

Chapter 6

Practical analysis of single channel records

DAVID COLQUHOUN

1. Introduction

The aim of this chapter is to discuss the sorts of things that can be measured in an experimental record of single channel currents, and how to make the measurements. What is measured will depend on the aims of the experiment. In some cases, interest may centre mainly on the *amplitudes* of the currents. For example, this might be the case (*a*) when channel conductance and subconductance levels are used as a criterion for a particular channel subtype, or (*b*) when measurements are made in solutions of different ionic composition and different membrane potentials, in order to investigate the mechanism of ion permeation through open channels. In other cases the *durations* of the open and shut times may be of primary interest, as when we want to know the nature of individual channel *activations* by a transmitter (the unitary event usually consists of more than one opening), or when the kinetic mechanism of channel operation is of primary interest.

The aims of a complete analysis are to measure (*a*) the amplitude(s) of the single channel currents, (*b*) the durations of shut periods, and the durations of sojourns at the various open channel current levels, and (*c*) the order in which the foregoing events occur. The amplitudes are, in the simplest cases at least, very nearly constant from one opening to the next. But, because we are looking at a single molecule, the durations of events and the order in which they occur are *random variables*; the information contained in them comes from measurements of their *distributions* (more strictly, their probability density functions - see Chapter 7 for more details).

Many measurements of individual durations have to be made in order to define these distributions properly. Single channel analysis must be one of the slowest known methods of generating an exponential curve from an experiment, because the averaging that normally results from having a large number of channels has not 'already been done for us'. Furthermore the analysis is particularly important; not much can be inferred by simply looking at the raw data, *because* of its randomness. These measurements can be rather time consuming; this leads to the temptation to use automatic or semiautomatic methods of analysis in which distributions are produced by a computer program with little intervention by the experimenter. It is not usual in

other fields for results to be produced without the data having been seen by the experimenter (though it is not entirely unknown). Personally I would be reluctant to accept for publication any paper in which the fit of each duration to the raw data had not been inspected by the experimenter, even if the experimenter had written the analysis program him or herself (and still more reluctant if the analyst was not even sure exactly what the program was doing to their data). Certainly it is important that reasonably complete details of analysis methods should be given in published papers so that the reader can make some sort of assessment of its reliability. There is a price for speed of analysis which is too high to be tolerable, and failing to look at your data exceeds that price, in my view. It would help if the writers of programs would not include options that allow this to be done.

2. Aims of analysis

It is usually supposed (with some reason) that the data can be well-represented as a series of rectangular transitions between discrete conductance states. The first aim of the analysis is to obtain an idealised version of the experimental record which resembles, as closely as possible, what would have been seen if the experiment had been free of noise and artefacts. This process is sometimes referred to as *restoration* of the observed record. The result will be a series of time intervals, each associated with the amplitude of the current during that interval, in the order in which the events actually occurred. Open and shut periods will not necessarily alternate; successive open intervals may occur if there are conductance sublevels, or if more than one channel opens simultaneously. Records in which two or more channels are open simultaneously are useful for checking the independence of channel openings (though tests of independence are insensitive - see Horn, 1991), but are *not* suitable for measuring lifetime distributions. A double opening may be omitted from the analysis of lifetimes by ignoring it (and simultaneously noting that the shut time between the preceding and following openings is to be ignored when forming the distribution of shut times); this procedure will bias both open and shut time distributions so it can be used only when double openings are rare.

The results of the analysis will not of course be entirely accurate. For example, if an opening is too short to be detected not only will its omission distort the open time distribution, but the shut periods on each side of it will appear to be one long shut period so the shut time distribution will be distorted too (and *vice versa* for missed shut times). This problem is discussed further below (section 10) and in Chapter 7.

Techniques are being developed that may allow direct fit of a mechanism to the original data without prior restoration (e.g. Chung *et al.* 1990; Fredkin & Rice, 1992), but these have not yet been developed to a point where they are useful in practice.

Computer programs for analysis

There are several commercial programs for single channel analysis. These are discussed briefly in Chapter 9. All of them can perform *some* of the types of analysis

discussed below, but none of them can do *all* of these methods (or other types of analysis that are not discussed in this chapter). If you want to use the more advanced methods of analysis, or to develop new methods, then you have only two options: (1) write your own program (or modify an existing one if you can get hold of the source code), or (2) get a program from somebody who has already written it.

3. Filtering and digitization of the data

Whether data are recorded on magnetic tape during the experiment, or recorded on-line (see below), it will almost always be necessary to filter the data before any analysis is done.

Filtering the data

The purpose of filtering is essentially to reduce the amount of high frequency noise in the record, so a 'low-pass' filter should be used. This passes frequencies from 0 Hz (DC) up to an upper limit specified by cut-off frequency, f_c , set on the filter. Butterworth filters pass relatively little noise with frequencies above f_c - they have a 'steep roll-off', which is why they are used to prevent aliasing in noise analysis. However, they are quite unsuitable for filtering single channel data because the price paid for the steep roll-off is that they 'ring', i.e. produce a damped oscillation, in response to a rectangular input. Single channel currents are essentially rectangular, and will therefore be distorted by such a filter. The type of filter that is normally used is a Bessel filter (usually an 8-pole Bessel filter). The types labelled 'damped mode', or 'low Q', on some commercial filters are similar. This sort of filter produces little or no ringing in response to a step input, though it is less effective in removing frequencies above f_c . It should be noted that some commercial filters of the Bessel type are calibrated on the front panel with a number (the corner-frequency) that is twice the -3 dB frequency; it is preferable that the values of f_c stated in papers should always be the -3 dB frequency; see Chapter 16). The question of the optimum setting for f_c is discussed below.

Filter risetime. Another convenient way to characterize the filter is by its rise time, t_r . This is given by

$$t_r = 0.3321/f_c, \quad (3.1)$$

so the higher the -3 dB frequency, the faster is the risetime. This expression, which is close to the 10-90% risetime, is actually derived for a type of filter known as a Gaussian filter, which behaves very like the 8-pole Bessel filter that is normally used in practice (see Colquhoun & Sigworth, 1983). Thus a 1 kHz filter has a risetime of 332 μ s, and *pro rata* for other values of f_c .

Combining filters. The relevant value of f_c in the discussions below is not that for the filter alone, but the effective value for the whole recording system. If, for example, the results were effectively filtered at $f_1 = 10$ kHz by the recording system,

and were then filtered again at $f_2 = 5$ kHz, then the effective overall filtering would be given, approximately, by

$$\frac{1}{f_c^2} \approx \frac{1}{f_1^2} + \frac{1}{f_2^2}, \quad (3.2)$$

i.e. at $f_c \approx 4.47$ kHz in the present example. This expression is exact for a Gaussian filter and a good approximation for an 8-pole Bessel filter, but it may not give good results for steep roll-off filters such as Butterworth or Tchebychef type (usually used, for example, in tape recorders).

Digitization of the data

All practical forms of analysis have to be done on a computer, so the first stage is to convert the observed current into a series of numbers by means of an analogue-to-digital converter (ADC). This is usually done after filtering through an analogue filter; another possibility is to use a digital filter *after* sampling (though some pre-filtering may be needed to prevent saturation of the ADC by high frequency noise).

Agonist-activated channels will generally give rise to long records which can, most conveniently, be recorded on magnetic tape during the experiment, and then replayed later for digitization. The algorithm used for digitization of records should be capable of writing the numbers directly to magnetic disk as it goes, in order to avoid frequent breaks in the record. The sampling rate should be 10-20 times the -3 dB frequency of the filter in use (though a factor of 5 is sufficient if the resulting points are then supplemented by interpolation).

For example, for a record filtered with a Bessel-type filter with $f_c = 5$ kHz (-3 dB), a suitable sampling rate would be 50 kHz. Samples are normally stored as 16 bit (2 byte) integers, so this corresponds to 6 megabytes of data per minute. A computer with plenty of hard disk space is needed. Some routines (e.g. that supplied by Cambridge Electronic Design) can sample continuously to disk at up to 80 kHz, but others are limited to about 30 kHz. Further details can be found in Colquhoun & Sigworth (1983).

Sampling on-line

When the opening of channels is caused by a step change in membrane potential or agonist concentration it will usually be convenient to do the experiment on-line. The computer will supply the command signal, through a digital-to-analogue converter (DAC) output, to change the membrane potential or concentration, and then immediately sample the resulting current through an ADC input. This procedure avoids problems in defining the exact moment at which the command step was applied. The sample length will usually be much shorter than is needed for steady-state (e.g. agonist-activated) channel records. For example, a 300 ms depolarization might be applied every 10 s, or a 1 ms concentration jump might be applied and the resulting channel activity record for 1 s subsequently. In such cases the sampled values can easily be accommodated in the computer memory, and there will be no

need to have a sampling routine that can write the numbers to a magnetic disc while sampling is in progress. In order to run DAC outputs and ADC inputs simultaneously, the computer routines should be interrupt-driven (see Chapter 9); the ability of different software/hardware combinations to do this varies greatly, but the latest version of the CED1401 interface (Cambridge Electronic Design) can do ADC sampling at 330 kHz simultaneously with DAC output at 105 kHz.

The length of the sample should be many times longer than the longest time constant that is to be investigated, so long samples will be needed for slow processes. If the sample is too short then the channel will appear sometimes to 'switch modes' between one sample and the next. Furthermore it is often forgotten in such experiments that it is not only the transient that follows the jump that is of interest, but also the equilibrium channel behaviour that is eventually attained. This is another reason for not making the sample too short. Preliminary analyses may be needed to determine how long the sample should be, and what recovery period is necessary between one pulse and the next.

If capacitative transients caused by the voltage step cannot be adequately compensated during the experiment, it may be desirable to do the compensation during the analysis. For example responses to pulses that happen to produce no channel openings can be averaged, and this average subtracted from each channel-containing record.

4. The measurement of amplitudes

As mentioned in the introduction, knowledge of channel amplitudes may be wanted as part of a complete analysis, or for particular purposes such as when measurements are made in solutions of different ionic composition and different membrane potentials, in order to investigate the mechanism of ion permeation through single open channels. The latter sort of study has been greatly facilitated by the ability to measure single channel currents. Before this was possible, such information had to be inferred indirectly from *instantaneous current-voltage relationships* in macroscopic experiments. In the latter, currents are measured as soon as possible ('instantaneously') after a step change in membrane potential, so that changes in the macroscopic current that resulted from changes in channel permeation could be distinguished from the (generally slower) changes in macroscopic current that result from (potential-dependent) changes in the *number* of channels.

The amplitude of channel openings may appear (incorrectly) to be reduced by molecules that block the open channel, if the blockages are frequent and too short to be resolved (e.g., Ogden & Colquhoun, 1985).

The problems that arise in the measurement of single channel current amplitudes are as follows.

Attenuation of brief events

A brief event (opening or shutting) will, because of the filtering of the record, produce

a response that does not reach full amplitude. This is illustrated in Figs 3 and 5-8. The fraction of the maximum amplitude that is attained (A_{\max}/A_0) can be calculated as

$$A_{\max}/A_0 = \text{erf}(2.668 f_c w) = \text{erf}(0.8860 w/t_r) \quad (4.1)$$

where $\text{erf}(\)$ is the *error function*, f_c is the -3 dB frequency of the filter recording system, t_r is the risetime of the filter (see (3.1) above), and w is the duration of the event. The error function, which is a function that is closely related to the cumulative Gaussian distribution, can be obtained from Tables (e.g. Abramowitz & Stegun, 1965), or calculated by a computer subroutine (one is given by Colquhoun & Sigworth, 1983), and all mathematical subroutine libraries contain the error function (e.g. NAG library; Press *et al.* 1986). The function in (4.1) is plotted in Fig. 3; it is correct for a Gaussian filter, and is a good approximation for the eight pole Bessel-type filter that is most commonly used in practice (Colquhoun & Sigworth, 1983). It must be emphasised again that some commercial filters show a value of $2f_c$ on the front panel; the manual must be checked to find the correct -3 dB frequency (or the rise time determined empirically). Table 1 shows the duration of events (openings or shuttings) that are required to reach 25%, 50% and 90% of maximum amplitude (which correspond, respectively, to $0.25 t_r$, $0.54 t_r$ and $1.3 t_r$).

Amplitudes of channel openings can be measured only for openings or shuttings that are long enough clearly to reach full amplitude (i.e. those with a length that is at least twice the rise time of the filter - see Fig. 3). There are two main approaches to

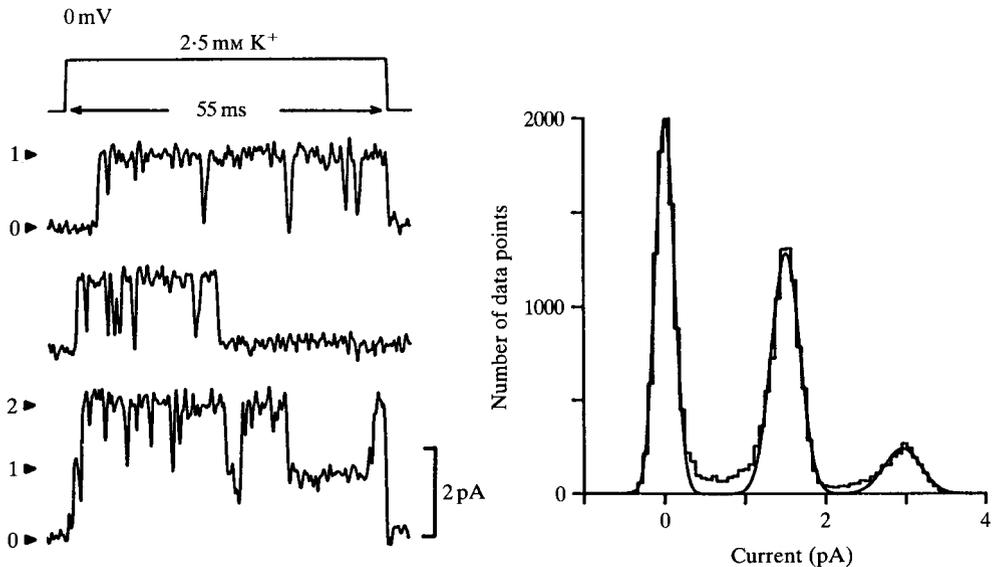


Fig. 1. Single channel records of potassium channels - delayed rectifier channels of skeletal muscle fibres - activated by a 55 msec depolarization from -100 mV to 0 mV. Two levels of opening are seen in the lowest record. On the right is shown a histogram of data points, showing peaks for the baseline and one and two levels of opening. Reproduced with permission from Standen *et al.* (1985).

