DATA ANALYSIS

This chapter provides guidelines to selecting certain key data acquisition parameters, and discusses current methods used in analysis of biological data, including fitting models to data and analyzing single ion-channel recordings. The scope of this discussion is limited to brief, practical introductions with references to more detailed discussion in this **Guide** and in the literature. The manuals for the various software programs can be used for more specific step-by-step instructions, and the text by Dempster (1993) is a useful collection of standard techniques in this field.

Choosing Appropriate Acquisition Parameters

Successful analysis of a data set requires that the signals be recorded using appropriate data acquisition parameters. Following are guidelines to selecting several critical recording parameters.

Gain and Offset

Sufficient gain and offset should be maintained on all transducers and amplifiers during data acquisition, because data review and analysis software can usually provide only limited amounts of additional ("software") offset and gain. More than 10x of software gain usually results in excessively jagged traces; this is because the digitized signal does not vary smoothly but is quantized (see **Chapter 9**), and too much amplification allows the difference between one quantization level and the next to become visible. Software offset is limited to the full range of the analog-to-digital converter (ADC) input, which is usually equivalent to about ±10 V referred to the input of the ADC. Signal resolution is best preserved when the signal fills this voltage range without exceeding it.

Sampling Rate

The sampling rate used should be selected considering the particular experiment. Use of excessively high sampling rates wastes disk space and will increase the time required for analysis. Furthermore, a higher sampling rate is usually associated with a higher filtering frequency, which in turn allows a larger amount of noise to contaminate the signal. Subsequent

analysis of the data may therefore require noise reduction using analysis-time filtering, which can be time-consuming. Guidelines to choosing the correct sampling rate are discussed in the following paragraphs (Colquhoun and Sigworth, 1983, and Ogden, 1987).

Biological signals are most commonly analyzed in the time domain. This means that the *time dependence* of the signals is examined, *e.g.* to characterize the membrane response to a voltage-clamp pulse. The usual rule for time-domain signals is that each channel should be sampled at a frequency between 5 and 10 times its data bandwidth. Knowing the value of the data bandwidth is required in order to set the filter cut-off frequency during acquisition and analysis.

For a sinusoidal waveform, the data bandwidth is the frequency of the sine itself. For most biological signals, the data bandwidth is the highest frequency of biological information of interest present in the recorded signal. This can be determined directly by examining a power spectrum of rapidly sampled unfiltered data, though this is rarely done. Alternatively, one can estimate the number of points per time interval required to give a data record whose points can be easily "connected" by eye and calculate the sampling rate directly. The data bandwidth and filter frequency can then be calculated from the sampling rate. For example, if a fast action potential (1 ms to peak) is to be recorded, 25 samples on the rising phase would yield a reasonably good 40 µs resolution, requiring a sampling rate of 25 kHz and an approximate data bandwidth of 5 kHz.

The rules are more straightforward in some special cases. Single-channel recording is discussed below. For signals with exponential relaxation phases, the sampling rate needed to estimate a time constant depends on the amount of noise present; for moderately noise-free data, at least 15 points should be taken per time constant over a period of 4 to 5 time constants. Many fitting routines will fail if sampling is performed over only 3 time constants, since the waveform does not relax sufficiently far towards the baseline. For a sum of multiple exponentials, the sampling rate is determined in this way from the fastest phase; sampling must extend to 4 time constants of the slowest phase. If this would result in too many samples, a split clock (as in the program CLAMPEX of Axon Instruments' pCLAMP suite) or other methods of slowing the acquisition rate during the acquisition, could be employed as long as at least 15 points are taken over each time constant.

When a set of several channels is recorded (*e.g.*, channels 0 through 3), most data acquisition systems sample the channels sequentially rather than simultaneously. This is because the system usually has only one analog-to-digital converter circuit that must be shared among the channels in the set. For example, if four channels are sampled at 10 kHz per channel, one might expect that they would be sampled simultaneously at 0 µs, 100 µs, 200 µs, etc. Instead, channel 0 is sampled at 0 µs, channel 1 at 25 µs, channel 2 at 50 µs, channel 3 at 75 µs, channel 0 again at 100 µs, channel 1 again at 125 µs, etc. There is therefore a small *time skew* between the channels; if this causes difficulties in analysis or interpretation, a higher sampling rate can be used to minimize the time skew (but this may cause problems associated with high sampling rates, as mentioned above).

