the effect of the error bars on the absolute errors of the estimation is not serious.

Estimation of the channel number

The knowledge of the channel number is of crucial importance for the generation of the predicted probabilities in Eq. 1, which in turn is being used for calculating the likelihood. In the examples above, this number was known when the algorithm was applied to simulated data. However, this number has to be extracted from the measured data in a real experiment. Horn (1991) discusses seven methods for the estimation of the channel number. Here, we show that the multi-channel analysis itself can reveal the correct channel number.

To test the influence of an incorrect estimation of the channel number, the analysis described above was repeated with wrong channel numbers and the resulting value of the maximum likelihood was used for further analysis.

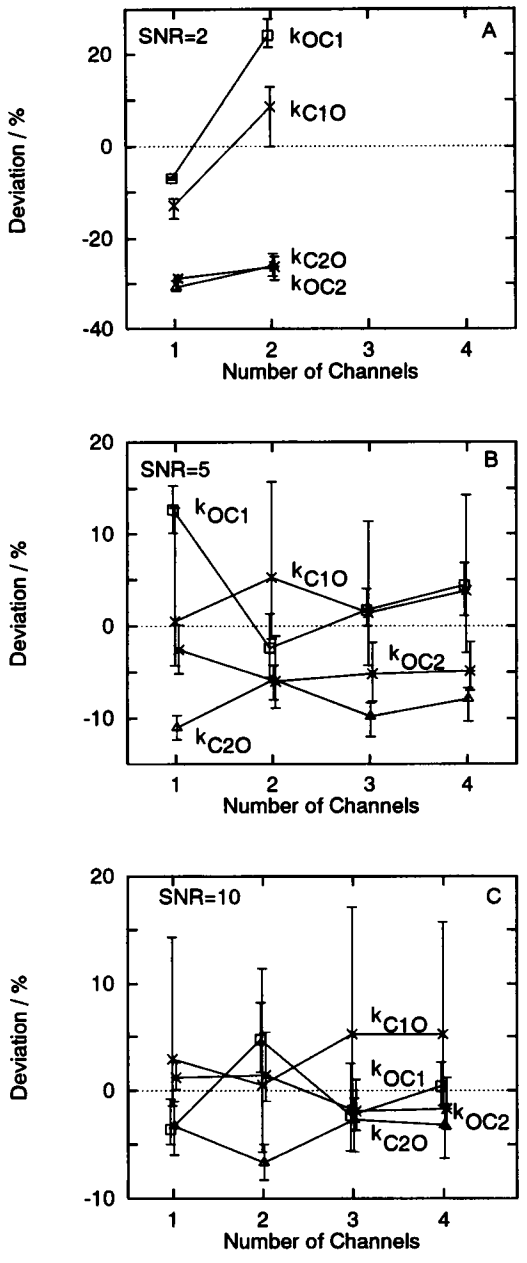


FIGURE 9 Relative fitting errors of the estimated rate constants $\Delta k_{ij}/k_{ij}$ versus channel number for different snr settings in the case of the three-state model with two closed states (Eq. 25). The open squares (k_{OC1}), crosses (k_{C1O}), asterisks (k_{OC2}), and triangles (k_{C2O}) represent the mean values obtained from fits with three different sets of starting values. The error bars give upper and lower limits, not standard mean errors. Fitting was done for three different signal-to-noise ratios. (*A*) snr = 2. (*B*) snr = 5. (*C*) snr = 10.