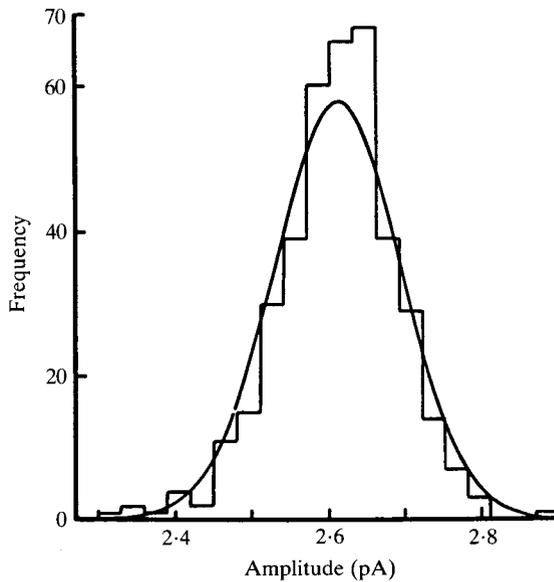


Fig. 2. An example of the distribution of the amplitude of single-channel currents. (Unpublished data of B. Sakmann & D. Colquhoun; *R. temporaria* end plate, $E_m = -91$ mV, suberyldicholine, 100 nM.) The mean of the 395 amplitudes was 2.61 pA. The continuous curve is a Gaussian distribution, which was fitted to the data by the method of maximum likelihood; it has a mean of 2.61 pA and a standard deviation of 0.08 pA. Note, however, that the observed distribution is rather more sharply peaked than the Gaussian curve. Reproduced with permission from Colquhoun & Sigworth (1983).

measurement of amplitudes, point amplitude histograms and amplitudes measured from each opening.

The point-amplitude histogram (distribution of data points)

The distribution of all the digitized current values can be plotted as a histogram. There will be a peak at the shut level and at each of the open levels. An example is shown in Fig. 1. The area under each peak is proportional to the *time spent* at that level.

This method has the advantage that the raw data are used without intervening processing, so the effects of the prejudices of the operator are minimized. It can also be calculated very rapidly. On the other hand, it has the disadvantages that (a) any drift in the baseline will distort the results (unless an appropriate correction is made), and (b) if there is more than one current amplitude there is no way to tell which *duration* (measured as below) is associated with which current amplitude.

Most real records contain some baseline irregularities, and often 'glitches' too, so in practice this method may be as slow to compute as the alternatives because of the necessity to correct properly for these imperfections. This method is also unsuitable if there are many openings or shuttings that are too short to reach full amplitude; in this case the histogram will be very smeared. It is useful, therefore, to confine the histogram to those points that correspond to periods when the channel is open, the

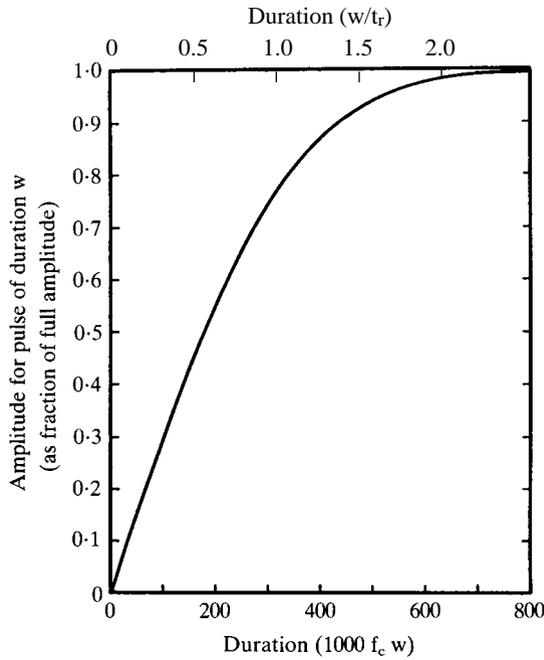


Fig. 3. Graph showing on the ordinate the maximum amplitude of the signal produced by a rectangular pulse of duration w . The amplitude is expressed as a fraction of the full amplitude (that produced by a long pulse). The abscissae are expressed in dimensionless units. The lower abscissa uses $1000 f_c w$ where f_c is the -3 dB frequency of the Gaussian filter (see text); These numbers represent the duration, w , in microseconds for $f_c = 1$ kHz, and *pro rata* for other f_c values. On the upper abscissa is plotted the duration of the pulse relative to the filter risetime, t_r . The fractional amplitude, up to a value of 0.5, is approximately equal to w/t_r (within 8%), e.g. for a fractional amplitude of 0.50, $w = 0.54 t_r = 0.179/f_c$. See also Fig. 5 and Table 1.

Table 1. Duration (w , in μs) of rectangular pulses required to produce a response that reaches the specified fraction of the full amplitude, for Gaussian filter with various cutoff frequencies (f_c , -3 dB)

f_c (kHz)	t_r (μs)	Fraction of full amplitude		
		0.25	0.5	0.9
		w (μs)		
1	332	83	179	432
2	166	42	90	216
3	111	28	60	144
4	83	21	45	108
5	66	17	36	86

For fractional amplitudes of 0.25, 0.5 and 0.9, the values of w correspond to $w/t_r = 0.25, 0.54$ and 1.30 respectively. See also Fig. 3.

open point-amplitude histogram, so the smearing effects of the open-shut transitions are avoided; in order to be sure that the channel is fully open, it will be necessary to allow at least two risetimes to elapse before collecting points for the histogram. Similarly a *shut-point amplitude histogram* can be made by including only points from periods when the channel is shut.

Separate amplitude estimates for each opening

In this method a separate estimate of amplitude is made for each period for which the channel is open.

The amplitude may be estimated by averaging points at the shut level, and points at the open level (after allowing long enough after opening for the filter transient to be completed). However the mean of points at the open level is liable to be biased if the opening actually contains brief undetected closures. It may therefore be better (and will in any case be acceptable) to fit the open and shut level by eye, using cursors on the computer screen. Alternatively, for events that are sufficiently long, the amplitude can be estimated by a least squares fit, simultaneously with the duration, by the time-course method described in the next section.

The amplitude estimates thus found can be plotted as a histogram (as in Fig. 2). In this case the area under each peak represents the *number of sojourns* at that level (rather than the time spent at that level). The form of the expected distribution is not exactly Gaussian (see Colquhoun & Sigworth, 1983) but the deviation is usually sufficiently small that little harm is likely to come from fitting the results (by maximum likelihood - see below) with Gaussian curves, as illustrated in Fig. 2.

One great advantage of this method is that the *duration* of each individual opening is known, as well as the amplitude. This allows, for example, the examination of open time distributions that are restricted to openings that are in a specified amplitude range. This cannot be done if the only information about amplitudes comes from a point-amplitude histogram.

5. The measurement of durations

The information about channel mechanisms that is contained in a single channel record resides largely in the durations of the channel openings and shuttings. Measurement of open and shut times will, therefore, usually be required. There are two main problems to be solved; firstly transitions from one current level to another must be *detected*, then the duration of time between one transition and the next must be *measured*. For optimum results different criteria should be used for these two jobs.

Detection of transitions

The optimum methods for detection of transitions are described by Colquhoun & Sigworth (1983); they are complex, and involve, for example, knowledge of the spectral characteristics of the background noise. Such methods are virtually never used in practice at present; instead transitions are located as the points where the

current crosses a preset threshold level. Fortunately, this simple method is not much worse than optimal methods.

In looking for transitions we aim to locate as many as possible of the genuine transitions, while rejecting, as far as is possible, any changes in current which, though they may look at first sight like transitions, are actually caused by random noise, or by door-slamming, tap-turning, seal breakdown or other such hazards of real life. Random noise can be coped with by setting the filter appropriately, and by imposition of a realistic resolution on the data, as described below. Disturbances resulting from things like door-slamming or refrigerators switching on and off, as well as baseline drift, can be dealt with *only by visual inspection of the data*. No foolproof method has yet been devised for automatically tracking a drifting baseline.

False event rate

The random noise in the record will, from time to time, result in fluctuations of the current sufficiently large to give the appearance of a transition (e.g. to cross a threshold level), even though no transition has actually occurred. Such *false events* can be kept to a minimum by filtering the data heavily, but if it is filtered too much important details may be obscured. The number of false events per second, λ_f , i.e. the number of times per second that the current departs from the baseline level by more than some specified amount, ϕ , is given, approximately (see Colquhoun & Sigworth, 1983), by

$$\lambda_f \approx f_c e^{-\phi^2/2\sigma_n^2}, \quad (5.1)$$

where f_c is the -3 dB frequency of the recording system, and σ_n is the background rms noise.

Note that the notation ‘exp()’ is often used to denote ‘e to the power’, so that complicated expressions need not be written as superscripts; for example (5.1) would often be written as $\lambda_f \approx f_c \exp(-\phi^2/2\sigma_n^2)$.

In the context of threshold-crossing analysis ϕ would be the threshold current level (taking the shut level as zero). The false event rate thus depends on the filter setting, and on the ratio, ϕ/σ_n , of the threshold level to the rms baseline noise. Thus, when $f_c = 1$ kHz, a ratio, ϕ/σ_n , of 3 will result in about 11 false events per second, on average. Similarly $\phi/\sigma_n = 4$ corresponds to about 0.33 false events per second (one every 3 seconds), and $\phi/\sigma_n = 5$ corresponds to about one false event every 270 seconds. These rates change, *pro rata*, for other filter settings.

Measurement of transitions

Once transitions have been located, we then wish to estimate the time interval between adjacent transitions, i.e. to estimate the durations of openings and shuttings (and, possibly, to estimate the amplitudes of the events at the same time). Two methods are in common use (1) threshold-crossing and (2) time course analysis. The latter provides better resolution, but the former will usually be faster, though this depends on the amount of manual checking and error correction that is done.

Whichever method is used, it will be necessary, before starting the analysis, to measure the following two quantities.

(a) The amount of baseline noise should be measured by finding a stretch of baseline free from obvious events, and calculating the standard deviation of these points, i.e. the root mean square (rms) baseline noise, which will be denoted σ_n .

(b) A preliminary estimate of the channel amplitude (denoted A_0) should be made, by choosing some long openings that are easy to measure (this will be needed to position the threshold line and to fit the durations of events that are too short for their amplitude to be measured).

The threshold crossing method

Usually a threshold is set halfway between the fully open and the shut current levels; every time the observed current crosses this '50 percent threshold' a transition is deemed to have occurred, and the duration of an event is measured as the length of time for which the current stays above (or below) this threshold. It is important, therefore, to filter the data so that spurious transitions (false events - see (5.1) above) are rare.

Setting the filter for threshold-crossing analysis. A 50% threshold corresponds to $\phi/A_0 = 0.5$, where ϕ is the threshold level and A_0 is the channel amplitude (see above). The setting of the filter can be illustrated by an example, shown in Fig. 4, in which the channel amplitude was found to be $A_0 = 3.8$ pA. In this case, the 50% threshold will be set at $\phi = 1.9$ pA. *The filter setting, f_c , is now chosen so as to produce an acceptable false event rate.* In Fig. 4, the same channel opening (downward deflection, followed by two brief shittings) is shown filtered at 1, 1.5, 2, 3 and 4 kHz (-3 dB). The standard deviation of the baseline noise (the r.m.s. noise) was, respectively, 0.10 pA, 0.14 pA, 0.19 pA, 0.27 pA and 0.33 pA. Thus, for the least filtered record (4 kHz), we have $\phi/\sigma_n = 1.9/0.33 = 5.8$. From (5.1), with $f_c = 4000$ Hz, we find $\lambda_f \approx 0.00025$ s $^{-1}$, i.e. roughly one false event in 66 minutes. This is a low rate, so filtering at 4 kHz would be suitable for threshold crossing analysis. It might be thought that this is an excessively low rate, and even less filtering would be safe. However, there are several reasons why it is better to be on the safe side. Firstly, the false event rate depends very steeply on ϕ/σ_n ; in this case it was 5.8, but if ϕ/σ_n were to fall only to 5.0 (e.g. if the rms noise rose by only 15%, from 0.33 pA to 0.38 pA) the false event rate would go up from about one per hour to about one per minute. Secondly, this is worked out from the baseline (shut channel) noise, but the current is usually noisier when the channel is open so it is likely that there will be more false shittings than predicted. And thirdly, most programs are not capable of holding the threshold accurately half-way between baseline and open level, which gives more scope for false events than calculated here. For time-course fitting the choice of filter is less critical; in a case such as that shown in Fig. 4, 2 or 3 kHz would be used.