An additional consideration arises from the fact that on many data acquisition systems, including the Digidata 1200 from Axon Instruments, the digital-to-analog converter (DAC) is updated whenever the ADC is read, even if there is no change in the DAC output. This means that the DAC is updated only at the sample rate over *all* channels. For example, if a stimulus is a 0 to

150 mV ramp and 50 samples are acquired from one channel at a sampling interval of 25 µs, the DAC output will appear as a series of steps each 25 µs long followed by an upward jump of 150 mV/50 = 3 mV, which may be too large for some electrophysiological applications. Therefore, if a rapidly changing continuous waveform is applied while acquiring slowly, the output waveform should be checked with an oscilloscope and, if necessary, the sampling interval should be increased. The computer preview of a waveform cannot be relied upon for this purpose because it does not account for the effect of sampling. Note, however, that since most users acquire significantly more samples per sweep than 50, this problem will not occur except in very unusual situations.

Filtering

The signal should be filtered using an analog filter device before it arrives at the ADC. As discussed in **Chapter 6** and in Colquhoun and Sigworth, 1983 and Ogden, 1987, this is done to prevent aliasing (folding) of high-frequency signal and noise components to the lower frequencies of biological relevance.

Acquisition-time filtering of time-domain signals is usually performed using a Bessel filter with the cut-off frequency (-3 dB point; see **Chapter 6**) set to the desired value of the data bandwidth. A 4-pole filter is usually sufficient unless excessive higher frequency noise requires the 6- or 8-pole version. The Bessel filter minimizes both the overshoot (ringing) and the dependence of response lag on frequency. The latter two effects are properties of the Chebyshev and Butterworth filters (see **Chapters 6** and **12**, or Ogden, 1987), which are less appropriate for time-domain analysis.

Filtering at Analysis Time

It is sometimes reasonable to sample data at higher rates than seems necessary, *e.g.*, when a greater bandwidth might be required during analysis. If excessive sample rates are used, the filter frequency must be set to a higher value. Since there is more noise at higher frequency, more noise is likely to contaminate the signal. Therefore, the data must be filtered further during analysis in order to reduce the noise and avoid aliasing of high-frequency signal content.

This analysis-time filtering is performed using filters implemented in software. The Gaussian filter is most commonly used for this purpose because of its execution speed, though the Bessel filter is employed as well. If one wants to write a computer program for a filter, the Gaussian is easier to implement than the Bessel filter (see program example in Colquhoun and Sigworth, 1983). Note that all filters alter the waveform; for example, a Gaussian-filtered step function deviates from the baseline *before* the time of transition of the unfiltered signal. The user can examine data records filtered at different frequencies to make sure that there is no significant distortion of the quantities of interest, such as time of transition or time to peak.

Another common software filter is *smoothing*, the replacement of a data point by a simply weighted average of neighboring points, used to improve the smoothness of the data in a record or graph. In contrast to the smoothing filter, the Bessel and Gaussian types have well-known filtering (transfer) functions, so that (*i*) it is easy to specify an effective cut-off frequency, and (*ii*) the effect of the filter may be compensated for by a mathematical procedure analogous to a high-frequency boost circuit in a voltage-clamp amplifier (Sachs, 1983). These advantages are

important if the frequency properties must be known throughout the analysis. If not, the smoothing filter is much faster to execute, especially on long data records, and easier to implement if one writes one's own software.

Integrals and Derivatives

The integral function is used to calculate the area under a digitized data trace. Applications include measuring the total charge transfer from records of membrane current, and measuring the area under a miniature endplate potential. The integral is generally calculated by direct summation of $y(x_i)\Delta x_i$ between two cursors placed along the x axis. For a data record, $y(x_i)$ is the amplitude at time point x_i, and Δx_i is the sampling interval at that time. For a histogram, $y(x_i)$ is the amplitude of the bin at location x_i, and Δx_i is the bin width of bin i (this allows for nonuniform bin width). The y values must be corrected for any superfluous baseline before integration using a fixed or slanting baseline, as appropriate. If the sample rate were low compared to the rate of change of the signal, so that y values show large changes from one x_i to the next, the integral is considerably less accurate than if significantly more samples were taken. Large inaccuracies can also occur if the signal does not return to baseline by the end of the data set or if part of the data is corrupted by an extraneous signal. In some cases, errors in the integral can be reduced by first fitting a smooth curve to the available data and then using the formula and parameters of the best fit to calculate the integral.