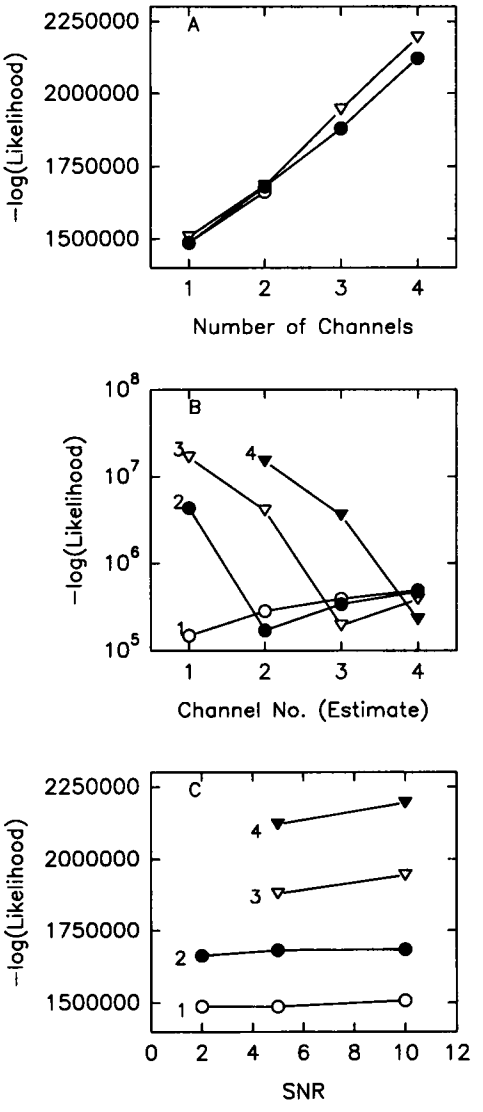


FIGURE 10 Dependence of the negative log likelihood on snr and channel number calculated for the three-state model with two closed states (Eq. 25). (*A*) negative log likelihood versus channel number. The data shown originate from fits of data sets with snr 2 (○), 5 (●), and 10 (△). In the case of snr = 2, the fits failed for more than two channels as already shown in Fig. 9 *A*. (*B*) Dependence of the negative log likelihood on the channel number used for the fitting routine (x axis), which may differ from the true channel number used in the simulations (parameter attached to the curves with one (○), two (●), three (△), and four (▲) channels). Each data set consisted of 10^5 samples. snr = 5. (*C*) −log likelihood versus snr: the likelihood for different channel numbers does hardly change for snr from 2 to 10. The curves differ by the number of channels, which was 1 in the bottom curve and 4 in the top curve as indicated by the attached numbers. In the two upper traces, only fits for snr of 5 and 10 succeeded, and so there are no data points for a snr of 2.

Fig. 10 shows the results of fits, done on the model of Eq. 25 (three-state model with two closed states). In Fig. 10 *A*, the true channel number was identical to the channel number of the model used for the prediction. The likelihood decreases with increasing channel number as indicated by the increase of the negative logarithm of the likelihood.

However, if the true channel number and the channel number of the model are different, a nonmonotonous dependence is found, as is shown in Fig. 10 *B*. The negative logarithm of the likelihood is plotted versus the assumed channel number. A line in Fig. 10 *B* is related to a fixed "true" channel number as attached to the curve. The important finding is that this curve takes a minimum (maximum of the likelihood) when the assumed channel number is identical to the "true" channel number. In the case of the four-channel data set, the likelihood value for the trial with one channel assumed is omitted, because the fitting routine failed. Anyway, this case is of minor interest because it is unlikely that only one open level is observed in a four-channel record.

The findings of Fig. 10 suggest a procedure for determining the "true" channel number. The fitting of the time series should be done with different channel numbers, and that number giving the maximum value of the likelihood should be taken as the "true" number of channels. In effect, a maximum likelihood procedure would be done over channel number as well as over rate parameters.

For this purpose, it is important that our simulations have shown that the snr has little influence on the likelihood (Fig. 10 *C*) as compared with the influence of channel number (Fig. 10 *A*).

In addition, we found that for the determinations of curves like those in Fig. 10 *B* the data set may safely be reduced to 10^5 samples. This speeds up the fitting procedure by a factor of 10. After the correct channel number has been determined, the rate estimates of these short-term fittings can be used as starting values for the subsequent fitting routine with the correct channel number, which now makes use of the whole time series.

