

# A New Algorithm for Idealizing Single Ion Channel Data Containing Multiple Unknown Conductance Levels

Antonius M. J. VanDongen

Department of Pharmacology, Duke University Medical Center, Durham, North Carolina 27710 USA

**ABSTRACT** A new algorithm is presented for idealizing single channel data containing any number of conductance levels. The number of levels and their amplitudes do not have to be known a priori. No assumption has to be made about the behavior of the channel, other than that transitions between conductance levels are fast. The algorithm is relatively insensitive to the complexity of the underlying single channel behavior. Idealization may be reliable with signal-to-noise ratios as low as 3.5. The idealization algorithm uses a slope detector to localize transitions between levels and a relative amplitude criterion to remove spurious transitions. After estimating the number of conductances and their amplitudes, conductance states can be assigned to the idealized levels. In addition to improving the quality of the idealization, this "interpretation" allows a statistical analysis of individual (sub)conductance states.

## INTRODUCTION

Ion channels are large integral membrane proteins that regulate the flux of ions across the plasma membrane. Early patch clamp recordings of single ion channel behavior (Neher and Sakmann, 1976) seemed to confirm the "unitary conductance" paradigm (Neher and Stevens, 1977), which suggests that channels switch stochastically between only two distinct permeability states: open and closed. Improvements in the resolution of patch clamp recordings (Hamill et al., 1981) have resulted in the observation of multiple open conformations for a large number of channels. In addition to a main conductance level, there appear to be sub- or superconductance states (Auerbach and Sachs, 1983; Hamill et al., 1983; Geletyuk and Kazachenko, 1985; Nagy, 1987; Matsuda, 1988; Nilius et al., 1989; Ferguson et al., 1993; Root and Mackinnon, 1994). Often there are direct transitions between the different conductance levels, resulting in complex opening patterns.

The analysis of single channel data usually starts with an "idealization" step, in which the behavior of the channel is extracted from the noisy data as a sequential list of amplitudes and durations of all levels. This event list can be used to graphically reconstruct a noiseless single channel record, while it also serves as the input for subsequent statistical analyses. Thus far, idealization of single channel behavior has been dominated by the classical 50% amplitude threshold algorithm (Sachs et al., 1982; Sachs, 1983; Colquhoun and Sigworth, 1983), which employs an amplitude detector set at half the unitary current to detect transitions between the open and closed states. However, this algorithm is not

suitable for analyzing single channel data containing multiple, unknown, possibly non-equidistant amplitude levels.

A new algorithm was designed that does not suffer from this limitation. The algorithm, TRANSIT, uses a slope detector to localize transitions between levels and a relative amplitude criterion to remove spurious transitions. It can be seen as an extension of the approach used in Sigworth's CATCH program (Sigworth, 1983). Part of the algorithm has been published previously in abstract form (VanDongen, 1992). Design objectives included the following desirable properties: 1) no a priori information about the number of conductance levels or their amplitudes is required; 2) no underlying model has to be assumed; 3) there is no intrinsic minimum duration requirement for the levels; 4) objectivity is maximized; and 5) the idealization is perfect in the absence of noise and filtering, no matter how complex the data.

## DESIGN OF THE ALGORITHM

Single channel data, no matter how complex, consist of two basic components: transitions and levels. The objective of an idealization algorithm is to accurately estimate the duration and amplitude of all levels from noisy data. If it were possible to localize the time points at which transitions occur, then idealization would be straightforward. The duration of a level could be determined by the time interval separating neighboring transitions, while the amplitude could be estimated by averaging the data points between two transitions.

Transitions can be distinguished from levels by their slope or first order time derivative. In the theoretical case of noise-free single channel data the slope of a level equals zero, while the slope of a transition is very large (Fig. 1). In this ideal case, all transitions are readily detected by differentiation. The duration and amplitude of each level can be found, and an idealized trace can be reconstructed. The idealization obtained with this approach will be perfect. No matter how complex the data, all levels will be found and

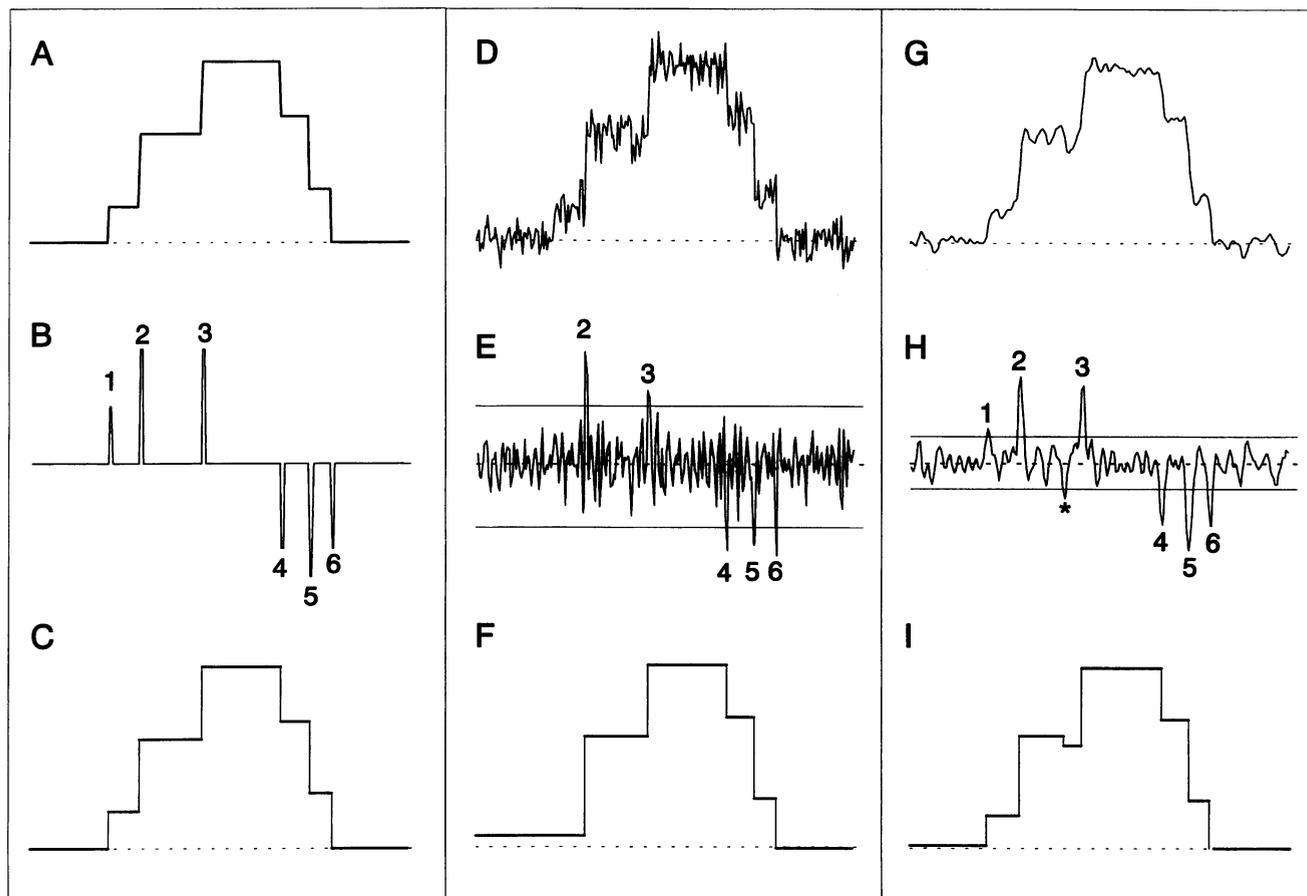
---

*Received for publication 21 August 1995 and in final form 18 December 1995.*

Address reprint requests to Antonius M. J. VanDongen, Department of Pharmacology, Duke University Medical Center, Durham, NC 27710. Tel.: 919-681-4862; Fax: 919-684-8922; E-mail: vando005@mc.duke.edu.

© 1996 by the Biophysical Society

0006-3495/96/03/1303/13 \$2.00



**FIGURE 1** Use of a slope detector to localize transitions. (A) A simulated ideal opening containing six transitions between five levels. (B) Slope or first derivative of A. The slope is zero, except at time points where transitions occur. (C) Idealization using the transitions localized in B. (D) Same record as in A, after addition of white noise. (E) Slope of the record in D. Levels have a non-zero slope due to current fluctuations. Slope threshold detects transitions 2–6, but transition 1 does not exceed the threshold. (F) Idealization using the transitions detected in E. The first level is missing because the first transition was not detected. (G) Same record as in D, after low-pass filtering. (H) Slope of G. All six transitions now exceed the slope threshold, but noise fluctuation in second level causes spurious transition. (I) Idealization using transitions detected in H. Spurious transition split the second level into two artificial levels.

their amplitudes and durations will be correct. Furthermore, it should be noted that no information about the number of levels or their amplitudes is required.