The derivative function is used to determine the rates of change of a digitized signal. This can be used to help in peak location (where the derivative will be near zero), transition detection (large positive or negative derivative), sudden departure from baseline, etc. The derivative will, however, amplify noise, making it difficult to determine trends. The signal should therefore be filtered beforehand or fit to a function which can then be used to obtain a noise-free form of the derivative. However, the worse the fit, the greater the error in the derivative.

Single-Channel Analysis

Goals and Methods

The goal of single-channel current analysis is to reconstruct the idealized current waveforms from which information about the mechanisms of channel function is derived. Specific information can be deduced with respect to channel state diagrams, kinetics of channel opening, closing and gating, channel barrier models, and the effects of channel blocking agents and membrane constituent on channel function.

Articles that present approaches and methods used in single-channel current analysis include Colquhoun and Hawkes, 1983; Colquhoun and Sigworth, 1983; McManus, Blatz and Magleby, 1987; Sigworth and Sine, 1987; and French *et al*., 1990.

The current from a single ion channel is idealized as a rectangular waveform, with a baseline current of zero (closed-channel), an open-channel current dependent on the conductance and membrane potential, and rapid transitions between these current levels. In practice, this idealized current is distorted by the limited bandwidth of the apparatus and contaminated by noise. Shifts

in the baseline may occur during the recordings, and capacitative current spikes are present in sweeps in which voltage changes are applied across the membrane. The current signal is further altered by filtering and by periodic sampling. These effects can impede making confident inferences from data regarding channel behavior.

Sampling at Acquisition Time

In order to adequately reconstruct the transitions, the sampling rate should be no less than 5 times higher than the cut-off frequency of the filter (Colquhoun and Sigworth, 1983; French *et al*., 1990). An even higher sampling rate is needed if interpolation of the data record is used to increase the accuracy of transition detection (unavailable in the pCLAMP single-channel analysis program).

Filtering at Acquisition Time

As discussed above, the input signal should be filtered with a Bessel filter during acquisition. The cut-off frequency must usually be set sufficiently low to prevent unacceptably frequent occurrences of noise, which cause false closings and false brief open-close events during the construction of idealized channel currents. French *et al*. (1990) discuss how to decide on the cut-off frequency for a particular situation: open events of longer duration are more likely to be falsely closed by noise, so it is convenient to specify the longest duration d_{max} that can be recorded with less than 1% occurrence of these false closings. Once d_{max} has been specified, the required cut-off frequency f_c can be calculated using the following combination of equations (18) and (19) in French *et al*. (1990):

$$f_c = \frac{100 f_c^*}{d_{max} FTC^*} \qquad (1)$$

Here FTC^* is the observed rate of false threshold crossings measured using recordings made with an arbitrary cut-off frequency f_C^*. FTC^* can be measured from idealized single-channel records generated using a threshold set on the side of the baseline opposite to where transitions are observed. This analysis can be achieved using the FETCHAN program in the pCLAMP suite.

Analysis-Time Filtering

The digital Gaussian filter can be used for additional analysis-time filtering of single-channel records. This introduces symmetrical time delays at both opening and closing and can therefore be used for the unbiased estimation of latencies using the 50% criterion (see below).

Generating the Events List

The first step during analysis of single-channel current records is to idealize the current records to a series of noise-free open and closed states having infinitely short (or at least shorter than the sample interval) transition times. This analysis, which can be performed by FETCHAN, results in a list of durations and amplitudes, called the *events list*. Several effects that tend to complicate this reconstruction are briefly discussed below; more thorough discussions are presented in the cited literature.

Setting the Threshold for a Transition

A transition between states of a channel occurs when the current amplitude passes through a threshold between two levels. The most commonly used threshold is 50% of the difference between the levels. The advantage of this threshold setting is that the event durations are not biased because the values of the threshold are the same for both opening and closing transitions.