DISCUSSION

The examples above show that the multi-channel analysis can reveal the rate constants of the involved model directly from the time series of the patch current at a signal-to-noise ratio that occurs in real experiments. The starting values of the k_{ij} for the simplex algorithm were not important for the final results of the fits. The success of the fitting routine was not better when the correct values were used or when values wrong by a factor of 3–10 were used as starting values. Starting values of the quality used in the simulations can easily be obtained from dwelltime histograms. This feature results from the fact that the first vertices of the simplex are generated automatically from the starting values (see theory above). The operations of the simplex algorithm imply that a correct point in the vertices cannot be reached directly by the strategies of the simplex algorithm.

The above examples reveal the benefits of the direct fit of the time series, which are as follows:

- Multi-channel analysis is possible without restriction to favorite conditions (e.g., burst kinetics, Draber et al., 1993).
- The analysis yields the rate constants k_{ij} themselves. There is no need for setting up and solving complicated equations that accomplish the conversion of amplitude factors and time constants into rate constants as has to be done in the case of approaches that first restore the noise-free time series and then create dwelltime histograms (Schultze and Draber, 1993). The approach presented here utilizes the advantage of a "target" analysis, in which the fit aims directly at the interesting parameters (k_{ij}) and not at intermediate quantities like time constants, which have no direct biophysical meaning.
- There is no need for a missed events correction (Draber and Schultze, 1994; Crouzy and Sigworth, 1990; Dabrowski and McDonald, 1993).
- Sublevel detection is automatically included.
- The true channel number can be determined according to Fig. 10 *B*.

Although the examples presented above are in favor of the method derived in this paper, the reader should keep in mind that there exists an underlying problem of nonidentifiability in the Markovian analysis of patch-clamp records: several distinct kinetic models give rise to the same aggregated current process. In consequence, one can never hope to estimate properly the rate constants in these cases (Ball and Rice, 1992).

A problem that is not analyzed in detail in the above examples is the assumption of white noise. It enters the analysis when the amplitude histogram is split into a sum of Gaussians for the determination of the $f_i(y)$. The simulation results above imply that a failure of this assumption does not bias the estimation of rate constants very much, as long as an adequate choice of the σ_i can account for the distorted Gaussian distribution with reasonable accuracy. This seems to hold for the blue noise of our examples and for the effect of the anti-aliasing filter. However, if fast switching (faster than the cutoff frequency of the anti-aliasing filter) leads to strong deviations from the form of a Gaussian distribution, β-distributions (FitzHugh, 1983) should be used for splitting the amplitude histograms into the $f_i(y)$ curves. Procedures for the fitting of amplitude histograms by β-distributions have been developed in our laboratory (Th. Rießner, personal communication) but have not yet been applied to the fitting of time series.

The method presented here can also be applied to reaction schemes possessing a cyclic or Y-shaped topology. This is not shown here so that we may restrict the number of applications. Cyclic reaction schemes do not impose new problems. In contrast, it is a benefit of our method that further constraints like microreversibility in cyclic schemes can easily be incorporated into the estimation procedure. In the case of a Y-shaped model, the algorithm can be used as is.

The major drawback of the method is its vast demand for computer time (i.e., several hours on a workstation for a 5-s record). Table 1 gives examples for some of the above fits.

This restriction can be overcome by the use of faster computers (high performance workstations) and smarter optimization routines. We are planning to investigate the use of genetic algorithms (GA) for the parameter search (Goldberg, 1989). By the use of GAs, parallel processing is enabled and makes it worthwhile to discuss the use of a cluster of workstations, where each machine computes the likelihoods for a subpopulation. The use of these more sophisticated techniques should cut down computer time needed to a more reasonable level.

TABLE 1 Computer time needed to estimate the parameters of the three-state model (Eq. 25) from a time series with 100,000 samples. The fit was started with "correct" initial guesses. Times are given in HH:MM:SS format

Channel number	DEC alpha	PC
1	21	2:47
2	1:15	13:27
3	6:39	1:09:44
4	2:10:11	9:26:20

The authors thank Dr. Silke Draber, Dr. Roland Schulze, and Dipl. Phys. Thilo Rießner for furnishing the computer programs for the simulation and the procedure for fitting amplitude histograms. We also thank Dr. Günter Radons and Dipl.-Phys. Volker Breuer for helpful discussions and comments concerning the theory of Hidden Markov Models. A. Albertsen was supported by a scholarship granted by the state of Schleswig-Holstein.