Real data usually contain generous amounts of noise (Fig. 1 D), which forces the use of a slope threshold (Fig. 1 E). This results in two kinds of problems: 1) transitions with small slopes are not detected, and 2) spurious transitions may appear when the fluctuating slope of a level exceeds the threshold. Missed transitions may be recovered by appropriate low-pass filtering, which increases the amplitude resolution at the expense of time resolution (Fig. 1, G–I). Spurious transitions are a fundamental problem associated with this approach. They make the idealization unreliable by introducing artificial levels (Fig. 1 I).

If the problem of spurious transitions could be resolved, then the slope threshold could be made relatively small and very few real transitions would be missed. A spurious transition originates from the noise that is superimposed on a level. Because the two artificial levels that a spurious transition introduces are actually two parts of the same

original level, their amplitudes will be very similar. Therefore, the difference in amplitude between any two consecutive levels is compared with a relative amplitude threshold. If the difference does not exceed the threshold, then the transition that separates them is probably spurious and the levels are concatenated. The final algorithm therefore employs two thresholds: a slope threshold to detect transitions and a relative amplitude threshold to eliminate spurious transitions.

To maximize objectivity, the threshold values are derived from the data. Fig. 2 illustrates how a record is analyzed to estimate the mean and standard deviation for both the amplitude ( $\sigma_{amp}$ ) and the slope ( $\sigma_{slope}$ ). A detailed description of the algorithm is given in Fig. 3.

## IMPLEMENTATION OF THE ALGORITHM

The idealization algorithm can be applied to real single channel data after filtering and digitization of the current

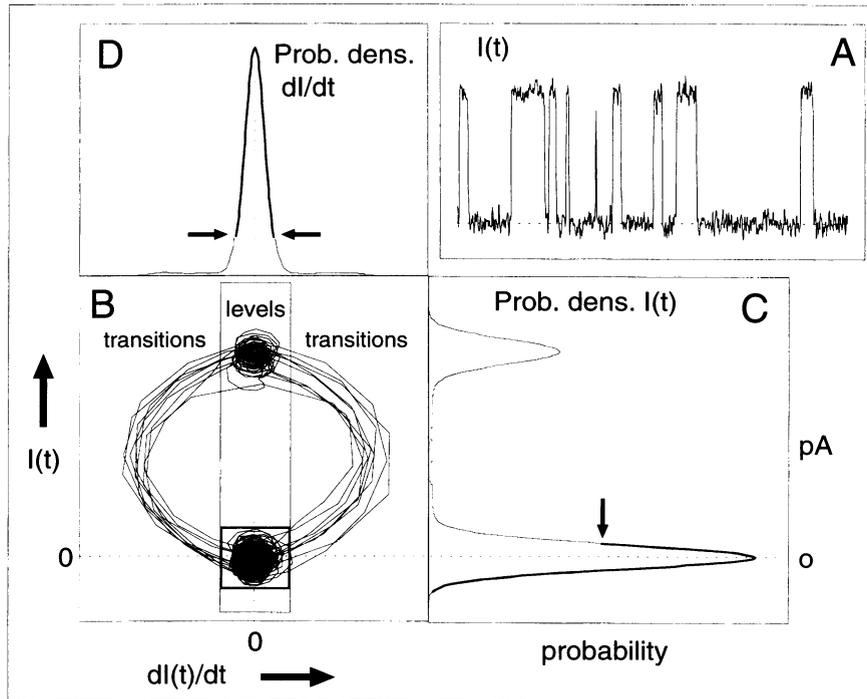


FIGURE 2 Analysis of baseline noise to estimate standard deviations. (A) Simulated single channel record showing baseline noise interspersed with openings. (B) Phase plane representation of the record in A. The current amplitude  $I(t)$  is plotted against the slope  $dl(t)/dt$ , estimated using the central difference (CDIF in Eq. 1). Levels have a relatively small slope, indicated by the large rectangular box. A bold box surrounding the closed level indicates  $\pm 3$  SD of the amplitude and the slope, estimated as described below. (C) Amplitude probability density function (pdf) estimated from the current amplitudes using a Gaussian kernel estimator (Silverman, 1986). It is shown as a projection of the phase plane to its y axis. The open and closed levels give rise to two peaks. The mean and standard deviation of the amplitude can be estimated by fitting a Gaussian curve to the most negative peak, which corresponds to noise in the baseline. All of the negative limb can be used, but only the top part of the positive limb (arrow) should be used, since it can become contaminated with noise from the open state(s) when the S/N ratio is poor. (D) Slope pdf estimated using CDIF. The large peak results from the levels that have a mean slope of 0 and follow a Gaussian distribution. Positive and negative transitions give rise to small tails. The standard deviation of the slope distribution can be estimated by fitting a Gaussian curve to the part of the peak that is not contaminated by the tails (in between the arrows).

record. This produces a discrete time series  $I_t$  ( $t = 1, \dots, N$ ), sampled at  $N$  time points with equidistant spacing  $\Delta t$ . Implementing the algorithm is straightforward. A number of issues deserve elaboration, however.

### Estimation of the first derivative

The time series  $I_t$  needs to be differentiated in order to estimate the first derivative (slope). Different advantages and disadvantages are associated with the two basic estimators for the first derivative at  $I_j$  (the current at time point  $j$ ): the forward difference (FDIF <sub>$j$</sub> ) and central difference (CDIF <sub>$j$</sub> ), which are defined as follows:

$$\text{FDIF}_j = (I_{j+1} - I_j)/\Delta t \quad \text{CDIF}_j = (I_{j+1} - I_{j-1})/2\Delta t \quad (1)$$

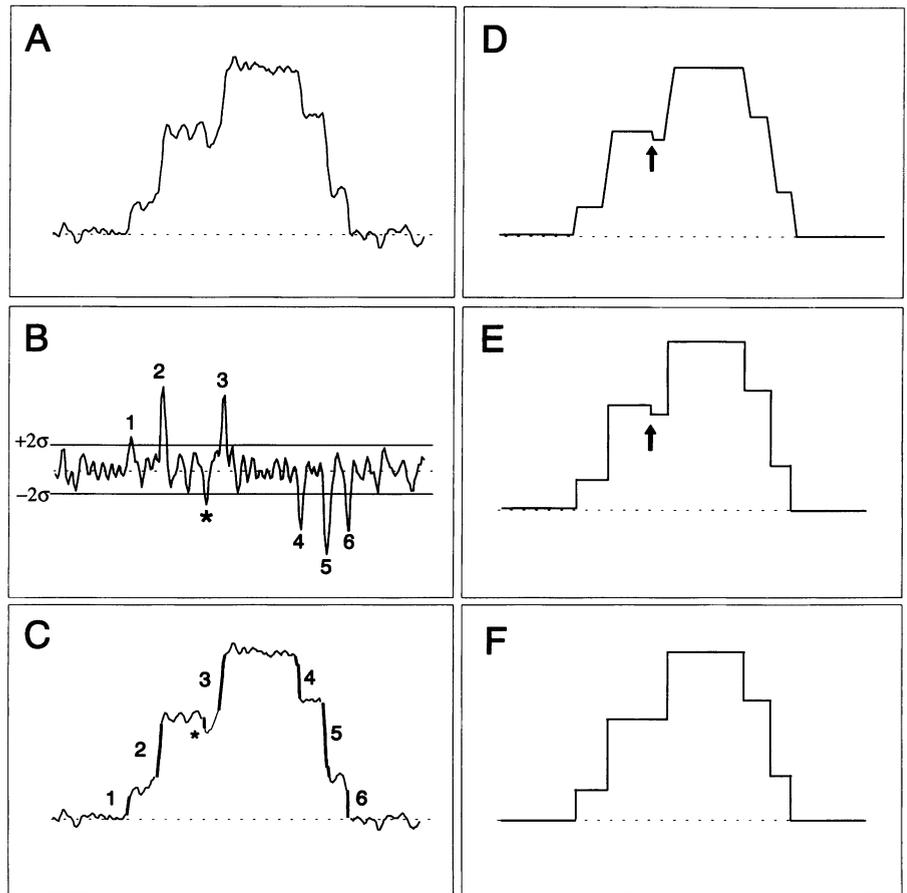
First we will consider the standard deviations  $\sigma_{\text{FDIF}}$  and  $\sigma_{\text{CDIF}}$ , estimated from the baseline noise. When this noise is “white,”  $\sigma_{\text{CDIF}}$  is half the size of  $\sigma_{\text{FDIF}}$ . A slope threshold that employs CDIF is therefore twice as sensitive for detecting real transitions, while being equally sensitive to spurious transitions arising from the noise. Filtering introduces correlation between neighboring points, and the difference in standard deviations becomes smaller than a factor

of two. In the limiting case of severe low-pass filtering, FDIF equals CDIF, and their standard deviations become identical. Simulations showed that for filter frequencies between  $1/3$  and  $1/10$  of the sample frequency,  $\sigma_{\text{FDIF}}$  was between 1.9 and 1.1 times as large as  $\sigma_{\text{CDIF}}$  (data not shown). CDIF is therefore superior to FDIF, because it has better noise-rejection properties.