Measuring the intervals between transitions. The time at which a transition occurred can be estimated by taking the data point on either side of the crossing of the threshold line, and interpolating between them to estimate the time at which the threshold is crossed. Some people simply count the number of data points between one threshold

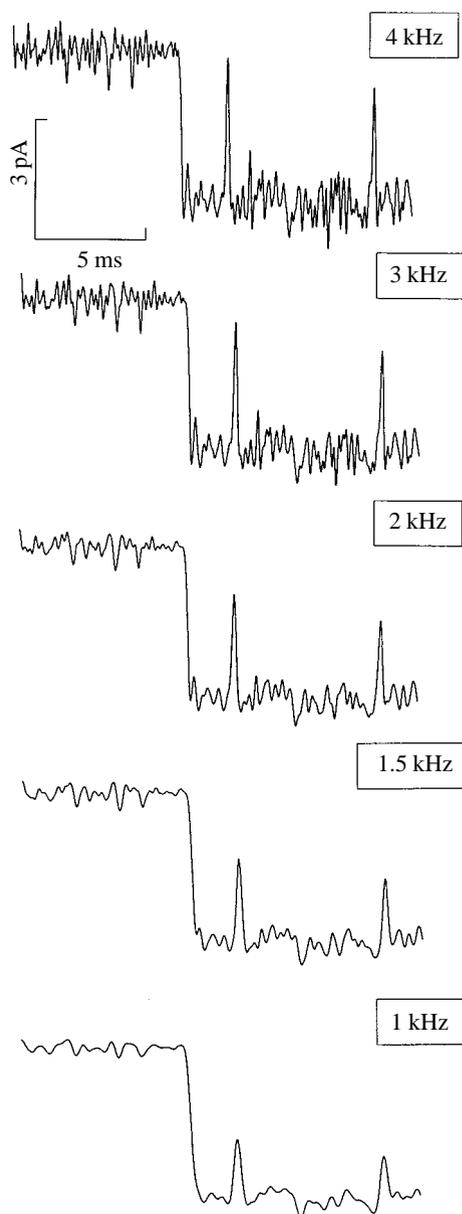


Fig. 4. Examples of a single channel record with various settings of the low pass filter. This channel comes from the record that was used for Fig. 10. The tape was replayed through an 8-pole Bessel filter set (top to bottom) at 4, 3, 2, 1.5 and 1 kHz (-3 dB). The channel amplitude is 3.8 pA, and the standard deviation of the baseline noise (the r.m.s. noise) is (top to bottom) 0.33 pA, 0.27 pA, 0.19 pA, 0.14 pA and 0.10 pA. The corresponding values of ϕ/σ_n are 5.8, 7.0, 10, 14 and 19 respectively. The data were originally recorded (after prefiltering at 20 kHz) on FM tape at 15 inches/second, giving a bandwidth up to 5 kHz. The output filter of the tape recorder was Tchebychev type, which is quite flat up to 5 kHz and then rolls off steeply. Nevertheless, the effective overall filtering, taking into account the patch clamp and tape recorder, will be somewhat more than is indicated (especially for the 4 kHz filter setting).

crossing and the next in order to estimate the duration, but, unless the data sampling rate is very fast, this method will be unnecessarily inaccurate for short events.

The *resolution* of the method, i.e. the shortest time interval that can be measured, is dictated by the signal-to-noise ratio of the data, and the filtering that must consequently be employed. Events that fail to reach 50 percent of full amplitude will, of course, be missed entirely by a 50 percent threshold detection method (see eq. 4.1, Fig. 3 and Table 1). Those that are just above 50 percent will be detected, but clearly the duration of intervals will be underestimated.

For example suppose that the filter is set at a -3 dB frequency of $f_c = 1$ kHz, so $t_r = 332 \mu\text{s}$ (see (3.1) and Table 1). Events (openings or shittings) of $179 \mu\text{s}$ or longer reach the 50 percent threshold (on average - the presence of random noise means that this will not happen every time). An event, of say, $190 \mu\text{s}$ duration would remain above the threshold for a short time only, and its duration would be seriously underestimated. It is shown by Colquhoun & Sigworth (1983) that the duration of events needs to be above roughly $1.3t_r$ (i.e. $430 \mu\text{s}$ in this example) before errors from this source become negligible. Although it is possible to correct for this effect, the correction is inexact in the presence of noise and is not usually used, so the resolution of the analysis is limited to 400 - $500 \mu\text{s}$ effectively, despite the fact that events much shorter than this can be *detected*.

Effect of duration on amplitude measurements. An opening must have a duration of at least $2t_r$ (of the order of 1 ms in the example above) before its amplitude can be measured reliably (see Fig. 3 - durations down to about t_r permit tolerable simultaneous fit of duration and amplitude, but an opening needs to last at least $2t_r$ before it is *clear* that the full amplitude has been reached; see Colquhoun & Sigworth, 1983).

For brief openings the response fails to reach full amplitude, as described above, and for any event much shorter than t_r , the shape of the response depends on the *area* of the pulse (i.e. the total charge passed during the opening), rather than on its amplitude. Any sort of brief current pulse will produce an observed response of the same shape; doubling the amplitude but halving the duration will result in indistinguishable responses (see Fig. 11-10 in Colquhoun & Sigworth, 1983). Brief events that do not reach full amplitude can therefore be fitted only if a value for the full amplitude is *assumed* (this applies equally to time course fitting). If there is only one sort of channel in the patch this is not a problem, but if the record contains large and small amplitude channel types then the duration of a brief event that falls short of the smaller amplitude cannot be estimated because there is no way to tell which sort of channel it originated from. In this case durations can be measured only for events that reach full amplitude so the resolution is drastically reduced (to about $2t_r$) for the whole analysis for the purposes of *fitting* of distributions to event durations, though the resolution for *detection* of events, and therefore the resolution to be imposed on the data (see below), may be much better than this.

The time course fitting method

The step-response function of the recording system. The time taken for the transition

from a shut channel to a fully open channel is very short (less than $10 \mu\text{s}$, Hamill *et al.* 1981), so the shape of the observed currents is almost entirely determined by the frequency response characteristics of the system. These characteristics may depend, for example, on the preparation itself, the patch clamp electrode, the patch clamp electronics, the tape recorder, and the filter that is used to limit high-frequency noise (see (3.2) above). The filter usually has the biggest effect. The response of a recording system to a step input can be measured as follows. A rectangular step input (intended to simulate the opening of a channel) can be induced by holding near to the headstage a wire connected to a high-quality triangular wave generator; the output is tape-recorded and filtered as in a real experiment. This output will be rounded, as illustrated in Figs 5 and 6. The result may depend on the tape speed (which controls

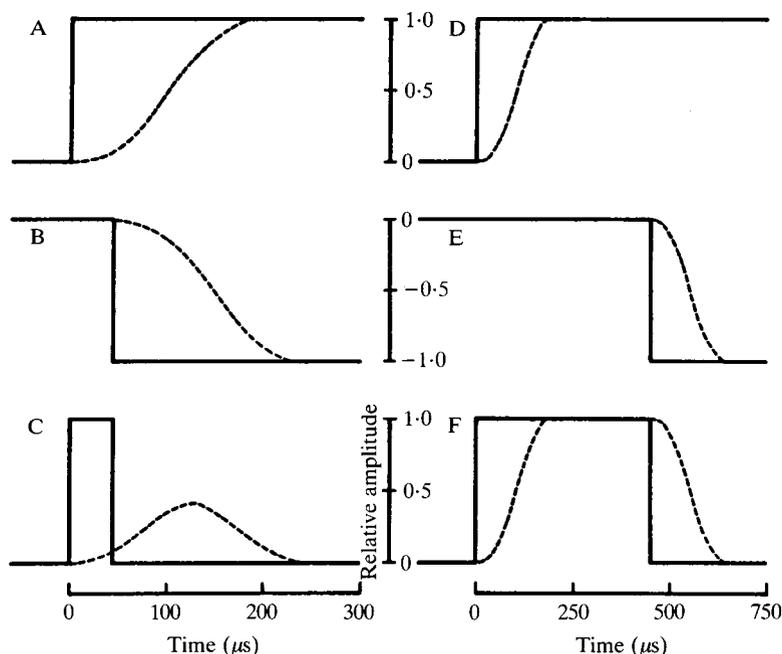


Fig. 5. Illustration of the method of calculation of the expected response to a step input. The left-hand column illustrates a short ($45 \mu\text{s}$) pulse, and the right-hand column a longer ($450 \mu\text{s}$) pulse. The dashed lines in A and D show (on different time scales) the experimentally measured response to a step input, shown schematically as a continuous line, for a system (patch clamp, tape recorder and filter) for which the final filter (eight pole Bessel) was set at 3 kHz (-3 dB). The rise time, t_r , of the filter is about $111 \mu\text{s}$ (see Table 1), so the pulse widths are $45 \mu\text{s} = 0.406t_r$ and $450 \mu\text{s} = 4.06t_r$. (A) The response to a unit step at time zero is shown. B shows the same signal but shifted $45 \mu\text{s}$ to the right and inverted. The sum of the continuous lines in A and B gives the $45 \mu\text{s}$ unit pulse shown as a continuous line in C. The sum of the dashed lines in A and B is shown as a dashed line in C and is the predicted response of the apparatus to the $45 \mu\text{s}$ pulse. It reaches about 41% of the maximum amplitude, which is very close to the value of 39% expected for a Gaussian filter (see Table 1 and Fig. 3). D, E and F show, except for the time scale, the same as A, B and C but for a $450 \mu\text{s}$ pulse, which achieves full amplitude. Reproduced with permission from Colquhoun & Sigworth (1983).

the bandwidth of the recorder), and it may depend on whether the output filters on the tape recorder are set to Bessel or Tchebychef type (on recorders where they are switchable, such as the Racal recorders).

Fitting channel transitions. The measured step response can be used to calculate the response to any pattern of channel opening and shutting, by the method shown in Fig. 5. In the case of long openings (or long shuttings) the effect of the filtering is merely to round off the square corners of the transition, as illustrated by the first opening shown in Fig. 6A. But a variety of other patterns can be produced when short openings and shuttings occur; some of these are illustrated in Fig. 6B,C,D. Such calculated responses ('convolutions of step responses' in the usual jargon) can be superimposed on the observed current, and the time intervals and amplitudes adjusted until a good fit is obtained.

The fit may be judged by eye or by a least squares criterion. A least squares fit

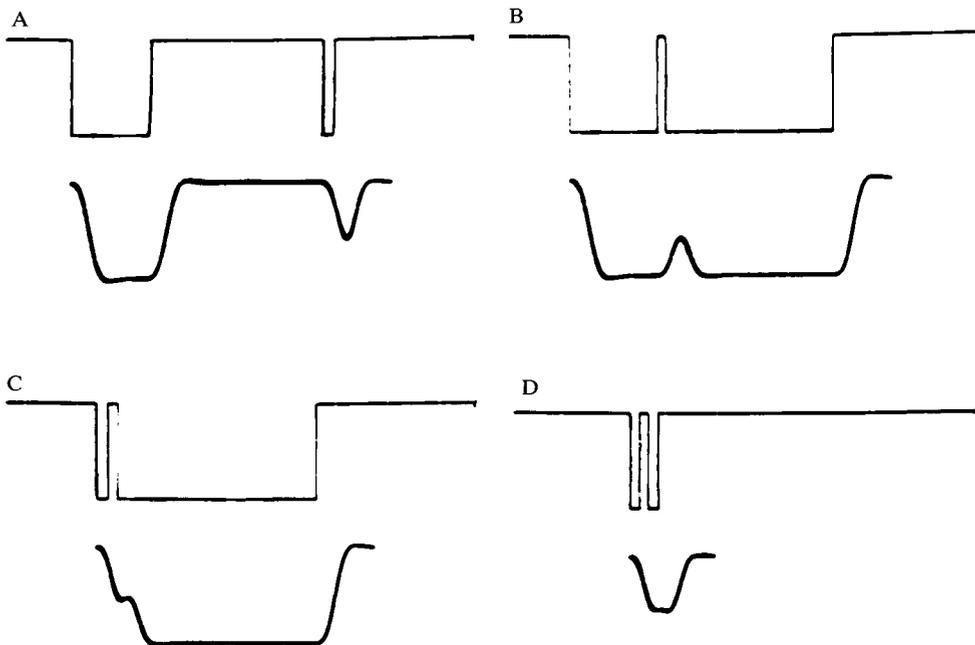


Fig. 6. Examples of the calculated output of the apparatus (lower traces) in response to two openings of an ion channel (upper traces). The curves are generated by a computer subroutine and were photographed on a monitor oscilloscope driven by the digital-to-analogue output of the computer. Openings are shown as downward deflection. (A) A fully resolved opening (435 μ s) and gap (972 μ s) followed by a partially resolved opening (67 μ s). (B) Two long openings (485 and 937 μ s) separated by a partially resolved gap (45.5 μ s). (C) A brief opening (60.7 μ s) and gap (53.1 μ s) followed by a long opening (1113 μ s); this gives the appearance of a single opening with an erratic opening transition. (D) Two short openings (both 58.2 μ s) separated by a short gap (48.1 μ s); this generates the appearance of a single opening that is only 55% of the real amplitude but which appears to have a more-or-less flat top, so it could easily be mistaken for a fully resolved subconductance level. Reproduced with permission from Colquhoun & Sigworth (1983).

(even when amplitudes are estimated simultaneously) can be quite rapid on a modern fast PC-compatible computer. Examples of such fits are shown in Figs 7 and 8.