Baseline Definition

The accuracy of the threshold method for transition detection depends on the stability of the baseline (*i.e.*, closed-channel) current or, if the baseline is unstable, on the ability of an automated procedure to correctly identify the baseline as it changes. A number of ways have been devised to track a moving baseline, including (1) averaging the baseline current level to get the new baseline; (2) defining the new baseline as that level which maximizes the number of times that the current signal crosses it during the closed state ("zero crossings" method; Sachs, 1983); and (3) defining the new baseline at the peak of the histogram of the baseline current amplitude (*e.g.*, G. Yellen, quoted in Sachs, 1983). FETCHAN uses a hybrid approach in which the average of the most recent closed channel current level is averaged with the old baseline level, weighted by a selectable factor. Regardless of the method used, the user must carefully monitor the baseline to ensure that any automatic procedure does not lose track of the baseline value.

Missed Events

Events will be missed if their true durations are shorter than the dead time of the system, which is determined by the filter cut-off frequencies used during acquisition and analysis. Events will also be missed if their superthreshold portions happen to miss the times when the signal is sampled even if their durations are longer than this dead time. The resulting error is minimal if the fastest time constant in the system is much longer than the sampling interval because few events will be shorter than the sampling interval. If this is not the case, the sampling rate must be increased with respect to the filter cut-off frequency (see the relevant sections above).

False Events

The probability of detecting false events depends on the amount and spectrum of noise in the system, the filter characteristics and the rate of sampling (French *et al*., 1990).

Multiple Channels

The presence of multiple channels complicates the determination of the kinetic behavior of a channel. If a record shows a transition from two open channels to one open, it cannot be determined if the transition was due to the closing of the first or the second channel. A number of methods have been proposed to deal with this ambiguity (French *et al*., 1990). As a precaution, the amplitude histogram of the raw data can be inspected to determine if multiple channels are present.

Analyzing the Events List

The second step in the analysis procedure is the extraction of model-specific information from the events list. Because the number of events is usually large, it is often convenient to sort the event data into bins. The binned data may then be fit to a mathematical function whose parameters are related to a state diagram of the channel. The most common histograms include

(1) the *dwell time histogram*, in which the duration of events is binned and which can be related to state models; (2) the *first latency histogram*, in which the period of time from stimulus onset to first opening is binned and which is used to extract kinetic information; and (3) the *amplitude histogram*, in which the amplitudes of the levels or of all points in the records are binned and yield information about the conductance, the system noise and the incidence of multiple channels.

Histograms

The simplest histogram is one in which a series of bins are defined, each of which has an associated upper and lower limit for the quantity of interest, *e.g.*, dwell time. Each dwell time will then fall into one of the bins, and the count in the appropriate bin is incremented by one. In the *cumulative* histogram, a bin contains the number of observations whose values are less than or equal to the upper limit for the bin.

Histogram Abscissa Scaling

Histograms may be scaled and displayed in a number of ways. The most common histogram has constant bin width, a linearly scaled abscissa (x axis), and counts displayed on the ordinate (y axis). A number of alternatives have been developed to improve the treatment of exponential distributions, in which low-time bins have many counts and high-time bins have few, yielding histograms that are difficult to visualize (*e.g.*, McManus *et al*., 1987; Sigworth and Sine, 1987). One solution is to increase the bin width logarithmically, so that for a single exponential, a constant number of events is expected per bin. This has the disadvantage that the presence of multiple time constants is not evident from the plot if the ordinate is not rescaled. These issues are discussed in the next section.

Histogram Ordinate Scaling

If a constant bin width histogram is used, linear scaling of the ordinate is the most common (Figure 1*A*). If a logarithmic bin width is used, several methods are common. In one of the methods, the number of counts in each bin is divided by the bin width and plotted on a logarithmic axis, in which case multiple time constants are evident (Figure 1*B*). The main difficulty with this scaling is that the bins do not have the same weight during the fit process (see below), because the variance of the bin depends on the number of counts it contains. This can be compensated for using the transformation of Sigworth and Sine (1987). In this transformation the bin width is constant and the square root of the number of counts per bin is plotted on the ordinate, with time plotted logarithmically on the abscissa (Figure 1*C*).