Unfortunately, this increased sensitivity comes at a price. A slope detector that uses CDIF is “blind” to patterns of fast switching containing consecutive levels of one or two sample points, as illustrated in Fig. 4. Other estimators for the first derivative with even better noise rejection properties (e.g., polynomial approximations (Savitsky and Golay, 1964; Steinler et al., 1972)) suffer from the same problem as the central difference.

To construct a slope detector that is relatively insensitive to fluctuations and has no minimal duration requirement, a combination of forward and central difference estimators has to be used. First, transitions are detected using CDIF and stored. Then, stretches of fast activity are localized using FDIF. For these segments, transition time points detected by FDIF override the previous central difference analysis. This strategy of combining forward and central

**FIGURE 3** The TRANSIT idealization algorithm. (A) A simulated, noisy, low-pass filtered complex opening containing six transitions between five levels. Same record as Fig. 1 G. (B) Slope of record in A, estimated by CDIF. Slope thresholds are set at plus and minus  $2\sigma_{\text{CDIF}}$  estimated from the baseline noise as described in Fig. 2. The threshold is exceeded seven times, resulting in the detection of six real and one spurious (\*) transition. (C) Same record as in A. Parts of the record for which the slope exceeded the threshold are marked as "transition," as indicated by a bold line. All other segments are "levels." (D) Amplitudes and durations of levels are estimated as described. Segments marked as levels have been replaced by a horizontal line with the estimated amplitude and duration. Segments marked as transitions have been replaced by lines connecting the levels. An arrow indicates the spurious transition. (E) Transitions in D have a finite duration because of low-pass filtering, which results in the durations of the levels being underestimated. The duration of each level is restored by adding half of the duration of each flanking transition. (F) In the final phase spurious transitions are removed by comparing the amplitudes of neighboring levels. Levels whose amplitude difference does not exceed the relative amplitude threshold set at three times the amplitude standard deviation ( $\sigma_{\text{amp}}$ ) are concatenated.



difference estimators in the slope detector utilizes the positive aspects of each, without incorporating their respective limitations.

### Unresolved levels

When the duration of a level does not exceed the rise time of the low-pass filter it will not be completely resolved. Such unresolved levels may be detected, if they cause the slope to change sign without going through a level. Two consecutive transitions with opposite sign is therefore an indication of an unresolved level. Both the amplitude and duration of such levels are unreliable, and they are flagged as such.

### The relative amplitude threshold

The probability that a spurious transition will survive the relative amplitude test depends on the duration of the two levels it creates. Since the amplitude of a level is obtained by averaging, the difference between the amplitudes of two levels containing  $N_1$  and  $N_2$  points corresponds to the difference between two means of sample size  $N_1$  and  $N_2$ , drawn from the same Gaussian distribution. The probability that a relative amplitude difference exceeds three standard deviations vanishes very quickly with increasing  $N_1$  and

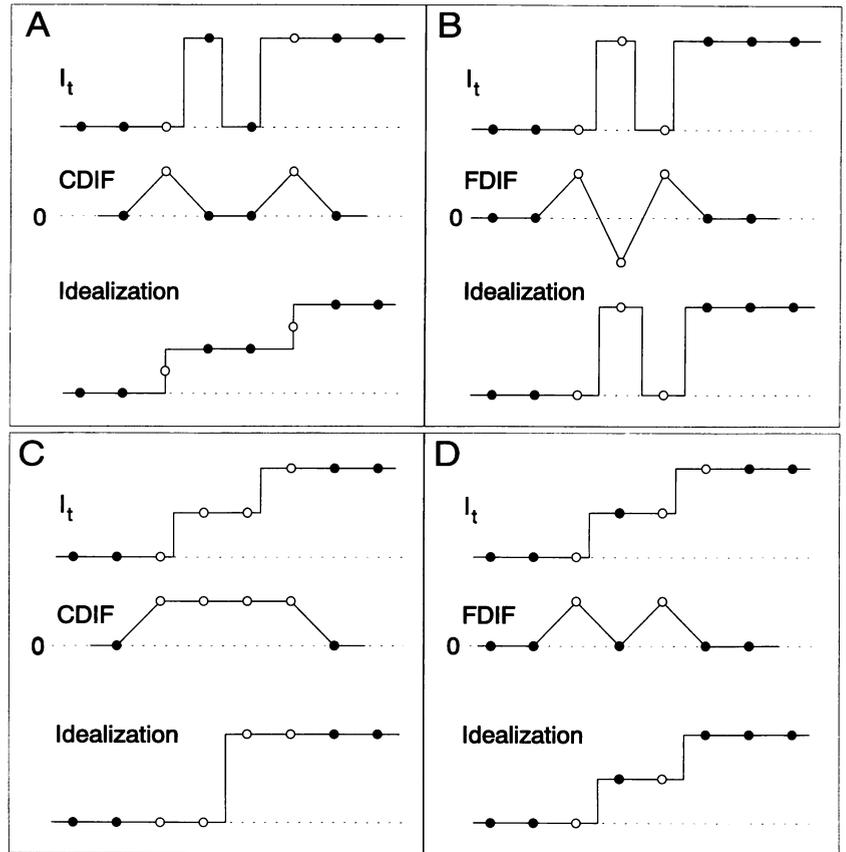
$N_2$ . However, low-pass filtering introduces correlation between neighboring points, increasing the probability that a spurious transition will survive the relative amplitude test. Simulations using filtered noise were used to empirically determine these probabilities. When one of the levels has a short duration (one to three sample points), then the probability of accepting a spurious transition becomes significant. This can be counteracted by increasing the relative amplitude threshold, whenever one of the two levels has a duration of three sample points or less. The effect of this will be further explored in the Results section.

The first and last point of a level may have been used by CDIF to estimate the flanking slopes. If they are used for amplitude estimation as well, then the amplitude and slope criteria are not statistically independent. Therefore, these points are omitted from the amplitude estimation for levels with a duration longer than three sample points.

### Concatenating levels

Neighboring levels are concatenated when their amplitude difference does not exceed the relative amplitude threshold. The amplitude of the new level becomes the average of the amplitudes of the two initial levels, weighted by their durations. Concatenation is an iterative process, the outcome of which depends on the order in which levels are concat-

FIGURE 4 Detection of fast switching activity: forward versus central difference. (A) Sampled ideal current segment  $I_t$  (top), its central difference CDIF (middle) and idealization (bottom). Open and closed levels with a duration of one sample point are flanked by longer levels. The central difference detects two transitions, (O), instead of three, resulting in an incorrect idealization. (B) The same eight sample points as in A are correctly handled by the forward difference FDIF, which detects three transitions, resulting in a correct idealization. (C) A second pattern of activity that CDIF cannot handle: a two-point intermediate-size opening flanked by a longer closed level and full-size opening. The central difference only detects a single transition with a duration of three points. (D) Again, the forward difference detects two transitions and produces a correct idealization.



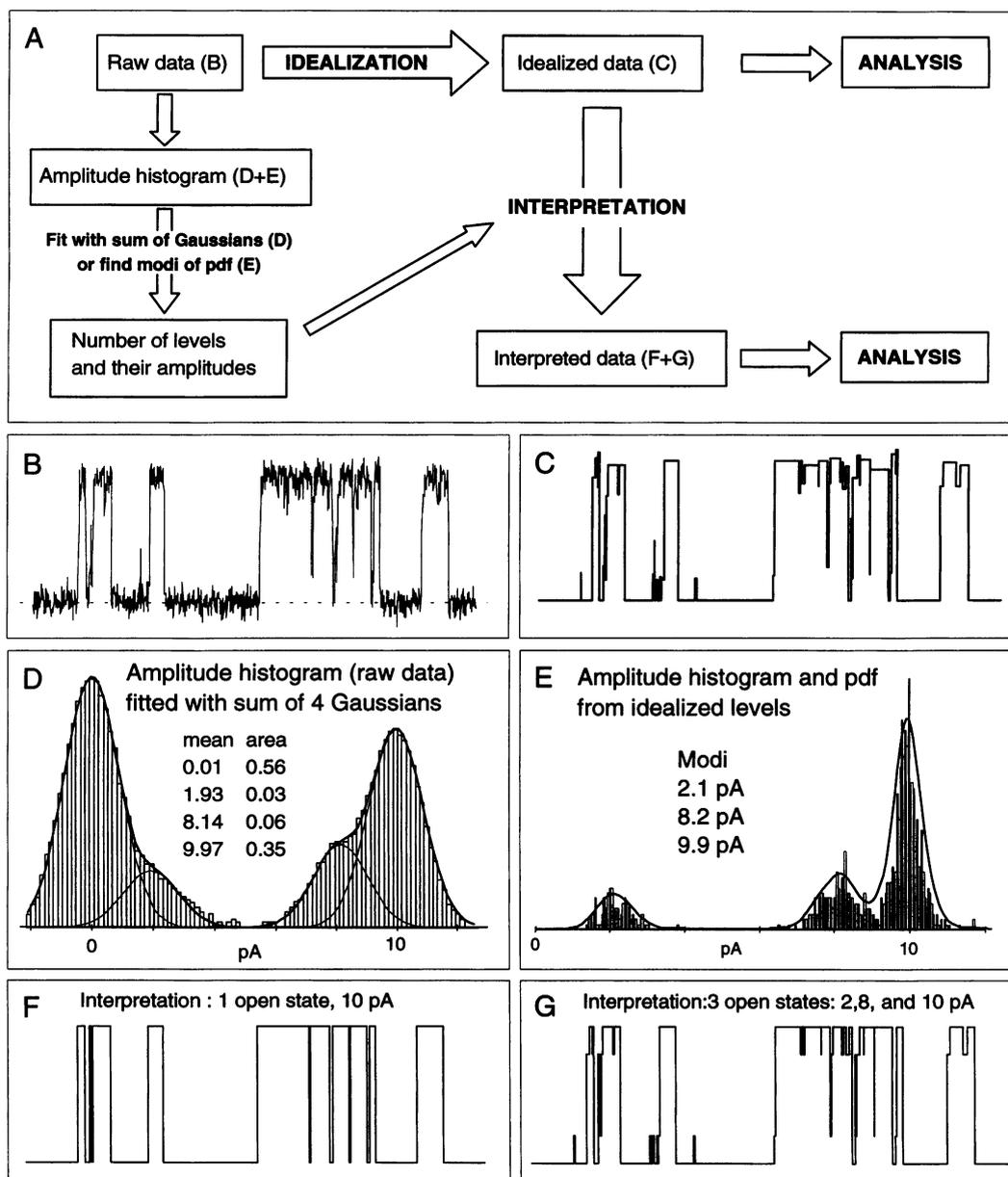
enated. In TRANSIT, the order is determined by the magnitude of the difference: the smallest difference, which is most likely to be spurious, is concatenated first. The flanking differences are then recalculated. This is repeated until all differences exceed the threshold.