Advantages and disadvantages of these two methods

Speed of analysis. As commonly practiced, the threshold crossing method is considerably faster than time course fitting, and when many channel openings have to be measured this is not a trivial consideration. However the speed difference depends

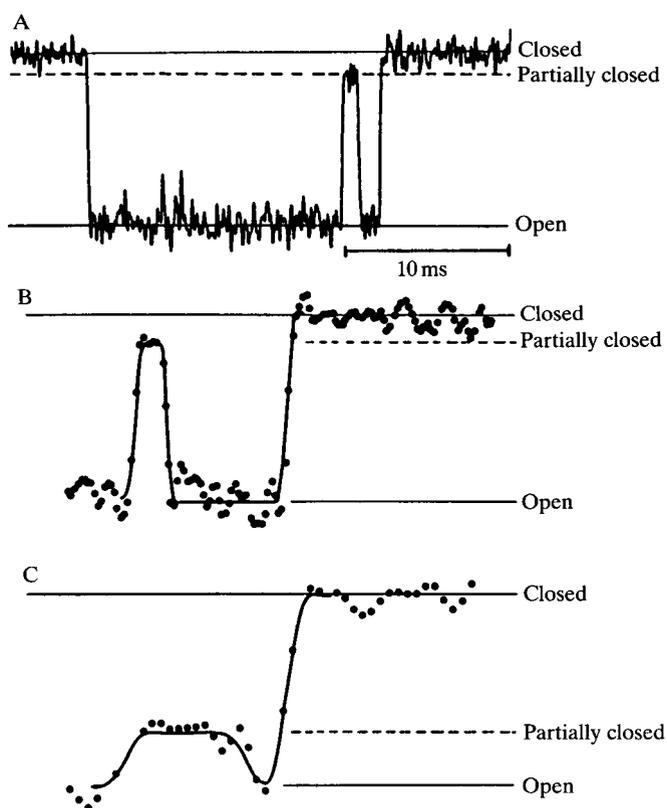


Fig. 7. Examples of single-channel current sublevels. (A) Elementary current activated by 100 nM ACh. -125 mV, 11°C . The two continuous horizontal lines, marked closed and open respectively, represent the patch current when a channel is either completely closed or completely open. The average amplitude of the current through the fully open channel is -3.71 pA. The dashed horizontal line represents the amplitude of a current sublevel. During the sublevel the channel is partially closed. The sublevel amplitude is -0.52 pA, i.e. 14% of the full amplitude. (B,C) Partial channel closures in another patch with 500 nM ACh at -178 mV and 10°C . The time course of the digitized current record is fitted (continuous line) by the same step response function as was used for full openings and closings. However, it was assumed that the amplitude of the current sublevels (marked partially closed) are 17% and 72%, respectively, of the full current amplitude. Sublevel amplitudes are indicated by the horizontal dashed line. The current through the fully open channel is -5.6 pA in (B) and -5.7 pA in C. The duration of the partial closures is 310 μs and 360 μs in B and C respectively. Filtered at 4 kHz (-3 dB). Reproduced with permission from Colquhoun & Sakmann (1985).

very much on the amount of checking that is done. It obviously takes time to check visually the fit to every opening, and the position of the baseline before and after the opening. Time-course fitting forces you to make these checks, but threshold crossing allows you to neglect them if you wish to do so. Furthermore, second generation time-course fitting programs will estimate both durations and amplitudes

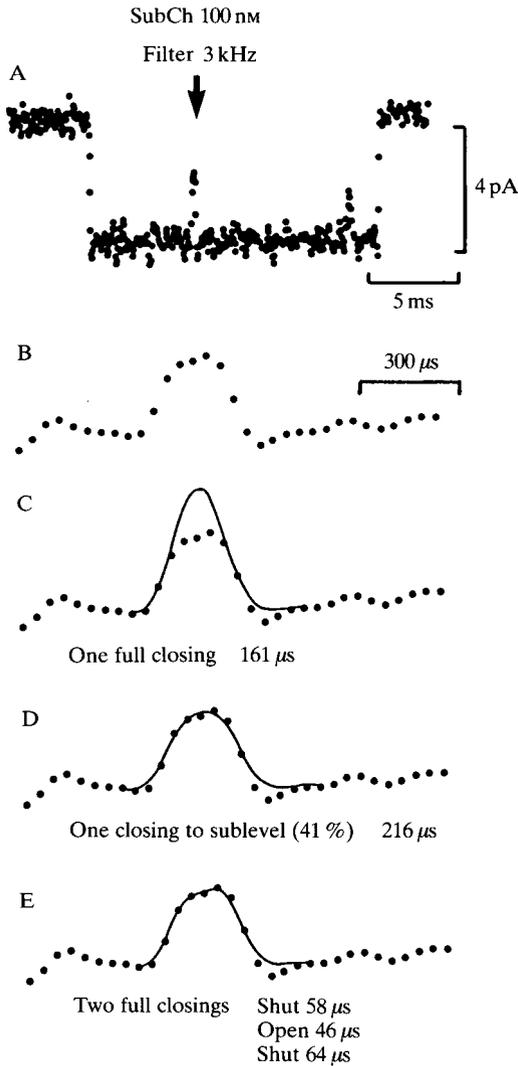


Fig. 8. Illustration of the problem of distinguishing subconductance states from multiple transitions. (A) Burst induced by SubCh, 100 nM, at -128 mV. Two putative brief gaps are visible. Low pass filter at 3 kHz (-3 dB). (B) The gap that is marked with an arrow in A shown on an expanded time scale. (C) Fit of data in B assuming a single complete closure of duration 161 μ s. (D) Fit of data in B assuming a single closure to a subconductance state, of duration 216 μ s. (E) Fit of data in B assuming that a full closure of 58 μ s is followed by a full opening of 46 μ s, and then another full closure of 64 μ s. Reproduced with permission from Colquhoun & Sakmann (1985).

simultaneously, whereas threshold-crossing programs usually require separate estimation of amplitudes. Estimation of amplitudes requires an additional, separate, job, e.g. making a point amplitude histogram; this itself can take some time unless the record is of such high quality that the baseline never drifts by more than a small fraction of the channel amplitude, and contains a negligible number of glitches. Even if this is done, information about the time spent at each amplitude level is lost.

Multiple conductance levels. Threshold-crossing methods are completely incapable of measuring channels that contain more than one conductance level. This is illustrated in Fig. 9, where a subconductance at about 50% of the full conductance level causes complete havoc. Lower subconductance levels would be missed entirely. The data in Fig. 9 are from a nicotinic channel; the problem is far more serious for GABA and glutamate channels which have frequent sublevel transitions.

Temporal resolution. The method of time course fitting clearly allows one to fit events that would be quite impossible with a threshold-crossing method, either (1)

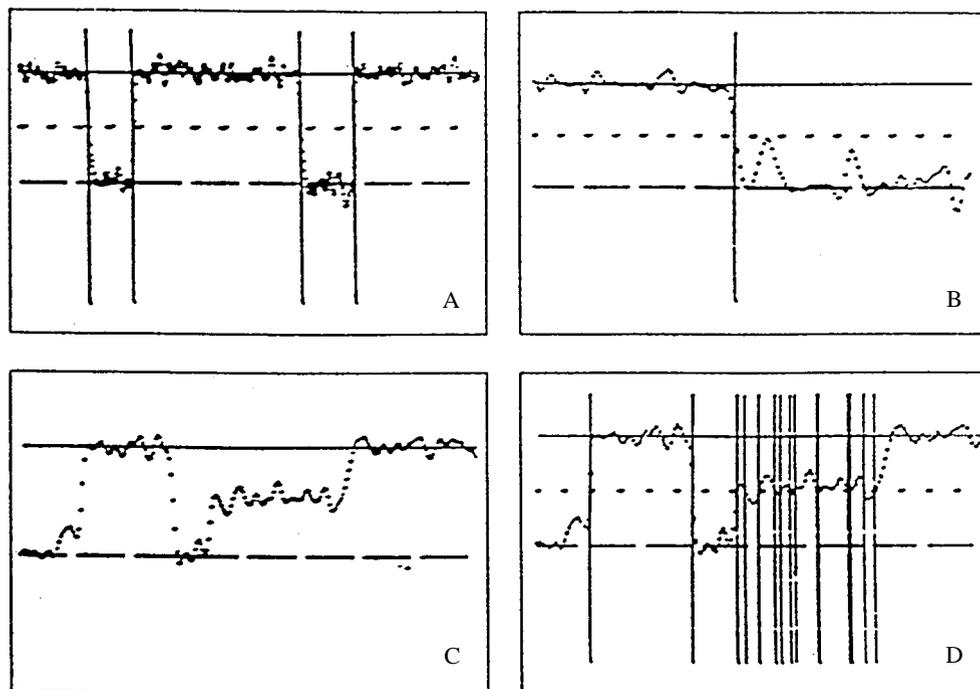


Fig. 9. Single channel currents recorded from *Xenopus* oocytes injected with mRNAs for α , β , γ and δ subunits of the BC3H1 nicotinic receptor. Cell-attached patch, acetylcholine 1 μ M, filtered at 4 kHz (-3 dB) and sampled at 40 kHz. The solid horizontal line represents the baseline (shut level), the long-dashed line is the open level, and the short-dashed line is the threshold level. Vertical lines mark the positions of threshold crossings. In A and B, transitions are successfully located by the threshold-crossing method, but the channel shown in C and D happens to have a sublevel close to the threshold level, so, as shown in D, a large number of entirely spurious transitions are generated. If the program was operating automatically, these would all go into your results! (C. G. Marshall, A. J. Gibb & D. Colquhoun, unpublished data.)

because they are too brief to reach the threshold at all (or too brief to have their duration measured accurately in this way, as discussed above) as in Fig. 6A,B, or (2) because brief events close to longer ones produce an apparent distortion as in Fig. 6C,D. An automated threshold crossing method would, of course, produce *some* result when faced with such events, but the result would be meaningless; this is one reason for not using excessively automated methods. Time course fitting allows a resolution that is of the order of three-fold better than the threshold crossing method; briefer events can be fitted.

The most obvious danger of time course fitting is that of over-ambitiousness. The risk of false events arising from an attempt to fit events that are too short is just the same as described above for the threshold crossing method, but the temptation to fit them may be greater. Not only genuine channel transitions, but *any* disturbance (artefactual or random noise), if sufficiently brief, will produce a signal that can plausibly be fitted by the time- course method. For example Fig. 8 shows a case in which it is impossible to be sure whether a subconductance level should be fitted, or whether there are actually two full shuttings in quick succession. One way to deal with this part of the problem is to go through the data twice; the first time full closings are fitted whenever possible, the second time sublevels are fitted whenever possible. It may then be possible to decide which is the more plausible e.g. by seeing whether the sublevel fits produce a consistent amplitude estimate. Such ambiguities are, of course, not avoided by a threshold crossing analysis; they are merely brushed under the carpet.

Deciding the resolution

It is desirable, for several reasons, that the data should have a well-defined time resolution, i.e. that it be known that all events above the specified resolution, but none below it, have been fitted. For example, it is this time resolution that fixes the false event rate, it is the time resolution that fixes the minimum duration of an open or shut time that can be fitted (but see below). Furthermore, a well-defined time resolution is essential if any form of correction for missed events is to be applied (see below and Chapter 7).

It might be thought that the resolution is automatically fixed in the threshold-crossing analysis (as the pulse duration required to reach 50% of full amplitude), but (a) this is not constant (see Colquhoun & Sigworth, 1983), and (b) the safe resolution will be greater than this, as described above. In the case of time course fitting the shortest durations that are fitted are decided subjectively, and are unlikely to be constant throughout an experiment. It is therefore highly desirable that a fixed resolution should be imposed on the data *after* analysis, as described below. First, though, a decision must be made as to the appropriate value(s) for the resolution.

The false event rate (per second) that is acceptable obviously depends on the value of the true event rate. If we are looking for openings that occur at only one per 5 seconds on average, then the record contains a lot of shut baseline in which false events can occur. If a false event rate of about 2% of the true rate were thought tolerable then we might aim for a false event rate of about one per 270 seconds. The

threshold would therefore be set to 5 times the baseline rms noise (see (5.1) above). In the example discussed above, with rms noise of 0.3 pA and a channel amplitude of $A_0 = 3$ pA, openings that have an amplitude less than $5 \times 0.3 = 1.5$ pA would therefore be ignored. Openings of this amplitude, which is $1.5/3.0 = 0.5$ (50%) of the full amplitude, would (on average) have a duration of about $0.54t_r$, from (4.1) and Fig. 3. For example if the overall filtering of the data corresponded to $f_c = 1$ kHz (-3 dB), so $t_r = 332 \mu\text{s}$, then openings of $0.54 \times 332 \approx 180 \mu\text{s}$ could be safely resolved (and *pro rata* for other values of f_c). Note, though, that although this resolution would be safe for *detection* of openings, it could be used for fitting them only if it was safe to *assume* an amplitude for such partially resolved openings; it requires a duration nearer to $2t_r$ for the amplitude to be resolved (see above).

For shut times, different values may apply. Suppose, in the example above, that openings occur in bursts, separated by short gaps, so there is relatively little open level in the record in which false gaps might be detected. For example Colquhoun & Sakmann (1985) observed brief gaps at a rate of about 500 per second of open time with acetylcholine. A false event rate of 2% of true rate, as before, now corresponds to about 11 false events per second, so the threshold could be set to only 3 times the baseline rms noise (see (5.1) above). In the example just discussed, shittings that have an amplitude less than $3 \times 0.3 = 0.9$ pA would therefore be ignored. Shittings of this amplitude, which is $0.9/3.0 = 0.3$ (30%) of the full amplitude, would (on average) have a duration of about $0.31t_r$, from (4.1) and Fig. 3. For example if the overall filtering of the data corresponded to $f_c = 1$ kHz (-3 dB), so $t_r = 332 \mu\text{s}$, shittings of $0.31 \times 332 \approx 100 \mu\text{s}$ could be safely resolved (and *pro rata* for other values of f_c), though the safe resolution for open times is at least $180 \mu\text{s}$, as just described.