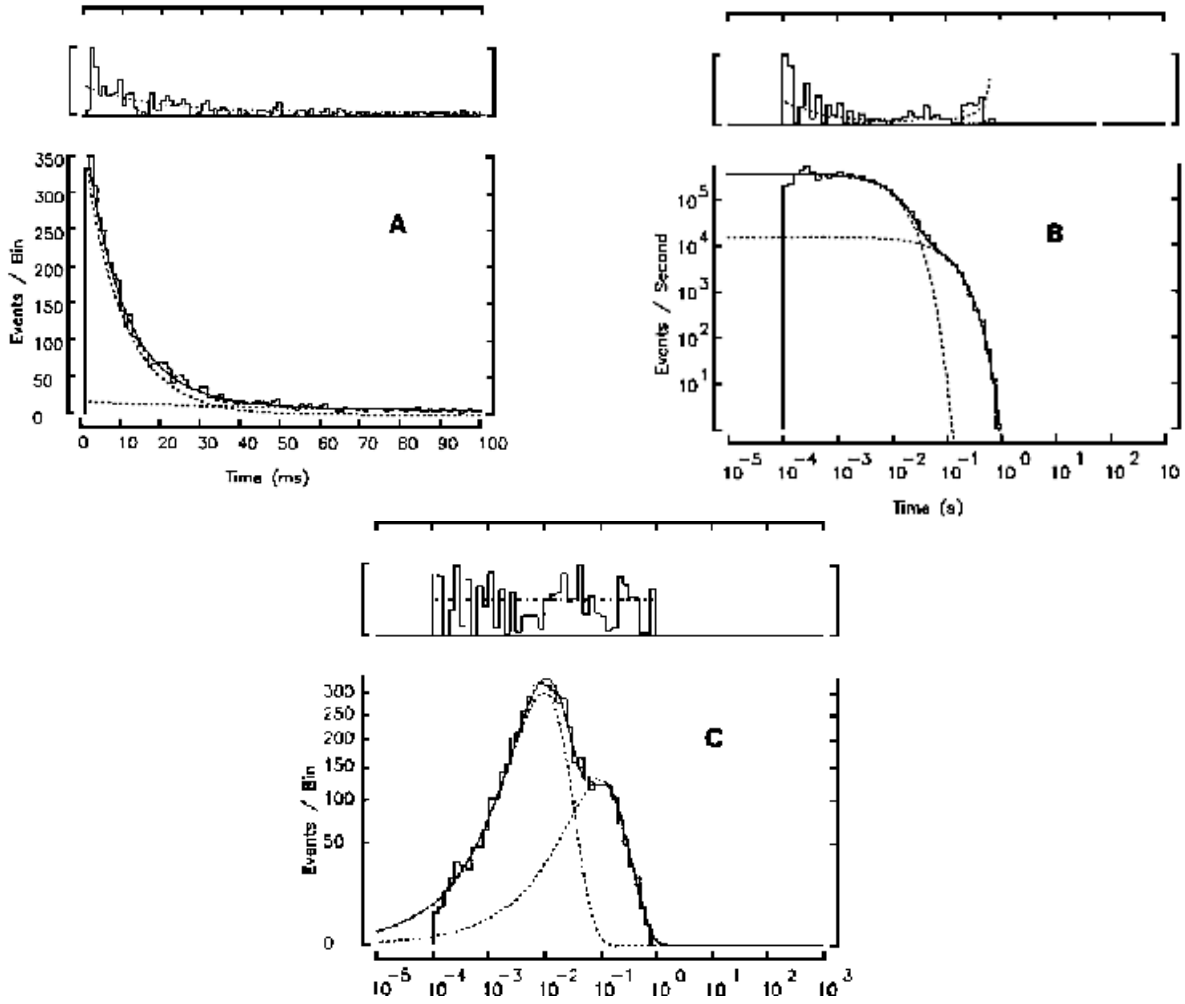

Figure 10-1. Histogram Scaling

[Three] representations of a dwell-time distribution with two exponential components. 5,120 random numbers were generated according to a distribution with time constants of 10 ms (70% of the events) and 100 ms (30%) and binned for display as histograms in the lower panel of each part of the figure. Superimposed are the theoretical probability density functions for each component *(dashed curves)* and their sum *(continuous curve).* In each part of the figure the upper panel plots the absolute value of the deviation of the height of each bin from the theoretical curve, with dashed curves showing the expectation value of the standard deviation for each bin. The upper panels were plotted with vertical expansion factors of 2.1, 5.4, 3.1, and 4.9, respectively. (*A*) Linear histogram. Events are collected into bins of 1 ms width and plotted on a linear scale. The 100-ms component has a very small amplitude in this plot. (*B*) Log-log display with variable-width (logarithmic) binning. The number of entries in each bin is divided by the bin width to obtain a probability density in events/s which is plotted on the ordinate... (*C*) Square-root ordinate display of a logarithmic histogram. Note that the scatter about the theoretical curve is constant throughout the display *(reproduced with permission from Sigworth and Sine, 1987).*

Errors Resulting from Histogramming Events Data

Problems may appear as a result of the histogram process. The first problem may occur in either amplitude or time histograms if the bin size is not an integral multiple of the resolution of the signal. For a dwell time, the resolution is the interval between successive samplings of the ADC channel; for an amplitude, the resolution is determined by the gain and the ADC. The symptom is that occasional or periodic peaks or valleys appear artifactually in the histogram. This problem is less severe if the bin width corresponds to many resolution intervals (*e.g.,* 10). If bin widths are variable, the smallest bin width must likewise include many resolution intervals.