**INTERPRETATION**

The TRANSIT algorithm as described thus far produces a sequential list of amplitude-duration pairs. The amplitudes and durations can now be used as input for statistical analysis (Fig. 5 A). All statistics that are generally computed from idealized data can be evaluated at this point. However, dwell time distributions for open and closed levels could be affected by unresolved closures. Also, spurious transitions that escaped the relative amplitude criterion will have produced artificial levels that could affect the dwell time statistics. The idealization can be further improved by using information about the number of current levels and their amplitudes. These parameters can be estimated by analyzing an amplitude histograms constructed from the idealized data or from the raw data (Fig. 5). Once the number of current levels is determined, the result of the idealization can be used in the INTERPRET phase of the TRANSIT analysis (Fig. 5, F and G).

INTERPRET assumes the existence of a closed level with zero amplitude (C) and allows one or more open levels

(O1–On) to be specified. INTERPRET compares the amplitude of each level with the defined amplitudes and assigns a number (0 for closed, 1 through  $n$  for O1–On) for the amplitude closest to that level. If the amplitude of a level exceeds the largest open amplitude (On), then more than one channel could be open simultaneously. In this case, the number of simultaneously open channels is estimated and assigned to the level as well. After all levels have been assigned a number, neighboring levels with identical numbers are concatenated: the dwell times are summed and the amplitudes averaged using dwell times as weights. The end result is a sequential list of levels, with four parameters for each level: the duration, the amplitude, the number indicating the conductance state (C, O1–On) and the number of simultaneously open channels. This data set may be used as input for the same statistical analysis as the original idealization. However, some of the previously mentioned problems arising from unresolved closures may have been remedied. Because every level now has a state number as well as the number of overlapping channels associated with it, new statistics become available. For instance, the probability of being in a particular conductance state as a function of time can be calculated. For channels with more than one open conductance state, a state transition diagram can be constructed that tabulates the probability of moving from each state to any other state.



**FIGURE 5** Idealization, level estimation and interpretation. (A) Diagram illustrating data flow during TRANSIT analysis. (B) Raw data, which are subjected to baseline analysis and idealization as described. (C) The result of idealization, which can be used for statistical analysis. Alternatively, the idealized data can be fed into the INTERPRET module, after the number of open levels and their amplitudes have been determined. Two different approaches are illustrated in *D* and *E*. (D) An all-point histogram obtained from the raw data was fitted with sums of 1–5 Gaussians. The sum of 4 Gaussians shown here minimized AIC. (E) An amplitude histogram from idealized levels and a kernel estimate of the pdf. Because the pdf is a continuous smooth curve, modi can be determined. (F and G) The result of two interpretations of the same idealization (C) assuming one open state of 10 pA (F) or three open states of 2, 8, and 10 pA (G).

## METHODS

### Simulations

Single channel behavior was simulated using continuous time, discrete state Markov models. The mean dwell time in any given state was calculated by inverting the sum of the rates leaving that state. During the simulation, the actual dwell time in a state is found by random drawing from an exponential distribution with a mean as calculated. This dwell time together with the associated current amplitude is stored. The channel then moves to a new state which is selected randomly, using the rate constants

associated with each move as weights. This process is repeated iteratively. The result is a consecutive series of duration-amplitude pairs. This time series is oversampled (8–12X), white noise is added, the record is filtered using a Gaussian filter (Colquhoun and Sigworth, 1983), and the result is resampled at the desired sample frequency.

### Maximum likelihood estimation

Parameters of dwell time distributions (time constants, areas) were estimated by a maximum likelihood procedure, using unbinned dwell

times. The likelihood was optimized using the variable metric Davidon-Fletcher-Powell method (Rao, 1984). The minimal number of exponential components required to describe the data was determined using the likelihood ratio test (Cox and Hinkley, 1974) at a  $p$  value of 0.05.

### Curve-fitting amplitude histograms

Amplitude histograms were fitted with a sum of Gaussians by minimizing the sum of squared differences (residuals) between the model and the data, using the variable metric method mentioned above. The number of Gaussian components was increased until a minimum in the asymptotic information criterion (AIC) was reached (Akaike, 1981).

## RESULTS

To test various aspects of TRANSIT, current records were simulated using continuous-time, discrete-state Markov models, as described in Methods.

### Noise sensitivity: rejection of spurious transitions

The role of the relative amplitude criterion in TRANSIT is to reject spurious transitions caused by noise. To evaluate

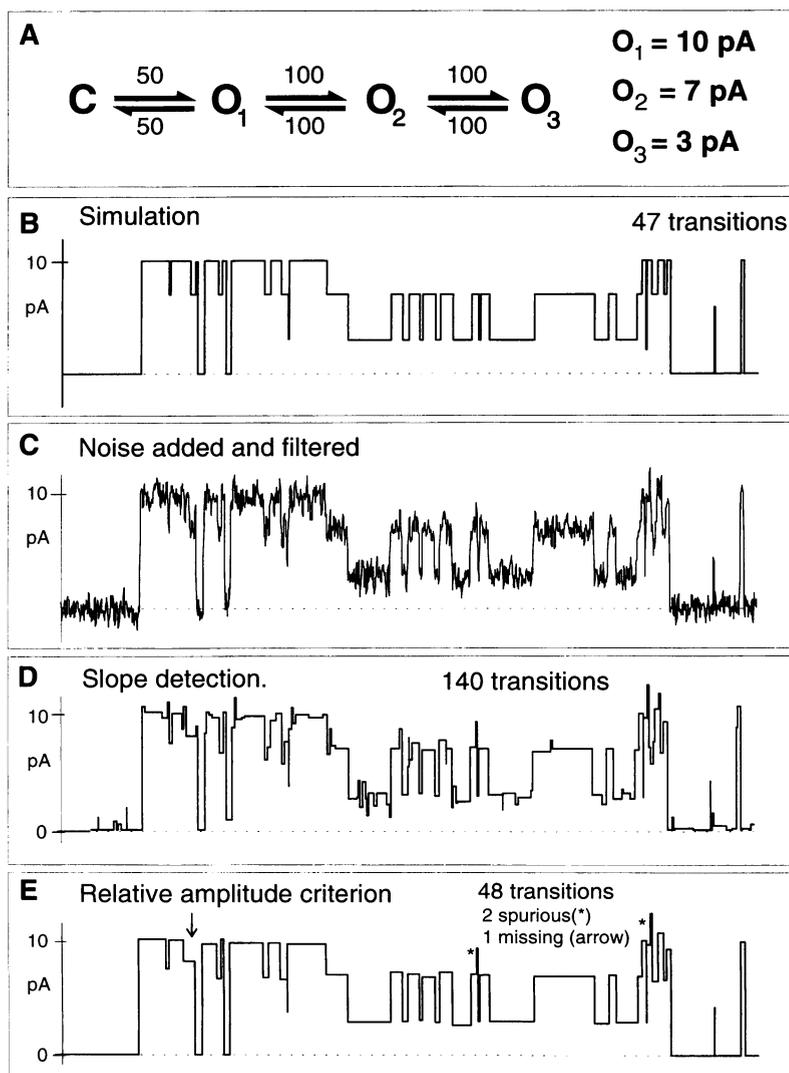
**TABLE 1** Probability  $P_{\text{spurious}}$  (%) as a function of threshold values