These values change *pro rata* for other filter settings; e.g. a resolution of $25 \mu\text{s}$ for shut times might be obtainable at $f_c = 4$ kHz, as long as the signal to noise ratio of the data was good enough to give 0.3 pA rms noise at this filter setting. Such resolution can be obtained only by time course fitting. Some improvements might be obtained in threshold crossing analyses by using separate thresholds for openings and for shittings, though this is not usually done, and would obviously cause problems if two or more successive events were brief.

Clearly the resolution that can be attained safely is not known until *after* the record has been analysed (so the optimum setting of thresholds for analysis is difficult). Our strategy with time course fitting evades this problem by fitting everything that could possibly be a real event while going through the data. At the end of the analysis the resolution must be specified, and this can be decided on the basis of several criteria: (a) the subjective feeling of the operator, during the analysis, concerning the shortest event that he or she can be sure is genuine, (b) calculations of false event rates of the sort illustrated above, based on the tentative analysis of the results, and (c) the appearance of the raw distributions of open and shut times, in particular the duration of interval below which there is 'obviously' a deficit of events. All of these criteria are, to some extent, subjective, but together they should allow realistic and safe values to be chosen for the best resolutions for open times and for shut times.

Imposition of the resolution. Once the resolution for open times has been decided

then every open time shorter than the chosen value is treated, along with the shut time on each side of it, as one long shut period. An analogous procedure is followed for all shut times that are shorter than the chosen shut time resolution. The result is a record with a consistent resolution throughout; it contains no openings shorter than the chosen open time resolution and similarly for shut times. The open and shut times in this record are now ready to have distributions fitted, as described below.

This simple procedure is always used in our laboratory, though it does not seem to be widespread. Its neglect clearly leads, in principle, to inconsistency. Suppose, for example, that the open time resolution is decided (as is common) simply by looking at the distribution of raw open times and choosing the resolution as the duration below which there is an 'obvious' deficit of observations. This may well give a realistic estimate of the resolution for open times, but if we then go ahead and fit all open times in the raw data that are longer than the resolution so chosen we shall be fitting as distinct openings some pairs of openings that are separated only by a shut period that is shorter than the shut-time resolution, and which we therefore have no right to regard as well-defined separate openings. An analogous inconsistency obviously arises for the fitting of shut times. The problem is easily avoided if consistent open time and shut time resolutions are imposed on the data, as described above, before any fitting is attempted.

Measurement of P_{open}

The probability, P_{open} , that a channel is open (also known, for brevity, as the *open probability*) can be estimated as the fraction of time for which a channel is open, i.e. the total open time divided by the total length of the record.

A useful estimate of P_{open} can be obtained only when there is only one individual channel contributing to the record, and this is usually not the case. Sometimes, though, there are *sections* of the record that originate from one channel only. For example, at high agonist concentrations many channels show long silent periods during which all the channels in the patch are desensitized. Periodically one channel emerges from the desensitized state, and opens and shuts at a high rate (because the agonist concentration is high). The lack of double openings during such periods of high activity shows that they originate from one channel only. Therefore the silent desensitized periods can be cut out from the record (provided that they are so long as to be obviously desensitized), and P_{open} calculated from the periods of high activity only (e.g. Sakmann, Patlak & Neher, 1980; Colquhoun & Ogden, 1988). This has been used as a method for obtaining equilibrium concentration-response curves that are corrected for desensitization (the desensitized periods are cut out). This method also has the advantage that the response, P_{open} , is on an *absolute* scale (the maximum possible response is known, *a priori*, to be 1).

Once a list of all the open and shut times has been measured, it is simple to calculate P_{open} for any specified part of the record (see, for example, section on stability plots, below). However it is usually preferable not to measure P_{open} in this way, but rather to measure it by integrating the record (either numerically on the

computer, or by playing the magnetic tape through an analogue integrator). Then P_{open} can be found as the area under the trace per unit time, divided by the single channel current amplitude. The great advantage of measuring P_{open} in this way is that it is insensitive to missed events (see section 10, below). The filtering of the signal attenuates and rounds the single channel signal (as shown in Figs 5-8), but *does not change the area* under the signal. Therefore the integration method is unaffected by the existence of undetected transitions.

6. The display of distributions

This section deals mainly with the display of measurements that have been made at equilibrium, so the average properties of the record are not changing with time (this can be checked by use of the stability plots described at the end of this section).

Following the analysis described above, we should have a list of our estimates of the durations of each (apparent) open period (together with the amplitude of the current) and of each (apparent) shut period in the order in which they occurred, each duration being greater than the chosen resolution. These durations are random variables so, in order to describe them quantitatively, a probability distribution must be fitted to them. We shall deal here mainly with the fitting of distributions that are described by the sum of one or more exponential (or geometric) components. This form of distribution is expected under the simplest (Markov) assumptions concerning the mechanism of channel opening; these assumptions are described in more detail in Chapter 7 and, for example, by Colquhoun & Hawkes (1983). Although this sort of distribution is what everybody uses it should be borne in mind that it will not be the correct form (a) when the resolution is such that many brief events are missed (though often this will not cause great deviations from exponential form) or (b) when the mechanism of channel opening does not obey the simple Markov assumptions (e.g. because membrane potential or ligand concentration are not held constant). Whatever the form of the distribution, it is described (for a continuous variable such as time) by a probability density function.

Probability density functions and histograms

The data consist of a list of times (e.g. open times or shut times or burst lengths); in order to display their distribution they must be displayed as a histogram. The histogram is usually described as showing the number of observations with durations that fall between the limits specified on the abscissa, i.e. its ordinate is dimensionless. The probability density function (p.d.f.), on the other hand, is a function such that the area under the curve (rather than the amplitude on the ordinate) represents probability, or number of observations. The p.d.f. therefore has dimensions of s^{-1} , and a total area of unity (see equation (1) below, and Chapter 7). Usually we will wish to superimpose a fitted p.d.f. on the histogram of the data, but the p.d.f. has dimensions that appear to be different from those of the histogram ordinate. The solution of this dilemma is that the histogram ordinate should be expressed not as

frequency of events, but as a *frequency density*, for example, as ‘frequency per 2 ms’ (where 2 ms may be the bin width); thus it is expressed in reciprocal time units like the p.d.f. This is not only correct, but also makes it clear that *area* must represent frequency in the histogram if all bins are not of the same width. To achieve superimposition on the histogram, the area under the p.d.f. must be made to correspond to the *area* of the histogram boxes, $N\Delta t$, by multiplying the p.d.f. by $N\Delta t$ where Δt is the bin width and N is the total number of observations (including the estimated number that are below the chosen resolution and therefore not detected; see below). Further details are given by Colquhoun & Sigworth (1983).

A simple exponential p.d.f. for an interval of duration t is defined by

$$f(t) = \tau^{-1}e^{-t/\tau} \quad (6.1)$$

where τ is the mean of the distribution (which is the same thing as what would be called the ‘time constant’ of the curve if it were a decaying current rather than a p.d.f.). The initial constant, τ^{-1} , ensures that the p.d.f. has unit area. If there is more than one exponential component the distribution is referred to as a *mixture of exponential distributions* (or a ‘sum of exponentials’, but the former term is preferred since the total area must be 1). If a_i represents the area of the i th component, and τ_i is its ‘mean’ then

$$\begin{aligned} f(t) &= a_1\tau_1^{-1}e^{-t/\tau_1} + a_2\tau_2^{-1}e^{-t/\tau_2} + \dots \\ &= \sum a_i\tau_i^{-1}e^{-t/\tau_i} \end{aligned} \quad (6.2)$$

The areas add up to unity, i.e. $\sum a_i = 1$, and they are proportional, roughly speaking, to ‘number of events’ in each component. The overall mean duration is

$$\text{mean duration} = \sum a_i\tau_i. \quad (6.3)$$

The cumulative distributions. The *cumulative* form of this distribution, the probability that an interval is *longer than* t , is, for a single exponential,

$$P(\text{interval} > t) = \int_t^\infty f(t)dt = e^{-t/\tau}, \quad (6.4)$$

or, for more than one component, the sum of such integrals, *viz.*

$$P(\text{interval} > t) = \sum a_i e^{-t/\tau_i}. \quad (6.5)$$

Occasionally the data histogram is plotted in this cumulative form with the fitted function (6.5) superimposed on it. This presentation will always look smoother than the usual sort of histogram (the number of values in the early bins is large), but it should *never* be used, because the impression of precision that this display gives is *entirely spurious*. It results from the fact that each bin contains all the observations in all later bins, so adjacent bins contain nearly the same data. In other words successive points on the graph are not independent, but are strongly correlated, and this makes the results highly unsuitable for curve fitting.

To make matters worse, it may well not be obvious at first sight that cumulative

distributions have been used, because the curve, (6.5), has exactly the same shape as the p.d.f. (6.2). There are no good reasons to use cumulative distributions to display data; they are highly misleading. In any case, it is much easier to compare results if everyone uses the same form of presentation.

Display of multi-component histograms

Figure 10A shows a histogram of shut times, with a time scale running from 0 to 1500 ms. This range includes virtually all the shut times that were observed.

The first bin actually starts at $t = 60 \mu\text{s}$ rather than at $t = 0$, because a resolution of $60 \mu\text{s}$ was imposed on the data (see above) so there are no observations shorter than

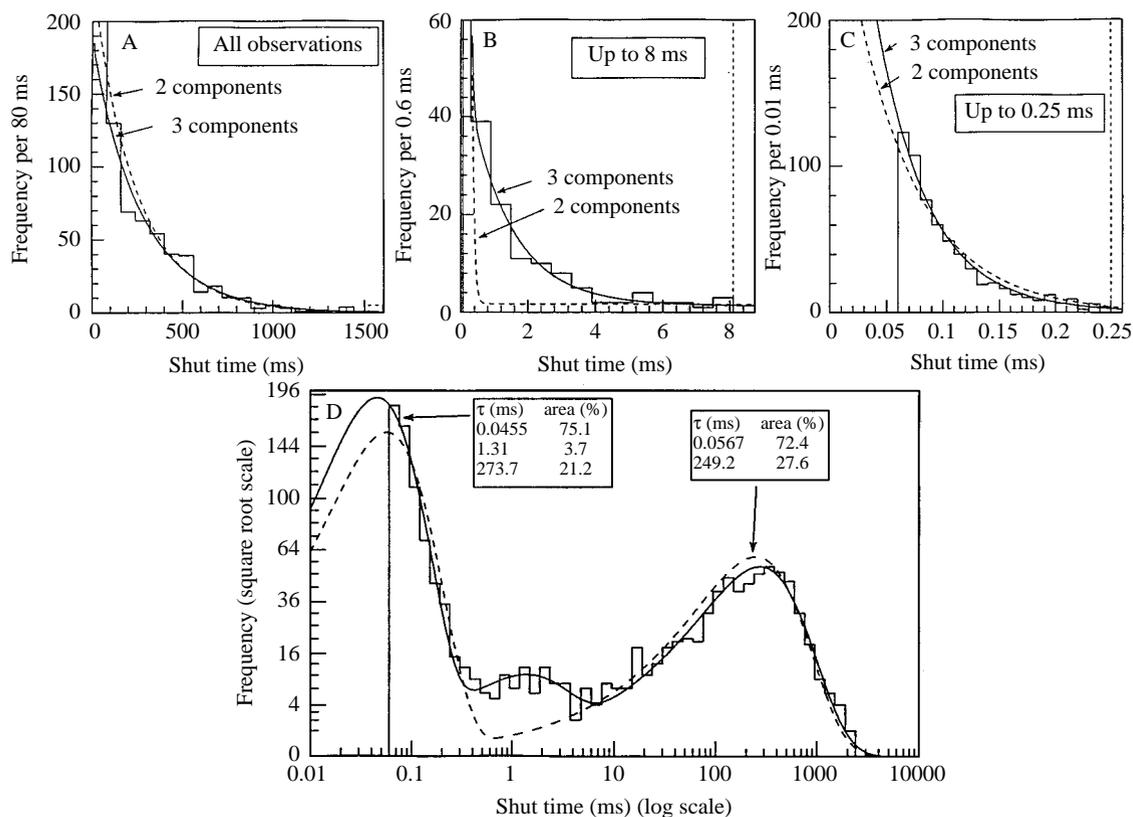


Fig. 10. Example of a distribution of shut times. In A, B and C the histogram of shut times is shown (on three different time scales), and in D the distribution of $\log(\text{shut times})$, for the same data, is shown. The data are from nicotinic channels of frog endplate (suberyldicholine 100 nM, -130 mV). Resolutions of $80 \mu\text{s}$ for open times, and $60 \mu\text{s}$ for shut times, were imposed as described in the text; this resulted in 1348 shut times which were used to construct each of the histograms. The dashed bins (which are off scale in B, C) represent the number of observations above the upper limit. The data were fitted by the method of maximum likelihood with either two exponentials (dashed curve) or three exponentials (continuous curve). The same fit was superimposed on all of the histograms. The estimated parameters are shown in D. (D. Colquhoun and B. Sakmann, unpublished data.)

this. All that is visible on this plot is a single slowly-decaying component with a ‘mean’ of about 250 ms, though the first bin, the top of which is cut off on the display, shows that there are many short shut times too. The same data are shown in Fig. 10C, but only shut times up to 250 μ s are shown here (so the 60 μ s resolution is now obvious). There are many shut times longer than 250 μ s of course, and these are pooled in the dashed bin at the right hand end of the histogram (the top of which is cut off). Again the histogram looks close to a single exponential, but this time with a ‘mean’ of about 50 μ s. Although it is not obvious from either of these displays, there is in fact a (small) third component in this shut time distribution. It is visible only in the display of the same data in Fig. 10B, in which all shut times up to 8 ms are shown, where an exponential with a mean of about 1 ms is visible. The data were not fitted separately for Fig. 10A,B,C, but one fit was done, to all the data (by maximum likelihood - see below) with either two exponential components (dashed line) or 3 exponential components (solid line). This same fit is shown in all four sections of Fig. 10. The deficiency of the 2 component fit is obvious only in the display up to 8 ms.