REFERENCES

Ball, F. G., and J. A. Rice. 1992. Stochastic models for ion channels: introduction and bibliography. *Math. Biosci.* 112:189–206.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.* 41:164–171.

Bharucha-Reid, A. T. 1960. Elements of the Theory of Markov Processes and their Applications. McGraw-Hill, New York. 468 pp.

Caceci, M. S., and W. P. Cacheris. 1984. Fitting curves to data—the simplex algorithm is the answer. *BYTE.* 5:340–362.

Chung, S. H., V. Krishnamurthy, and J. B. Moore. 1991. Adaptive processing techniques based on Hidden Markov Models for characterizing very small channel currents buried in noise and deterministic interferences. *Phil. Trans. R. Soc. Lond. B.* 334:357–384.

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

Colquhoun, D., and A. G. Hawkes. 1982. On the stochastic properties of bursts of single ion-channel openings and of clusters of bursts. *Phil. Trans. R. Soc. Lond. B.* 300:1–59.

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

Dabrowski, A. R., and D. McDonald. 1992. Statistical analysis of multiple ion channel data. *Ann. Stat.* 20:1180–1202.

Dabrowski, A. R., and D. McDonald. 1993. Consistent estimation in the presence of bandwidth limitation. *J. R. Stat. Soc. B.* 55:267–289.

Desai, S. A., D. J. Krogstad, and E. W. McCleskey. 1993. A nutrient-permeable channel on the intraerythrocytic malaria parasite. *Nature.* 362: 643–646.

Draber, S., and R. Schultze. 1994. Correction for missed events based on a realistic model of the detector. *Biophys. J.* 66:191–201.

Draber, S., R. Schultze, and U.-P. Hansen. 1993. Cooperative behavior of K^+ channels in the tonoplast of *Chara corallina. Biophys. J.* 65: 1553–1559.

FitzHugh, R. 1983. Statistical properties of the asymmetric random telegraph signal with application to single channel analysis. *Math. Biosci.* 64:75–89.

Fredkin, D. R., and J. A. Rice. 1992a. Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. R. Soc. Lond. B.* 249:125–132.

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

Goldberg, D. E. 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Publishing Co., Reading, MA. 412 pp.

Golub, G. H., and C. F. Van Loan. 1983. The matrix exponential. *In* Matrix Computations. G. H. Golub, editor. Johns Hopkins University Press, Baltimore, MD. 396–403.

Horn, R. 1991. Estimating the number of channels in patch recordings. *Biophys. J.* 60:433–439.

Horn, R. A., and C. R. Johnson. 1991. Topics in Matrix Analysis. Cambridge University Press, Cambridge. 607 pp.

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

Huang, X. D., Y. Ariki, and M. A. Jack. 1990. Hidden Markov Models for Speech Recognition. Edinburgh University Press, Edinburgh, Scotland. 276 pp.

Jackson, M. B. 1985. Stochastic behavior of a many-channel membrane system. *Biophys. J.* 47:129–137.

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

Levinson, S. E., L. R. Rabiner, and M. M. Sondhi. 1983. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell System Tech. J.* 62:1035–1074.

Rabiner, L. R. 1989. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. Inst. Elect. Eng.* 77: 257–285.

Schultze, R., and S. Draber. 1993. A nonlinear filter algorithm for the detection of jumps in patch-clamp data. *J. Membr. Biol.* 132:41–52.

Sigworth, F. J. 1983. Electronic design of the patch clamp. *In* Single-Channel Recording. B. Sakmann and E. Neher, editors. Plenum Press, New York. 3–35.

Steeb, W.-H. 1991. Kronecker Product of Matrices and Applications. Bibliographisches Institut & F. A. Brockhaus AG, Mannheim, Germany. 135 pp.

Yellen, G. 1984. Ionic permeation and blockade in Ca^{2+} activated K^+ channels of bovine chromaffin cells. *J. Gen. Physiol.* 84:157–186.