	$\sigma_{\text{slope}}$	$2\sigma_{\text{slope}}$	$3\sigma_{\text{slope}}$		$\sigma_{\text{slope}}$	$2\sigma_{\text{slope}}$	$3\sigma_{\text{slope}}$
$\sigma_{\text{amp}}$	4.7	0.61	<b>0.002</b>	$\sigma_{\text{amp}}$	18.7	1.91	<b>0.030</b>
$2\sigma_{\text{amp}}$	0.13	<b>0.065</b>	<b>0.001</b>	$2\sigma_{\text{amp}}$	2.95	0.98	<b>0.026</b>
$3\sigma_{\text{amp}}$	<b>0.010</b>	<b>0.006</b>	<b>0.001</b>	$3\sigma_{\text{amp}}$	0.12	<b>0.089</b>	<b>0.009</b>

(use  $4 * \sigma_{\text{amp}}$  if  $N1, N2 \leq 3$ )

Five data sets of 200 records containing 2048 points were constructed. The total number of data points per data set was 409,600. The filter frequency (5 kHz) was  $\frac{1}{4}$  the sample frequency (20 kHz). A DC offset was added to simulate an open channel, from which transition in both directions are possible.  $P_{\text{spurious}}$  is the probability (in %) that TRANSIT fails to reject a spurious transition.  $p$  values  $< 0.1\%$  are printed in bold: in those cases the probability for a spurious transition to occur is  $< 1$  per 1000 sample points.

how the two thresholds affect the probability of a spurious transition contaminating the idealization, current records containing only pink noise were simulated. Baseline analysis was performed to obtain the required standard deviations. Thresholds were set at 1, 2, and 3 times the standard



**FIGURE 6** The importance of the relative amplitude criterion. (A) Single channel behavior was simulated using the linear Markov model shown. (B) Example of a noise-free, unfiltered current trace simulated from the model. The record length is 400 ms (2048 points sampled at 5.12 kHz). There are 47 transitions between 48 levels. (C) The simulated time series was oversampled 10 times (51.2 kHz), and white noise was added; the data were low-pass filtered at 1 kHz and resampled at 5.12 kHz. This resulted in baseline noise with a standard deviation of 1.0 pA. S/N ratios are therefore 3, 4, and 10 for the three jump heights. (D) Slope detection produced 140 transitions, most of which are spurious. The idealization at this point is shown. (E) The relative amplitude criterion removes all but two spurious transitions (\*), while one transition (arrow) is missing.

deviation for both amplitude and slope criterion, giving a total of nine combinations. Since these records contain only noise, all transitions are spurious. Table 1 tabulates the probability that fluctuations in the baseline will result in a spurious transition, for the nine different threshold settings. The effect of using a higher threshold of  $4\sigma_{\text{amp}}$  for transitions involving short durations was also evaluated. Table 1 shows that if the slope threshold is at least  $2\sigma_{\text{slope}}$  and the amplitude threshold is at least  $3\sigma_{\text{amp}}$ , then the probability of a noise fluctuation giving rise to a spurious transitions is  $<1$  in 1000 sample points ( $P_{\text{spurious}} < 0.1\%$ ). This probability goes down 10-fold by using the more stringent amplitude test for short durations.

### Detection of real transitions: importance of the relative amplitude criterion

The ability of TRANSIT to detect transitions between multiple current levels in the presence of noise was tested next. A four-state Markov model with three open states (Fig. 6 A) was used to generate the same single channel data with and without noise. A typical noise-free record is shown in Fig. 6 B, which contains 48 transitions, most of which are 3 or 4

pA jumps. After addition of noise and filtering (Fig. 6 C), the record was idealized. Fig. 6 D shows the result of the slope detection phase of the algorithm; 140 transitions were detected, the majority of which are spurious. The relative amplitude criterion removes all but two spurious transitions (Fig. 6 E\*), while 1 real transition (Fig. 6 E, arrow) is missing.

### Comparison of TRANSIT with the 50% amplitude threshold detector

The above analysis suggests that the TRANSIT algorithm can produce a reasonable idealization from data for which the signal-to-noise (S/N) ratio is as low as 3 or 4. The 50% amplitude threshold detector does not work reliably under such conditions. This point is illustrated in Fig. 7 using a two-state (closed-open) model. The single channel amplitude is 3.5 pA, while the standard deviation of the noise (after filtering) is 1.0 pA, yielding an S/N ratio of 3.5 (Fig. 7 A). Under these extremely noisy conditions, TRANSIT is still able to extract the underlying single channel behavior (Fig. 7, B and E), whereas the 50% amplitude threshold

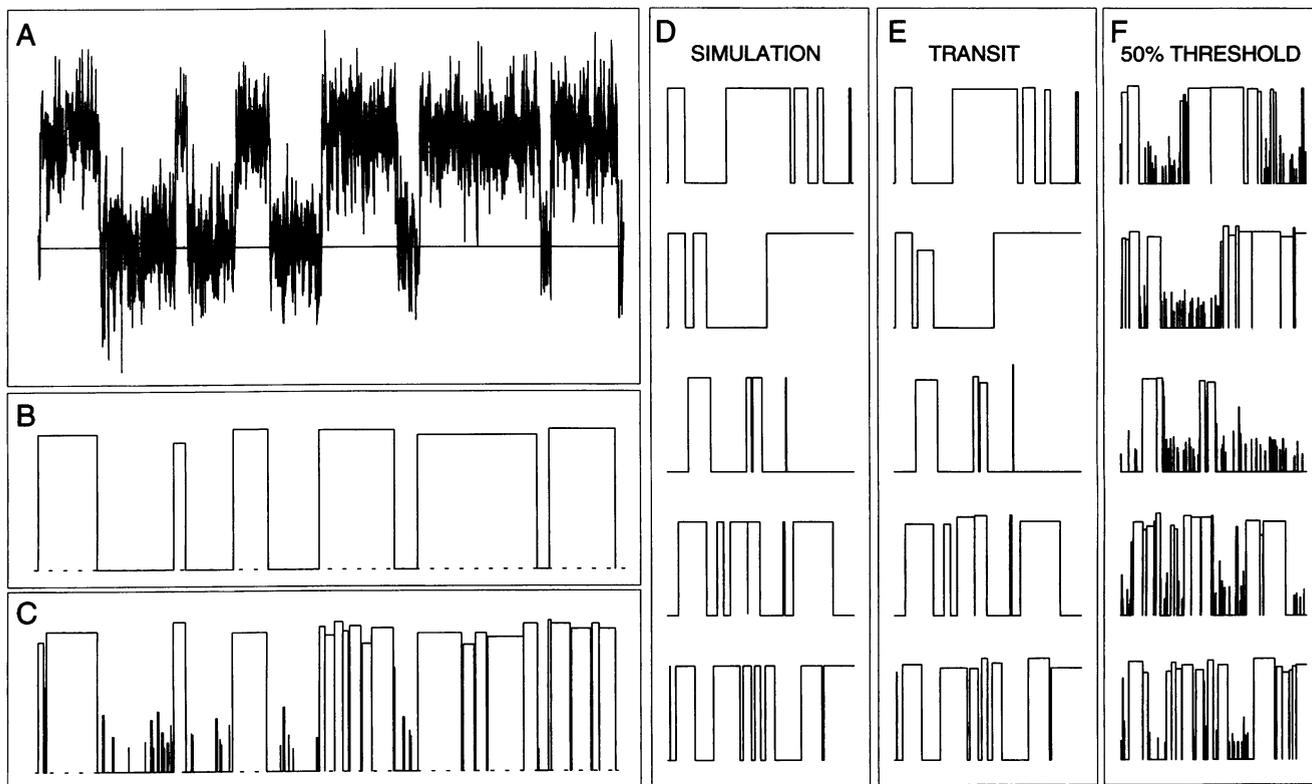


FIGURE 7 Comparison of TRANSIT with 50% amplitude threshold algorithm. (A) Single channel data were simulated using a two-state (C-O) model with both rate constants being  $100 \text{ s}^{-1}$ . The open state had an amplitude of 3.5 pA. Records were 100 ms long (2048 points sampled at 20.48 kHz) and filtered at 5.12 kHz. The standard deviation of the baseline noise was 1 pA resulting an S/N ratio of 3.5. (B) Idealization by TRANSIT of the record shown in A. The relative amplitude threshold was raised to  $4\sigma_{\text{amp}}$  for transitions involving levels with a duration  $\leq 3$  sample points. (C) Idealization of record A by the 50% amplitude threshold algorithm. Noise fluctuations cause many spurious transitions both from the baseline and the open state. (D-F) Idealization of five more traces with the same amount of noise superimposed (raw traces not shown) by TRANSIT (E) and the 50% amplitude threshold algorithm (F). The noise-free simulations are shown in D for comparison.

detector produces a large number of spurious transitions (Fig. 7, *C* and *F*), as expected.