Clearly the conventional histogram display is inconvenient for intervals that cover such a wide range of values. The logarithmic display described next is preferable.

Logarithmic displays: the distribution of log(duration)

It was suggested by McManus, Blatz & Magleby (1987), and by Sine & Sigworth (1987), that it might be more convenient, when intervals cover a wide range (as in the preceding example), to look at the distribution of the logarithm of the time interval, rather than the distribution of the intervals themselves. Note that this is not simply a log transformation of the conventional display, because this would have bins of variable width on the log scale, whereas the distribution of $\log(t)$ is shown by bins of constant width on the log scale. In addition, Sine & Sigworth suggested using a square root transformation of the frequency density (to keep the errors approximately constant throughout the plot).

The distribution has the following form. If the length of an interval is denoted t , and we define

$$x = \log(t) ,$$

then we can find the p.d.f. of x , $f_x(x)$, as follows. First we note that if a t is less than some specified value t_1 , then it will also be true that $\log(t)$ is less than $\log(t_1)$. Thus

$$\text{Prob}[t < t_1] = \text{Prob}[\log(t) < \log(t_1)] = P, \text{ say.} \tag{6.6}$$

In other words the cumulative distributions for t and $\log(t)$ are the same. Now it is pointed out in chapter 7 (equations 3.7-3.9), that the p.d.f. can be found by differentiating the cumulative distribution. Thus, denoting the probability defined in (6.6) as P ,

$$\begin{aligned}
 f_x(x) &= \frac{dP}{dx} = \frac{dP}{d \log(t)} = \frac{dt}{d \log(t)} \cdot \frac{dP}{dt} \\
 &= t f(t) \\
 &= \sum a_i \tau_i^{-1} \exp(x - \tau_i^{-1} e^x)
 \end{aligned} \tag{6.7}$$

The second line here follows because dP/dt is simply the original distribution of time intervals, $f(t)$; it shows, oddly, that the distribution of $x = \log(t)$ can be expressed most simply not in terms of x , but in terms of t . When $f(t)$ is multi-exponential, as defined in (6.2), and we express $f_x(x)$ in terms of x by substituting $t = e^x$, we obtain the result in (6.7). This function is not exponential in shape, but is (for a single exponential component) a negatively skewed bell-shaped curve, the peak of which, very conveniently, occurs at $t = \tau$.

The same data, and the same fit, that was displayed in Fig. 10A,B,C, are shown in Fig. 10D as the distribution of $\log(\text{shut times})$. The same fitted curves are also shown, and the three component fit shows three peaks which occur at the values of the three time constants. It is now clearly visible, from a single graph, that the two-exponential fit is inadequate. (The slow component of the 2-exponential fit also illustrates the shape of the distribution for a single exponential, because it is so much slower than the fast component that the two components hardly overlap.) This sort of display is now universally used for multi-component distributions. Its only disadvantage is that it is hard in the absence of a fitted line, to judge the extent to which the distribution is exponential in shape.

Bursts of channel openings

It is often observed that several channel openings (a *burst* of openings) occur in rapid succession, the individual openings being separated only by brief shut periods, and that then a much longer shut period is seen before the next burst. This may occur spontaneously, or as a result of brief channel blockages. This phenomenon is evident in the shut time distribution shown in Fig. 10, from which it is clear that about 75% of shut times are very short (around 50 μs on average), and almost all the rest are much longer. The former are the ‘shut times within a burst’, and the latter are the long shut times that separate one burst from the next (‘shut times between bursts’).

Definition of bursts in practice. If some critical time, t_c , is defined, such that shut times shorter than t_c are deemed to be ‘within bursts’, then the experimental record can be divided into bursts (the end of each burst being signalled by occurrence of a shut time longer than t_c . Such a division can never be totally unambiguous when dealing with a random process, but various criteria exist for choosing an optimum value for t_c (see, for example, Colquhoun & Sakmann, 1985). However, as long as the ‘means’ of the long and short components differ by a factor of at least 50, and preferably over 100, the dangers of misclassification are acceptable. In the example illustrated in Fig. 10, this criterion is met, even when the small component of shut

times with a mean of around 1 ms is deemed to be ‘within bursts’. In this case $t_c = 5$ ms result in less than 2% of shut times being misclassified.

Distributions based on bursts. Once bursts have been defined, it is possible to define many new sorts of distribution, which can be helpful in the interpretation of single channel records (see Chapter 7). One reason for their usefulness is that one can usually be sure that all the openings in a burst come from the same individual channel, so shut times within bursts can be interpreted in terms of channel mechanisms, even under conditions where there is a large and unknown number of channels in the patch (so consecutive bursts may not originate from the same channel, and the shut time separating them is therefore not interpretable).

For example, the distribution of the burst length, or of the total open time per burst, can be defined. For many sorts of channel it is the mean burst length, rather than the mean length of the individual opening, that constitutes the ‘unitary event’ for physiological purposes (it would be irrelevant for the function of a synapse that the burst actually contained some very short closures within it). Another advantage of measuring quantities such as these becomes clear when we consider the effect of failing to detect brief shuntings; the burst lengths will be far less sensitive to such failures than the individual open times (see below, and Chapter 7).

It is expected that multi-exponential distributions will also fit distributions such as those of burst length, open time per burst and so on. Some other forms of distributions are also encountered. For example the distributions of the sum of any *fixed* number of exponentially distributed intervals is described by a *gamma distribution* (used, for example, by Colquhoun & Sakmann, 1985), though the distribution of the sum of a *random* number of exponentially distributed intervals is itself exponentially distributed (which is why the burst length, in some cases, has an approximately exponential distribution).

But we can also define a different sort of distribution on the basis of division of the record into bursts. For example, the *number of openings per burst* is a discrete variable (it can take only integer values, 1, 2, 3, . . . etc), rather than a continuous variable like time. Under the simplest (Markov) assumptions it is expected to be described by a mixture of *geometric distributions* (see also Chapter 7). The geometric distribution is the discrete analogue of the exponential distribution, and is described next.

Geometric distributions

This sort of discrete distribution can be exemplified by the distribution of the number of openings per burst. The probability, $P(r)$, that a burst will contain r openings is given by

$$P(r) = \sum a_i (1 - \rho_i) \rho_i^{r-1}, \quad r = 1, 2, 3, \dots, \infty \quad (6.8)$$

where a_i represents the area of each component, as before, and the ρ_i are constant coefficients (less than 1) that give the constant factor by which $P(r)$ is reduced each time r is increased by 1. The distribution therefore, apart from being discontinuous, has the same shape as an exponential distribution, as shown in Fig. 11.

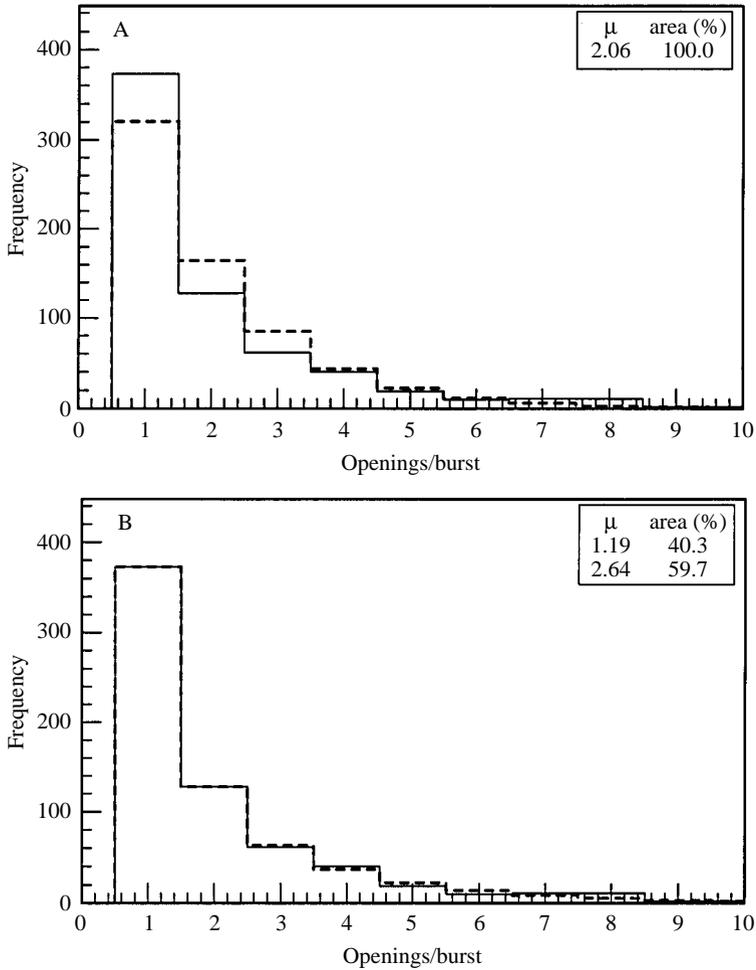


Fig. 11. Example of the distribution of the number of (apparent) openings per burst (frog muscle endplate, suberyldicholine 20 nM, -139 mV). Resolution was set as in Fig. 10; this resulted in 1355 resolved openings, and a critical shut time of 3 ms was used to divide the record into 659 bursts. The same data are shown in A and B. The data were fitted, by the method of maximum likelihood, with either a single geometric distribution (dashed line in A), or by a mixture of two geometric distributions (dashed line in B). The fitted parameter values are shown on the graphs. (D. Colquhoun and B. Sakmann, unpublished data.)

The mean number of openings per burst for each component, μ_i say, is related to the ρ_i values thus

$$\mu_i = \frac{1}{1 - \rho_i} \quad (6.9)$$

Fig. 11 shows the distribution of the number of openings per burst for similar data to that used for the shut time distribution in Fig. 10. A critical shut time of 3 ms was

used to divide the record into bursts. As discussed above, many of the shut times are too short to be resolved, so this distribution should preferably be referred to as ‘the number of *apparent openings* per burst’. If all shut times were detected there would be more shittings, and more openings, than are detected here.

In Fig. 11A the distribution has been fitted with a single geometric component, and the fit is not good (there are too many values with one opening per burst, and too few with 2 or 3 openings per burst, for a good fit). In Fig. 11B the same data has been fitted with two geometric components, and the fit is good.

The areas and means can be predicted from a specified kinetic mechanism, as described in Chapter 7, so, conversely the fitted values of these parameters can be used to estimate the rate constants in the underlying mechanism.

Stability plots

This section has dealt mainly with the display of measurements that have been made at equilibrium, so the average properties of the record are not changing with time. In practice it is quite common for changes to occur with time. This can be checked by constructing a *stability plot* as suggested by Weiss & Magleby (1989). In the case, for example, of the measured open times, the approach is to construct a moving average of open times, and to plot this average against time, or, more commonly, against the interval number (e.g. the number of the interval at the centre of the averaged values). A common procedure is to average 50 consecutive open times, and then increment the starting point by 25 (i.e. average open times 1 to 50, 26 to 75, 51 to 100 etc). The overlap between samples smooths the graph (and so also blurs detail). An exactly similar procedure can be followed for shut times, and for open probabilities. In the case of open probabilities, a value for P_{open} is calculated for every each set of 50 (or whatever number is chosen) open and shut times, as total open time over total length.

Fig. 12A shows examples of stability plots for open times, shut times and P_{open} which was calculated from the same experimental record (from frog muscle nicotinic channels) as that used for the shut time distribution shown in Fig. 10. It can be seen at once that all three quantities are reasonably stable throughout the recording. In contrast Fig. 12B shows similar plots for a recording from the NMDA type of glutamate receptor channel (Gibb & Colquhoun, 1991). In this case, though the open times are stable throughout the recording, there are two periods when the shut times suddenly become short, and P_{open} correspondingly increases to nearly 1.

Plots of this sort can be used to mark (e.g. by superimposing cursors on the plot) to mark sections of the data that are to be omitted from the analysis. For example, this approach has been used to inspect, separately, the channel properties when the channel is in a ‘high P_{open} period’, and when it is behaving ‘normally’.

It should be noted that when the average P_{open} value (the value for the whole record) is plotted on the stability plot, it can sometimes appear to be in the wrong position. This may happen when the record contains a very long shut period which reduces the overall P_{open} , but which affects only one point on the stability plot (which is normally constructed with ‘interval number’ on the abscissa, rather than time).