A second problem is called *sampling promotion error*. Sampling promotion error (Sine and Steinbach, 1986) occurs because data are sampled periodically. Suppose data were acquired with 1 ms sampling rate and dwell times binned into a histogram with bin width equal to the sampling rate (1 ms) and centered around 3 ms, 4 ms, 5 ms, etc. The 4 ms bin would therefore contain events whose true dwell times lie between 3 and 5 ms (Figure 2). If the dwell times fall exponentially with increasing times, the 4 ms bin would contain more events from 3 to 4 ms than from 4 to 5. The subsequent fit would treat all these events as if they had occurred at 4 ms (the midpoint), thereby resulting in an error. In a single exponential fit, this affects only the intercept and not the time constant; but there may be more subtle effects in a multi-exponential fit. The error is small when the bin width is much smaller (perhaps by a factor of 5) than the fastest time constant present.

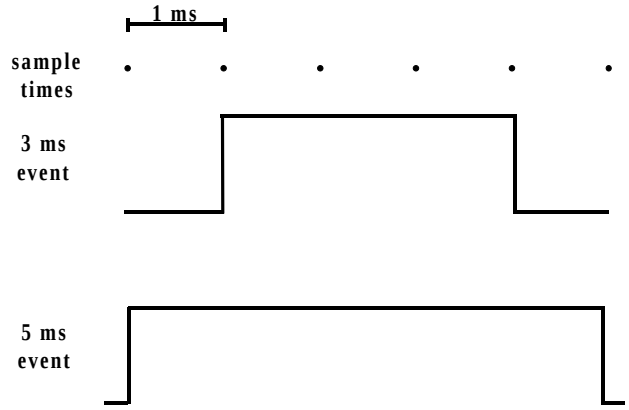

Figure 10-2. Sampling Error

This figure illustrates how a 1 ms sampling rate can result in an apparent 4 ms open time for events of between 3 and 5 ms. The dots in the top part of the figure represent times when the signal is sampled. The lower part of the Figure shows two events whose waveforms are high at 4 sample times.

A third problem is termed *binning promotion error*. In an exponentially falling dwell time distribution, a bin is likely to contain more events whose true dwell times are at the left side of the bin than are at the right side. The average dwell time of the events in that bin is therefore less than the midpoint time *t*. The error occurs when a fit procedure assumes that all the bin events are concentrated at a single point located at the midpoint *t*, instead of at the average, which is less than *t*. Binning promotion error can occur in addition to sampling promotion error because binning takes place in a different step. Both of these errors are due to the asymmetric distribution of true dwell times about the bin midpoint. A correction procedure has been proposed by McManus et al. (1987).