The TRANSIT algorithm has also been tested on more realistic models containing multiple closed states, some of which are very brief. At a noise level where the 50% threshold detector works reliably ( $S/N = 6$ ), TRANSIT correctly idealizes all resolved levels, but also detects many unresolved closures (flickers) that do not reach the 50% amplitude threshold (data not shown).

### TRANSIT performance and noise

A desirable property of an idealization algorithm is that in the absence of noise and filtering, the idealization is perfect. This was tested using a six-state Markov model, which produced complex single channel records (Fig. 8). Idealization by TRANSIT correctly detected all transitions and accurately estimated the dwell times and amplitudes of all levels. Degradation of the performance resulted from increasing amounts of noise and filtering (Fig. 9 and 10). The idealizations can be seen to deteriorate progressively; increased noise results in an increase in the number of spurious transitions. Increased filtering results in short levels being lost.

### Estimating the number of current levels and their amplitudes

Before a statistical analysis of individual (sub)conductance levels can be undertaken, INTERPRET needs to assign state numbers to all levels. A necessary step is then the estimation of the number of current levels and their amplitudes. This could be done from the raw data after construction of an

all-points histogram, in which points with a large slope are excluded (Tyerman and Findlay, 1992). Alternatively, an amplitude histogram can be constructed from the idealized levels. The two approaches are compared in Fig. 11 using the six-state model shown in Fig. 8 *A*. The  $S/N$  ratio varied between 2 and 9 for the various jump heights. All six current levels are correctly extracted with both methods. Since amplitudes of idealized levels are obtained by averaging, their standard deviation is smaller than that of an individual point. Consequently, the Gaussian components in an idealized amplitude histogram are usually much narrower than those in an all-points histogram (compare Fig. 11, *C* and *D*), which can help separation of peaks that are close. Also, amplitudes of levels with a relatively short duration, which can get lost in an all-points histogram, are better represented in the idealized histogram since every level contributes one observation, irrespective of its duration.

### Subconductance analysis

After estimation of the number of levels and their amplitudes, INTERPRET can assign state numbers to levels, as described above. This information can then be used to construct dwell time histograms for individual open states, as well as state transition diagrams. This is illustrated in Fig. 12 with the use of a five-state circular Markov model with only forward rate constants. The amplitudes and mean dwell times of the five states are listed in Fig. 12 *A*. The shortest mean open time was 0.5 ms, while the time between samples was 0.05 ms. Fig. 12 *B* illustrates a typical idealization of data produced by the model. Amplitude histograms of both raw and idealized data showed well defined peaks at 1, 2, 3, and 4 pA (not shown). INTERPRET was used to assign

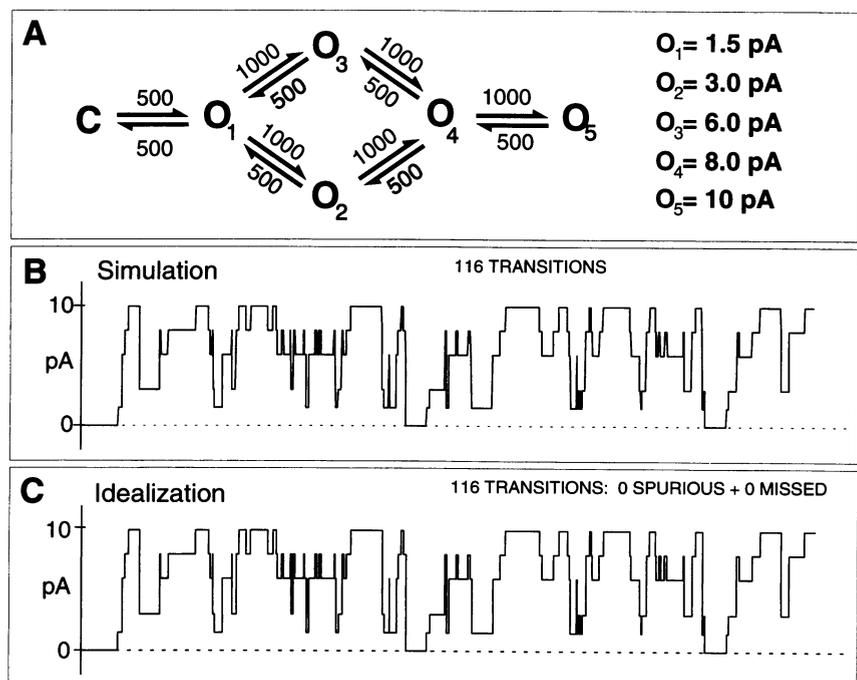
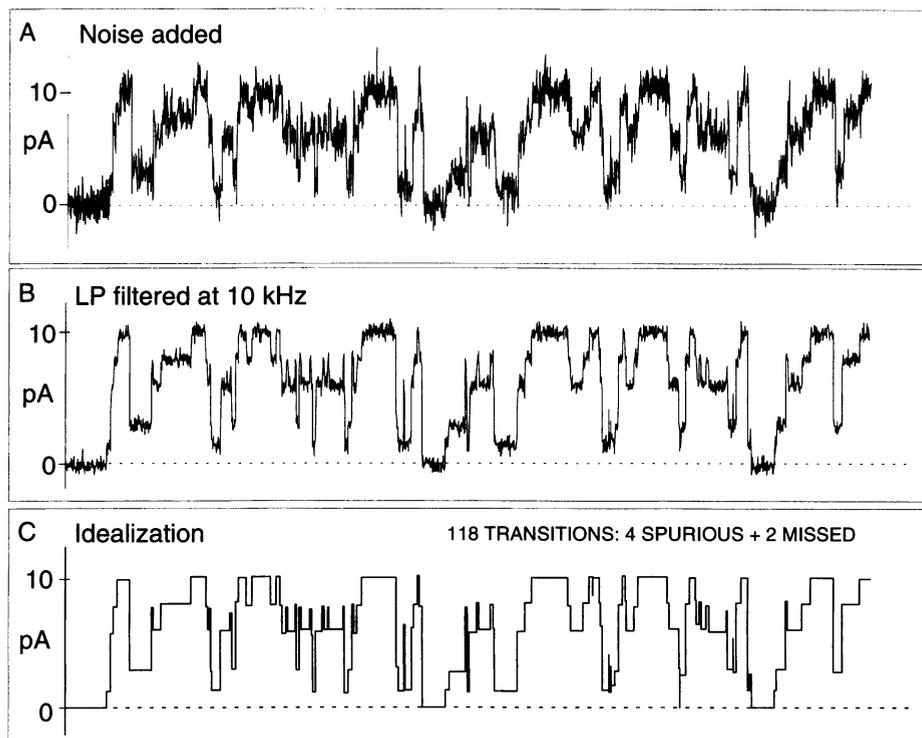


FIGURE 8 Idealization by TRANSIT is perfect in the absence of noise. (A) Six-state Markov model, containing five open states with amplitudes 1.5, 3, 6, 8, and 10 pA and jump heights of 1.5, 2, 3, and 4.5 pA. (B) Noise-free simulation of the model in A. The record is 50.12 ms long (2048 points sampled at 40 kHz) and contains 116 transitions. (C) Idealization of record in B. All 116 transitions are detected and all 117 levels correctly idealized.

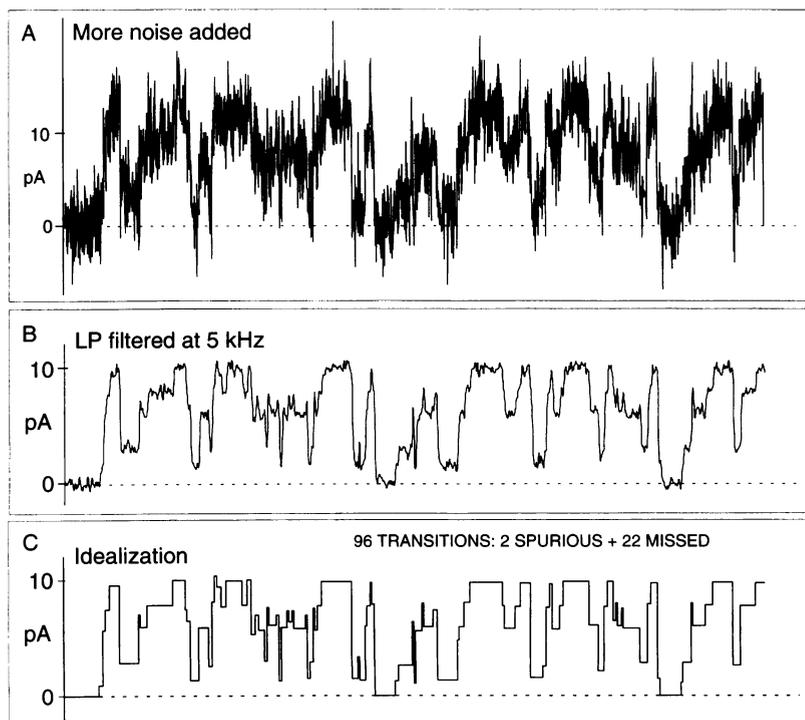
**FIGURE 9** Introduction of noise results in spurious and missed transitions. (A) Same record as in Fig. 8 B, after addition of noise (1 pA). (B) Low-pass filtering (10 kHz) was necessary to increase the S/N ratio. The standard deviation of the (filtered) baseline noise was 0.25 pA. (C) Idealization of record in B: 114 transitions are recovered (out of 116), and spurious transitions resulted in the introduction of two spurious levels.



conductance states (0–4) to the levels for 100 idealized records. The idealization produced 2632 levels, which INTERPRET reduced to 2623. Dwell time distributions were single exponential as shown in Fig. 12, C and E–H. In addition, a Markov state transition diagram was constructed, which tabulates the probabilities of going from each state to

any other state (Fig. 12 D). The circular connectivity and the absence of reverse rate constants can be extracted from the transition diagram, although missed transitions from O3 to O4 due to the short lifetime of O4 suggest a connection between O3 and C, which doesn't exist in the original model.