Amplitude stability plots. Exactly similar plots can be constructed for the channel

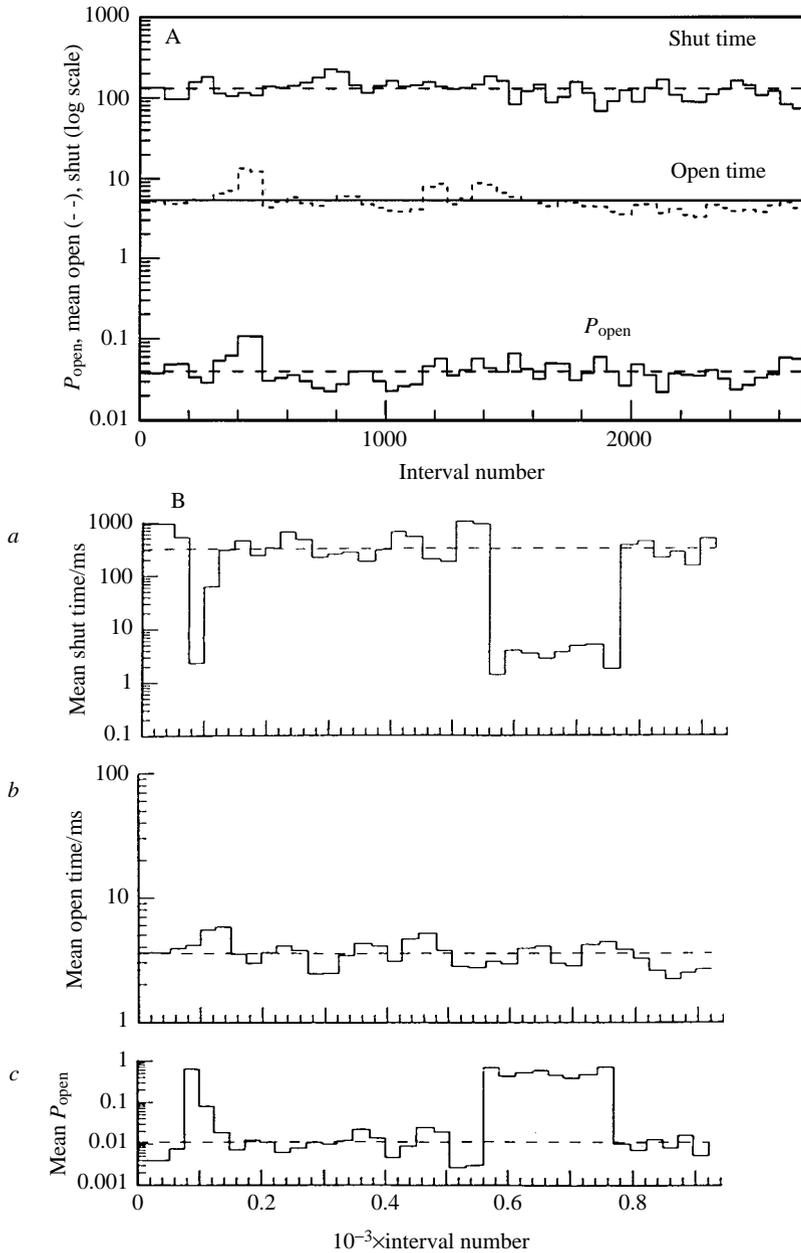


Fig. 12. Examples of stability plots. (A) The stability plots for shut times, open times and P_{open} are shown for the same data that were used for the shut time distribution in Fig. 10. All three plots are shown on the same graph by using a logarithmic ordinate. A running average of 50 values was calculated, the starting point being incremented by 25 values for each average. The overall average values of shut time, open time and P_{open} are plotted as horizontal lines. (B) Similar plots are shown (this time, as three separate graphs) for a recording from the NMDA-type glutamate receptor (outside-out patch from hippocampal CA1 cell, glutamate 20 nM + glycine 1 μ M, -60 mV; Gibb & Colquhoun, 1991). In this case a running average of 20 values, with an increment of 10 values, was used.

Amplitude stability plots. Exactly similar plots can be constructed for the channel amplitudes, in order to check whether they stay constant throughout the experiment.

7. The fitting of distributions

Fitting the results: empirical fits

At this stage we have a histogram that displays the experimentally-measured distribution of, for example, channel amplitudes, or open times, or number of openings per burst. The conventional approach is next to fit to these data a theoretical distribution (usually a mixture of Gaussian or exponential or geometric distributions, respectively). In the case of fitting exponentials, for example, the problem is how to find the values of the parameters (i.e. the ‘means’, τ_1 , τ_2 . . ., and the relative areas, a_1 , a_2 . . .) that provide the best fit the experimental data. The question of what mechanism might account for the observations is then considered (if at all) retrospectively.

Fitting a mechanism

In cases where a specific kinetic mechanism is being postulated for the ion channel it is possible to do better than this; all the data can be fitted simultaneously, the parameters being the underlying rate constants (as defined by the law of mass action) in the mechanism, rather than a set of empirical values of τ and a . This more sophisticated approach can be used only if allowance is made, *during* the fitting process, for events that are too short to be detected, and methods for doing this are discussed in Chapter 7.

Fitting exponentials

For the moment we shall consider only the first case, and the problem will be exemplified initially by the case where exponentials are to be fitted.

If a simple exponential will suffice then the traditional way of estimating the time constant, from the slope of a semilogarithmic plot of frequency against time will, if not optimum, be quite satisfactory in practice. In practice, hardly any observations can be fitted by a single exponential, so the problem gets a bit more difficult; the traditional method of ‘curve stripping’ will rarely be satisfactory and is, in any case, little or no faster than doing the job properly.

All satisfactory methods involve the minimization (or maximization) of some function, so the first thing that has to be done is to get hold of a general purpose computer program that can minimize a specified function. The easiest methods to use are based on simple search procedures; for example Patternsearch (see Colquhoun, 1971) or the Simplex method (Nelder & Mead, 1965; O’Neill, 1971; Hill, 1978). Search methods will usually converge even when give rather poor initial guesses (they are ‘robust’), and constraints can easily be incorporated into the fitting. However they are usually not as fast as more complex (but less robust) gradient methods. Most commercial subroutine libraries have a selection of minimization routines (e.g. the NAG PC library).

The next thing to do is to decide what is meant by ‘best’ fit. Two methods are commonly used. The first of these, the minimum chi-squared method, used the histogram frequencies as the data. This method is described, for example, by Colquhoun & Sigworth (1983); although it is quite satisfactory in practice it will not be discussed further here because it is generally believed that the *method of maximum likelihood* is preferable to other criteria for best fit. This method produces the values for the parameters ($\tau_1, \tau_2 \dots$, and $a_1, a_2 \dots$) which make the observation of our particular set of data more probable than it would be with any other parameter values (see Edwards, 1972; Colquhoun & Sigworth, 1983). The principle is easy. Suppose that we have a set of measurements (e.g. open times) denoted $t_1, t_2 \dots t_n$. If these observations are independent of each other then the probability (density) of making all n observations is proportional to the product of the separate probability (densities) for each observation, a quantity known as the likelihood (of a particular set of parameters), viz.

$$Lik = f(t_1)f(t_2) \dots f(t_n) \quad (7.1)$$

where f is the p.d.f., defined in (6.1) or (6.2) as appropriate. For example, if a single exponential as in (6.1) is sufficient then $f(t_1) = \tau^{-1}e^{-t_1/\tau}$, and so on. Most commonly we work with the log(likelihood), L , which from (7.1) can be written as

$$L = \log(Lik) = \sum_{i=1}^n \log f(t_i) \quad (7.2)$$

The optimization program finds the values of the parameters that make L as large as possible (these same values will, of course, also make Lik as large as possible); if the program is designed for minimization (as most are) we simply minimize $-L$ in order to maximize L . Notice that for the purpose of this calculation the observations are regarded as constants (the particular values of $t_1, t_2 \dots$ that we happen to have observed) while the parameter estimates are regarded as variables (the τ_i and a_i values are adjusted until L is maximized). The method is easily adapted to cope with the case that there are no observations below some specified minimum durations (e.g. the resolution), or above a specified maximum duration (see Colquhoun & Sigworth, 1983). Notice also that this method uses the original observed time intervals, *not* the histogram frequencies found from them. This has the advantages that (a) values of, say, 1.1 and 1.9 ms are not treated as though they were identical just because they happen to fall in the same histogram bin, and (b) the values found for the parameters will *not* depend on how we choose to divide up the observations to form a histogram. The histogram is still needed to display the fit once we have got it, but the fit is independent of how the bins are chosen. This adds considerably to the objectivity of the analysis.

How to superimpose a fitted curve on the histogram

The result of fitting two exponential components is shown in Fig. 13. In this case the data consisted of values of the *total open time per burst* for bursts of openings produced at the frog endplate by a very low concentration (4 nM) of suberyldicholine.

Under these conditions there are many brief openings. In this experiment the resolution was set to $60 \mu\text{s}$ for openings (and hence for the data in Fig. 13), and $40 \mu\text{s}$ for shut times.

Maximum likelihood fitting gave the faster time constant as $\tau_f = 0.157 \text{ ms}$, and the slower as $\tau_s = 22.8 \text{ ms}$; the corresponding areas were $a_f = 0.686$ (i.e. 68.6 percent of area), and $a_s = 0.314$ (31.4 percent of area). The fitted p.d.f. was thus

$$f(t) = w_f e^{-t/0.157} + w_s e^{-t/22.8}, \quad (7.3)$$

where $w_f = a_f \tau_f^{-1} = 4369.4 \text{ s}^{-1}$ and $w_s = a_s \tau_s^{-1} = 13.8 \text{ s}^{-1}$ are the *amplitudes* (at $t = 0$) of the components, and t is in milliseconds. All the data are shown in Fig. 13A, in a histogram with a bin width of $\Delta t = 4 \text{ ms}$. The estimated total number of observations (i.e. the number of bursts in this case) was $N = 279.7$. The area under the histogram is therefore $N\Delta t = 1.119 \text{ s}$, and the continuous curve plotted in Fig. 13A is $g(t) = 1.119f(t)$, as explained above, i.e.

$$g(t) = 4888.5 e^{-t/0.157} + 15.4 e^{-t/22.8} \quad (7.4)$$

The continuous curve fits the slow component of the histogram quite adequately in Fig. 13A. However the fit to the fast component in Fig. 13A appears poor at first sight;

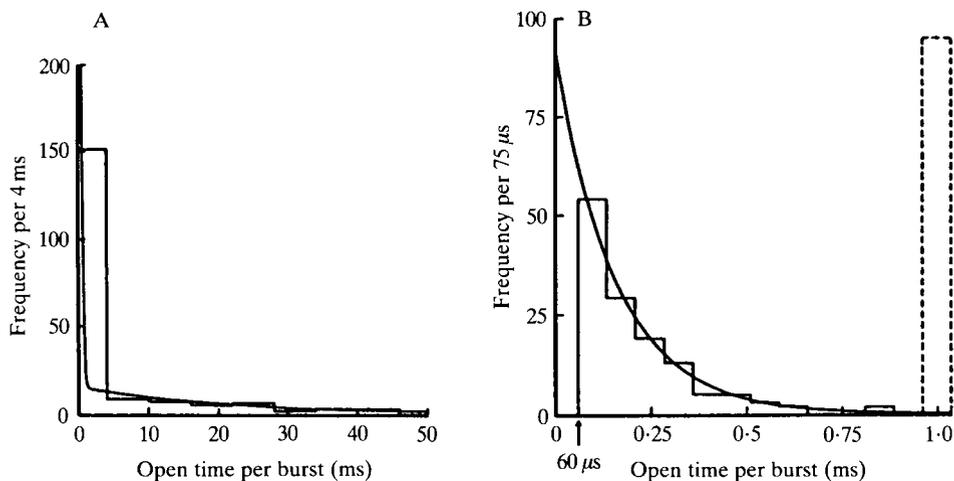


Fig. 13. Distribution of the total open time per burst (histogram) fitted, by the method of maximum likelihood, with a two-exponential probability density function (continuous line). The results are for frog endplate channels activated by suberyldicholine (4 nM, membrane potential, -188 mV). See text for details of the fitted parameters. The same data, and the same fitted curve, are shown in A and B on two different time scales. In A almost all of the data are shown (there were a few values longer than 50 ms). In B only values up to 1 ms are shown (the dashed bar represents all values greater than 1 ms). There are no observations below $60 \mu\text{s}$, the open time resolution, because this resolution was imposed on the data before the histogram was formed (see text). In B the ordinate of the fitted curve at $t = 0$ is 91.9; in this graph the bin width is $\Delta t = 75 \mu\text{s}$ so $N\Delta t = 0.0210 \text{ s}$ and $N\Delta t(a_f \tau_f^{-1} + a_s \tau_s^{-1}) = 91.9$; The area below $60 \mu\text{s}$ under the fitted curve suggests that there were 61.2 values that were too short to be seen. Modified from Colquhoun & Sakmann (1985) with permission.

the continuous curve does not pass centrally through the first bin, but is squashed up to the left hand side of it. In fact the fit is seen to be perfectly good when the *same* p.d.f., equation (7.3), is plotted on the expanded histogram in Fig. 13B; this shows data only up to 1 ms with 75 μ s bins, i.e. it shows mainly the fast components of the distribution. This example shows that although only one fit is done (resulting in equation (7.3)), the results either have to be displayed as two separate histograms (on different time scales), or displayed in the logarithmic manner described in Fig. 10D, in order to be able to judge visually the goodness of fit. In fact the fit to the leftmost bin in Fig. 13A is perfectly good (though this is certainly not obvious); it is the *areas* rather than the ordinates that must match, and the amplitude of the continuous curve at $t = 0$ is, from (7.4), $4888.5 + 15.4 = 4903.9$, i.e. it lies about 3 metres off the top of the page. The ordinate of the first histogram bin is 151 so its area is $151 \times 4 \text{ ms} = 604 \text{ ms}$. Now the open time resolution was 60 μ s so the first bin extends from 0.06 ms to 4.06 ms (N.B. *not* from 0 to 4 ms); the area under the continuous curve over this range is

$$\int_{0.06}^{4.06} g(t)dt = 524 + 56 = 580 \text{ ms}, \quad (7.5)$$

which is close to the observed bin area of 604 ms. Put another way, the continuous curve predicts a bin height of $580/4 = 145$, close to the observed value of 151.