**FIGURE 10** Increased noise further degrades performance. (A) Same record as Fig. 8 B, after addition of more noise. (B) More severe low-pass filtering was required (5 kHz) to reduce the standard deviation of the baseline noise to 0.36 pA. (C) Idealization recovers 96 transitions (out of 116) and introduces 2 spurious ones.



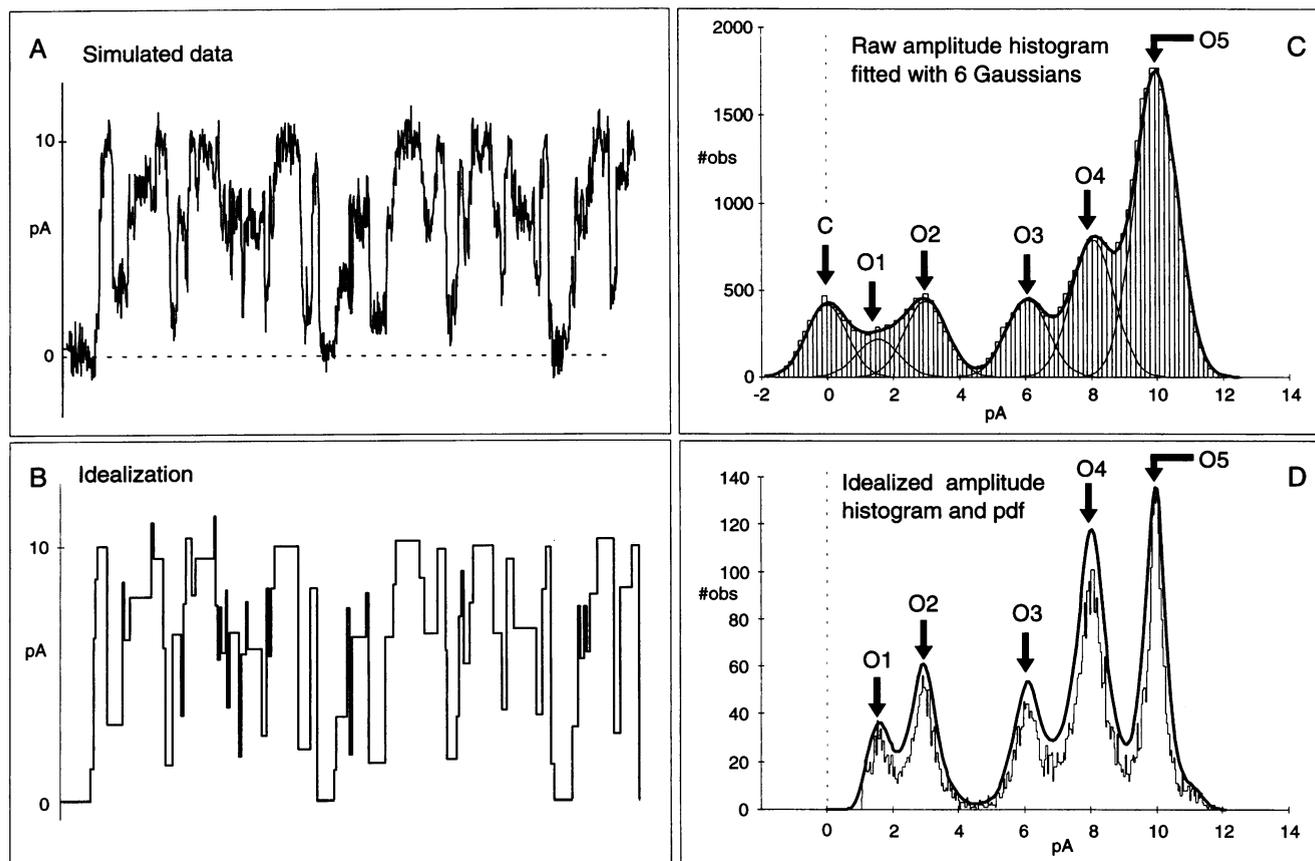


FIGURE 11 Estimation of the number of levels and their amplitudes. (A) Simulated current trace using the six-state model in Fig. 9 A. There are five open states with amplitudes 1.5, 3, 6, 8, and 10 pA. Jump heights are 1.5, 2, 3, and 4.5 pA. The noise standard deviation after filtering is 0.75 pA, resulting in S/N ratios between 2 and 6. (B) Idealization of the record in A. The relative amplitude threshold was raised to  $4\sigma_{\text{amp}}$  for transitions involving levels with a duration  $\leq 3$  sample points. (C) All-points amplitude histogram constructed from 20 records similar to record A. Sample points with slopes exceeding the threshold were discarded to prevent transitions from contaminating the histogram. Sums of Gaussians were fitted to the histogram, and six Gaussians were found to minimize AIC. Their means were 0.06, 1.54, 3.00, 6.06, 8.05, and 9.97 pA. (E) An amplitude histogram was constructed from the idealized levels using 100 records. Levels with durations shorter than 3 sample points were excluded, since their amplitude is unreliable. Every level contributed 1 observation to the histogram, irrespective of its duration. The smooth bold curve shows the pdf obtained by a kernel estimator, which it is 6 times more efficient and produces a smooth curve. The modi were 1.59, 2.95, 8.04, and 9.94 pA.

## DISCUSSION

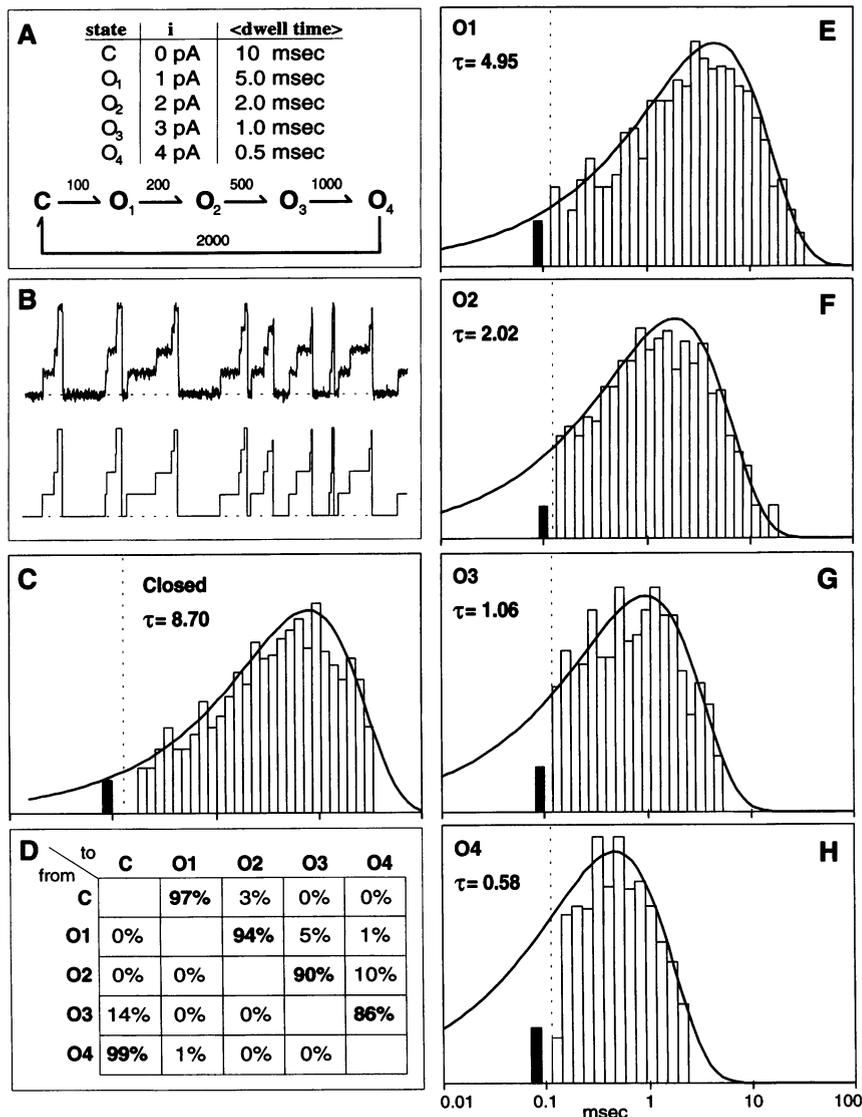
The objective of this paper was to design a single channel idealization algorithm that would have the desirable properties mentioned in the introduction. The first two are clearly met: no a priori information about the number of levels or their amplitudes is required by TRANSIT and no underlying model has to be assumed. The only assumption that TRANSIT makes is that transitions between current levels are fast.