Errors in the parameters

The best way of assessing the error in, for example, a value of τ is to repeat the experiment several times and observe how consistent the values turn out to be. If this cannot be done, or an internal estimate of error from a single experiment is thought to be desirable for some other reason, then there are various methods of calculating errors. The two that have been used in practice are (a) calculation of approximate standard deviations and correlations and (b) calculation of ‘support’ or ‘likelihood’ intervals. The details of the calculations are given by Colquhoun & Sigworth (1983); the discussion here will be limited to some comments on the use of these methods.

Approximate standard deviations of parameter estimates. These will give realistic estimates of error only when there are rather a lot of observations, and the parameters are well-determined (i.e. when precise estimates of error are not usually needed). The same calculations also give an estimate of the *correlations* between parameter estimates, and these can be rather useful. A strong positive correlation between two parameters suggests, for example, that the quality of the fit will be little changed if the value of one parameter is increased as long as the value of the other is also increased. This may mean that the *ratio* of the two parameters is better determined than the values of either of them separately.

Likelihood intervals. When the maximum likelihood method is used the parameter estimates found are the most likely values (i.e. those that make our data most probable). A pair of values for a parameter can be found (one above the most likely value and one below it), that are less likely by some fixed amount, and this pair of

values constitutes an interval within which we can reasonably expect the true value of the parameter to lie. For non-linear problems (such as fitting exponentials) we cannot attach an exact probability value to what we mean by 'reasonably'. However as a rough guide we can note that for a linear problem '0.5 unit interval' (values of the parameter for which L is 0.5 units below its maximum) would correspond to plus or minus one standard deviation, and a 2-unit interval would correspond to plus or minus two standard deviations.

The lower and upper likelihood limits will not generally be symmetrical about the best (maximum likelihood) value, and they are well-suited to expressing the range of possible values for rather ill-determined parameters. They have been found useful for this purpose in other curve-fitting problems also.

How many components are needed?

It is very often asked how one can decide whether a particular set of data requires, for example, three exponential components to fit it, or whether two components will suffice. A number of statistical methods (all approximate) have been proposed, to calculate 'whether the fit is significantly better' when an extra exponential is added (see, for example, Horn, 1987). In my view these methods are of very little value. The way to answer this question is to repeat the experiment several times, and each time fit the data with both two and three components, as illustrated in Fig. 10. In this example the third component, which had a mean of 1.3 ms, accounted for only 3.7% of the area under the distribution, so its reality might be doubted. However, many repetitions of this experiment revealed that a third component, *with approximately the same mean and area*, could be fitted in almost every case. It is this *consistency*, from one experiment to the next that provides convincing evidence that there are really three components. If in fact there were only two components, it would of course always be possible to fit the same data with three components, but the mean and area of the third component would be quite *inconsistent* from one experiment to the next, because we would not be fitting anything real, but just random noise in the data.

This approach shows, in another way, the undesirability of trying to decide the number of components by statistical tests on individual experiments. A small component such as that just illustrated, might well not reach 'statistical significance' in any one experiment, and so be missed altogether if one relied on such tests.

The example just discussed is interesting also because it shows that even a small extra component can have a surprisingly large effect on the fit of the major components. In Fig. 10 the mean for the fastest component was 56.7 μ s when only two components were fitted, but addition of the small third component changed this estimate to 45.5 μ s. Similarly, the mean for the component of long shut times was altered from 249 ms to 274 ms.

8. Correlations

The record of open and shut times that is used to fit distributions can also be used to calculate various correlation coefficients. For example one might be interested in the correlation coefficient between successive open times, successive burst lengths, or the lengths of successive openings within a burst. Such measurements can give information about the routes between various states of the system (see Chapter 7). When such correlations exist the distributions may *not* be the same for the first, second, etc. opening following a perturbation such as a voltage jump; care may therefore be needed in analysing such results. The display, and interpretation of correlations is considered in Chapter 7.

9. Transients: single channels after a voltage- or concentration-jump

So far the discussion has centred mostly around recordings made at equilibrium. Some new, and more complex, considerations arise when the record is not at equilibrium. For example, following a sudden change in concentration (a *concentration-jump*) or membrane potential (a *voltage-jump*), it will take some time for a new equilibrium to be established.

The reasons for wanting to measure the amplitude and duration of single channel currents after a jump include all the reasons already discussed, but such experiments also allow new sorts of question to be addressed.

A common motive for doing such jump experiments is simply to average a large number of responses to produce a more-or-less smooth curve. This average will be proportional to the probability that the channel is open as a function of time following the concentration or voltage step (though the absolute probability can be found only if it is known how many channels are active in the patch). It should, therefore, have the same shape as the macroscopic current relaxation found, for example, by the whole-cell clamp method. The big advantage of doing the experiment this way is that the individual single channel currents that underlie the macroscopic responses can be seen, and may be identifiable as a particular sort of channel. Thus the channels that carry the macroscopic current can be identified, with much greater certainty than could be done from the macroscopic current alone.

Another motive for measuring channel openings after a jump is to cast light on the channel mechanism. The sort of information about mechanisms that can be obtained from such experiments is different from, and complementary to, that which can be obtained from equilibrium measurements. Some of the principles involved are discussed in Chapter 7. For example, measurements of the *first latency* (the time from the moment of the jump to the first channel opening) can give valuable information about kinetic mechanisms that is not obtainable from equilibrium measurements.

First latency measurements can also help to explain the time course of synaptic currents (e.g. Chapter 7, and Edmonds & Colquhoun, 1992).

In the case, particularly, of voltage-activated channels it has been common practice to investigate *only* voltage jumps. But, because the jumps are relatively brief, this precludes the measurement of any slow kinetic processes (which may consequently, and usually unnecessarily, be referred to as 'modes').

If correlations are present (section 8, above) the distributions of the 1st, 2nd, 3rd, . . . open times (shut times, or burst lengths etc) after the jump may not be the same (e.g. Colquhoun & Hawkes, 1987). Information from this source has yet to be exploited experimentally.

10. Effects of limited time resolution

It is very often true that some openings and shittings are too short to be detected, and their omission will obviously cause errors. It is clear, for example, from Fig. 10 that many brief shittings have not been detected. The three component fit in this case gave a mean of 45.5 μs for the fast component, but no shittings shorter than 60 μs could be detected with confidence in this experiment. It follows from (6.4) that about 73% of the short openings were missed, and only 27% were detected and measured.

Corrections for missed events

When only *either* openings *or* shittings (but not both) are short enough to be missed in substantial numbers, approximate corrections can be made, without having to know about the details of the channel mechanism. Take, for example, the case where most openings are long enough to be detected but substantial numbers of brief shut periods (gaps) are missed. In this case it is the *open* time distributions which will be in serious error because two openings separated by an unresolvable gap will be counted as a single opening. The shut time distributions will be accurate in the region where they can be measured, i.e. for those gaps that are long enough to be detected. The shut time distribution can therefore be extrapolated to zero time to get an estimate of the number of shut times that have escaped detection; this will of course work only if a sufficient number of the short shittings (the ones that are longer than the resolution) are detected to allow accurate extrapolation. A corrected mean open time can then be found by taking the total observed 'open' time (minus the short time spent in undetected gaps) and dividing it by the total number of openings, i.e. by the corrected total number of gaps (those seen plus those undetected). Of course only the *mean* open time can be so corrected; the true *distribution* of open times cannot be found. If the open time distribution itself has two components, then we cannot say to what extent the inferred missed shut times were missed from 'short openings' or from 'long openings'. In order to say anything about this sort of question we must make some postulate concerning the underlying channel mechanism (see Chapter 7).

If the data contain openings and shut periods that are so short that substantial

numbers of *both* are missed then there is no way of correcting the results without making some postulate about the detailed channel mechanism. When this can be done, methods for allowing for missed events have been developed, and these will be discussed in Chapter 7.

Details of the simpler corrections mentioned above, and the modifications that are needed when openings occur in bursts, are given, for example, by Ogden & Colquhoun (1985) and by Colquhoun & Sakmann (1985). When possible it is a good idea to work with values that are not sensitive to loss of brief events, such as the burst length (or, as in Fig. 13, the total open time per burst).

Minimizing the problem of missed events

In view of the problems surrounding corrections for missed events, it is desirable to circumvent the problem, whenever it is possible, by making measurements that are insensitive to the event omission. Some examples of such measurements follow.

(1) Measurements of burst length are clearly less sensitive than measurements of open or shut times (e.g. failure to detect all the brief shittings within a burst will have little effect on the measurement of its overall length).

(2) One motive for looking at the distribution of open times is that the number of components in this distribution is, in principle, equal to the number of different open states. However the same is true of the distribution of the total open time per burst. If there are many brief shittings, the latter distribution can be measured more precisely, and so should be used in preference to the distribution of (apparent) open times.

(3) As was pointed out at the end of section 5, if P_{open} values are measured by integration of the experimental record, rather than by measurement of individual open and shut times, the errors resulting from missed events are largely eliminated.

References

- ABRAMOVITZ, M. & STEGUN, I. A. (1965). *Handbook of Mathematical Functions*. Dover Publications, New York.
- CHUNG, S. H., MOORE, J. B., XIA, L., PREMKUMAR, L. S. & GAGE, P. W. (1990). Characterization of single channel currents using digital signal processing techniques based on hidden Markov models. *Phil. Trans. Roy. Soc. London B* **329**, 265-285.
- COLQUHOUN, D. (1971). *Lectures on Biostatistics*. Oxford: Clarendon Press.
- COLQUHOUN, D. & HAWKES, A. G. (1983). The principles of the stochastic interpretation of ion channel mechanisms. In *Single Channel Recording* (ed. B. Sakmann & E. Neher). New York: Plenum Press.
- COLQUHOUN, D. & HAWKES, A. G. (1987). A note on correlations in single ion channel records. *Proc. Roy. Soc. London B* **230**, 15-52.
- COLQUHOUN, D. & OGDEN, D. C. (1988). Activation of ion channels in the frog end-plate by high concentrations of acetylcholine. *J. Physiol.* **395**, 131-159.
- COLQUHOUN, D. & SAKMANN, B. (1983). Bursts of openings in transmitter-activated ion channels. In *Single Channel Recording* (ed. B. Sakmann & E. Neher). New York: Plenum Press.
- COLQUHOUN, D. & SAKMANN, B. (1985). Fast events in single-channel currents activated by acetylcholine and its analogues at the frog muscle end-plate. *J. Physiol.* **369**, 501-557.
- COLQUHOUN, D. & SIGWORTH, F. L. (1983). Fitting and statistical analysis of single channel records. In *Single Channel Recording* (ed. B. Sakmann & E. Neher). New York: Plenum Press.

- EDMONDS, B. & COLQUHOUN, D. (1992). Rapid decay of averaged single-channel NMDA receptor activations recorded at low agonist concentration. *Proc. Roy. Soc. London B* **250**, 279-286.
- EDWARDS, A. W. F. (1972). *Likelihood*. Cambridge University Press.
- FREDKIN, D. R. AND RICE, J. A. (1992). Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. Roy. Soc. London B* **249**, 125-132.
- GIBB, A. J. AND COLQUHOUN, D. (1991). Glutamate activation of a single NMDA receptor-channel produces a cluster of channel openings. *Proc. Roy. Soc. London B* **243**, 39-45.
- HAMILL, O. P., MARTY, A., NEHER, E., SAKMANN, B. & SIGWORTH, F. J. (1981). Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archiv*. **391**, 85-100.
- HILL, I. D. (1978). A remark on algorithm AS47. Function minimization using a simplex procedure. *Appl. Statist.* **27**, 380-382.
- HORN, R. (1987). Statistical methods for model discrimination. Applications to gating kinetics and permeation of the acetylcholine receptor channel. *Biophys. J.* **51**, 255-263.
- HORN, R. (1991). Estimating the number of channels in patch recordings. *Biophys. J.* **60**, 433-439.
- McMANUS, O. B., BLATZ, A. L. & MAGLEBY, K. L. (1987). Sampling, log-binning, fitting and plotting durations of open and shut intervals from single channels and the effects of noise. *Pflügers Arch.* **410**, 530-553.
- NELDER, J. A. & MEAD, R. (1965). A simplex method for function minimization. *Computer J.* **7**, 308-313.
- OGDEN, D. C. & COLQUHOUN, D. (1985). Ion channel block by acetylcholine, carbachol and suberyldicholine at the frog neuromuscular junction. *Proc. Roy. Soc. London B* **225**, 329-355.
- O'NEILL, R. (1971). Algorithm AS47. Function minimization using a simplex procedure. *Appl. Statist.* **20**, 338-345.
- PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1986). *Numerical Recipes*. Cambridge University Press.
- SAKMANN, B., PATLAK, J. & NEHER, E. (1980). Single acetylcholine-activated channels show burst-kinetics in presence of desensitizing concentrations of agonist. *Nature* **286**, 71-73.
- SIGWORTH, F. J. & SINE, S. M. (1987). Data transformations for improved display and fitting of single-channel dwell time histograms. *Biophys. J.* **52**, 1047-1054.
- STANDEN, N. B., STANFIELD, P. R. & WARD, T. A. (1985). Properties of single potassium channels in vesicles formed from the sarcolemma of frog skeletal muscle. *J. Physiol.* **364**, 339-358.
- WEISS, D. S. & MAGLEBY, K. L. (1989). Gating scheme for single GABA-activated Cl-channels determined from stability plots, dwell-time distributions, and adjacent-interval durations. *J. Neurosci.* **9**, 1314-1324.