A level only becomes part of the idealization if both transitions flanking the level 1) exceed the slope threshold, and 2) survive the relative amplitude test. Consequently, a level whose amplitude differs less than  $3\sigma_{\text{amp}}$  from the level preceding or following it will not be properly idealized. Therefore, TRANSIT should not be used on data with S/N ratio  $< 3$ . TRANSIT has no intrinsic minimum duration requirement; levels as short as a single sample interval will be reliably idealized, as long as they are resolved (i.e., they

meet the detection criteria). However, short levels will be missed by TRANSIT if their amplitude is sufficiently degraded by low-pass filtering. In that case, the current deflections caused by the transitions flanking the level have become indistinguishable from noise fluctuations.

Objectivity was maximized by basing the two important thresholds on standard deviations obtained from the baseline noise. The optional INTERPRET phase is not objective, however, since it requires the user to decide on the number of levels and their amplitudes. This subjectivity is inevitable when one wants to calculate statistics of subconductance levels. Finally, it was illustrated that in the absence of noise and filtering, idealization by the TRANSIT algorithm is perfect even for complex single channel behavior. Gradual degradation of performance resulted from increasing amounts of noise and filtering.

TRANSIT differs from previously published approaches for analyzing data containing subconductance levels. Pat-



**FIGURE 12** Subconductance analysis. (A) A five-state circular Markov model with four open states (1, 2, 3, and 4 pA) and only forward rates is shown. The mean dwell times are tabulated for all states. (B) Example of a simulated current record and the corresponding idealization. The standard deviation of the baseline noise was 0.12 pA, resulting in S/N ratios of 8.3 and 33 for the 1 and 4 pA jumps, respectively. The amplitude pdf showed well resolved peaks at 1, 2, 3, and 4 pA (not shown). These amplitudes were used by INTERPRET to assign conductance state numbers to levels for 100 idealized records. Dwell time distributions for the five states (C, O1–4) were best described by a single exponential, using a maximum likelihood analysis. The results are shown in C and E–H using a log transform (Sigworth and Sine, 1987). Time constants for the exponential functions are shown in each panel. (D) A Markov state transition diagram was constructed from the interpreted data. Probabilities of going from each state to any other state are tabulated.

lak's mean-variance analysis (Patlak, 1988, 1993) is useful for estimating the number of conductance levels, their amplitudes and kinetic properties, but it is not an idealization algorithm, inasmuch as it does not generate a sequential list of amplitudes and durations. The sublevel Hinkley detector (Draber and Schultze, 1994) requires the number and amplitudes of all current levels and has an intrinsic minimal duration requirement of six samples or even more for sub-conductance levels. The hidden Markov analysis (Chung et al., 1990, 1991) circumvents the problem of idealization completely by directly estimating Markov model parameters from the raw data. This method should be used whenever the objective is to estimate the parameters of a Markov model.

Currently, TRANSIT employs a fixed relative amplitude threshold. The optional increase in stringency for transitions involving short (1–3 sample point) durations by raising the amplitude threshold is a first step toward making the algorithm perform at a fixed probability of accepting a spurious

transition. Additional work will be needed to further investigate this point.

In conclusion, a new algorithm has been described that can reliably idealize single channel data, containing any number of unknown conductance levels. The idealization can be used to estimate the number of levels and their amplitudes. Using this information, conductance states can be assigned to current levels, and statistical properties of individual (sub)conductance levels can be evaluated.

The author would like to thank Arthur M. Brown in whose lab TRANSIT was conceived. This work was supported by a grant from the National Institute of Neurological Disorders and Stroke (NS-31557 to A. M. J. V. D.).

## REFERENCES

- Akaike, H. 1981. Modern development of statistical methods. *In* Trends and Progress in System Identification. IFAC Series for Graduates, Re-

- search Workers and Engineers, Vol. 1. P. Eykhoff, editor. Pergamon, Oxford. 169–184.
- Auerbach, A., and F. Sachs. 1983. Flickering of a nicotinic ion channel to a subconductance state. *Biophys. J.* 42:1–10.
- 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. *Philos. Trans. R. Soc. Lond. Biol.* 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. *Philos. Trans. R. Soc. Lond. Biol.* 329:265–285.
- Colquhoun, D., and F. J. Sigworth. 1983. Fitting and statistical analysis of single-channel records. In *Single-Channel Recording*. E. Neher and B. Sakmann, editors. Plenum Press, New York. 191–263.
- Cox, D. R., and D. V. Hinkley. 1974. *Theoretical Statistics*. Chapman and Hall Ltd., London.
- Draber, S., and R. Schultze. 1994. Detection of jumps in single-channel data containing subconductance levels. *Biophys. J.* 67:1404–1413.
- Ferguson, W. B., O. B. McManus, and K. L. Magleby. 1993. Opening and closing transitions for BK channels often occur in two steps via sojourns through a brief lifetime subconductance state. *Biophys. J.* 65:702–714.
- Geletyuk, V. I., and V. N. Kazachenko. 1985. Single Cl channels in molluscan neurones: multiplicity of conductance states. *J. Membr. Biol.* 86:9–15.
- Hamill, O. P., J. Bormann, and B. Sakmann. 1983. Activation of multiple-conductance state chloride channels in spinal neurones by glycine and GABA. *Nature*. 305:805–808.
- Hamill, O. P., A. Marty, E. Neher, B. Sakmann, and F. J. Sigworth. 1981. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pfluegers Arch.* 391:85–100.
- Matsuda, H. 1988. Open-state substructure of inwardly rectifying potassium channels revealed by magnesium block in guinea-pig heart cells. *J. Physiol.* 397:237–258.
- Nagy, K. 1987. Subconductance states of single sodium channels modified by chloramine-T and sea anemone toxin in neuroblastoma cells. *Eur. Biophys. J.* 15:129–132.
- Neher, E., and B. Sakmann. 1976. Single-channel currents recorded from membrane of denervated frog muscle fibers. *Nature*. 260:799–802.
- Neher, E., and C. F. Stevens. 1977. Conductance fluctuations and ionic pores in membranes. *Annu. Rev. Biophys. Bioeng.* 6:345–381.
- Nilius, B., J. Vereecke, and E. Carmeliet. 1989. Different conductance states of the bursting Na channel in guinea-pig ventricular myocytes. *Pfluegers Arch.* 413:242–248.
- Patlak, J. B. 1988. Sodium channel subconductance levels measured with a new variance-mean analysis. *J. Gen. Physiol.* 92:413–430.
- Patlak, J. B. 1993. Measuring kinetics of complex single ion channel data using mean-variance histograms. *Biophys. J.* 65:29–42.
- Rao, S. S. 1984. *Optimization: theory and applications*. Wiley Eastern Limited, New Delhi.
- Root, M. J., and R. Mackinnon. 1994. Two identical noninteracting sites in an ion channel revealed by proton transfer. *Science*. 265:1852–1856.
- Sachs, F. 1983. Automated analysis of single-channel records. In *Single-Channel Recording*. E. Neher and B. Sakmann, editors. Plenum Press, New York. 265–285.
- Sachs, F., J. Neil, and N. Barkakati. 1982. The automated analysis of data from single ionic channels. *Pfluegers Archiv. Eur. J. Physiol.* 395:331–340.
- Savitsky, A., and M. J. E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36:1627–1639.
- Sigworth, F. J. 1983. An example of analysis. In *Single-Channel Recording*. B. Sakmann and E. Neher, editors. Plenum Press, New York. 301–321.
- Sigworth, F. J., and S. M. Sine. 1987. Data transformations for improved display and fitting of single channel dwell-time histograms. *Biophys. J.* 52:1047–1054.
- Silverman, B. W. 1986. *Density estimation for statistics and data analysis*. Chapman and Hall, New York.
- Steinler, J., Y. Termonia, and J. Deltour. 1972. Comments on smoothing and differentiating data by simplified least square procedure. *Anal. Chem.* 44:1906–1909.
- Tyerman, S. D., and G. P. Findlay. 1992. Multiple conductances in the large K channel from *Chara corallina* shown by a transient analysis method. *Biophys. J.* 61:736–749.
- VanDongen, A. M. J. 1992. TRANSIT: a new algorithm for analyzing single channel data containing multiple conductance levels. *Biophys. J.* 61:A256.