The degree of these bin-width-related errors may be reduced independently of the corrective procedures mentioned above, if the fit procedure explicitly uses the fact that a bin actually represents an area instead of an amplitude. Some errors may be eliminated if the individual data points are used without ever using a histogram, as in the maximum likelihood fitting method.

Lastly, as discussed in the section on missed events, short events may not be detected by the 50% threshold criterion. This can give rise to a number of errors in the extracted fit parameters, which relate specifically to state models. For further details, consult the references cited in French *et al*., 1990.

Amplitude Histogram

Amplitude histograms can be used to define the conductances of states of single channels. Two kinds of amplitude histograms are common: *point histograms* and *level histograms*. The former use all the acquired data points; they are useful mainly for examining how "well-behaved" is a data set. Abnormally wide distributions may result from high noise if the signal were over-filtered, or if baseline drift had been significant. Similarly, the number of peaks will indicate the presence of multiple channels or subconductance states. The all-points histogram will probably not be useful for determining the conductances unless baseline drift is small. Level histograms use only the mean baseline-corrected amplitudes associated with each event in the events list. Such histograms can be fitted to the sum of one or more Gaussian functions in order to estimate conductances.

Fitting to Histograms

Amplitude and dwell time histograms can be fitted by appropriate functions, usually sums of several Gaussian functions for the former and sums of exponentials for the latter (see the section on Fitting below). The time constants and amplitudes can be related to the parameters of several single-channel models, but this will not be described here.

Histogram bins containing zero counts should usually be excluded from a fit because the chi-square function (see below) is not defined when a bin i contains $N_i = 0$ counts and therefore has $\sigma_i = 0$. Alternatively, adjacent bins can be combined to yield a nonzero content.

Fitting

Reasons for Fitting

Fitting a function to a set of data points, such as a histogram or a time series, may be done for any of the following reasons:

(1) A function could be fitted to a data set in order to describe its shape or behavior, without ascribing any "biophysical" meaning to the function or its parameters. This is done when a smooth curve is useful to guide the eye through the data or if a function is required to find the behavior of some data in the presence of noise.

(2) A theoretical function may be known to describe the data, such as a probability density function consisting of an exponential, and the fit is made only to extract the parameters, (*e.g.*, a time constant). Estimates of the confidence limits on the derived time constant may be needed in order to compare data sets.

(3) One or more hypothetical functions might be tested with respect to the data, *e.g.*, to decide how well the data were followed by the best fit function.

The fitting procedure begins by choosing a suitable function to describe the data. This function has a number of free parameters whose values are chosen in order to optimize the fit between the function and the data points. The set of parameters that gives the best fit is said to describe the data, as long as the final fit function adequately describes the behavior of the data. Fitting is best performed by software programs; the software follows an iterative procedure to successively refine the parameter estimates until no further improvement is found and the procedure is terminated. Feedback about the quality of the fit allows the model or initial parameter estimates to be adjusted manually before restarting the iterative procedure. Fitting by pure manual adjustment of the parameters (the so-called "chi by eye") may be effective in simple cases but is usually difficult and untrustworthy in more complex situations.

The two following topics will be briefly discussed below: *statistics*, *i.e.*, how good is the fit and how confident is the knowledge of the parameters, and *optimization*, *i.e.*, how to find the best fit parameters. The statistical aspects are well discussed in Eadie *et al.* (1971); Colquhoun and Sigworth (1983) provide examples relevant to the electrophysiologist. A number of aspects of optimization are presented in Press *et al.* (1988).

Statistical Aspects of Fitting

Statistics deals with the probability of occurrence of events. Probability is difficult to define; there are two ways in which the word is used. (1) *Direct* probability: If we observe that N_1 of N single-channel events have open channel durations between 10 and 20 ms, we say that the probability p_1 of this occurring is N_1/N, as long as N is very large. The probability density function (pdf) is an algebraic expression that when summed or integrated between 10 and 20 ms gives the value of p_1. (2) *Inverse* probability: If you are told by your physician that you have one of three possible diseases D_1, D_2 or D_3, and you ask what the probability is that you have D_2, the physician might say, "It's either 0 or 1," meaning that either you already have D_2 or you do not have it, but the physician cannot yet determine which of these two situations exists. To be helpful, your physician might give an *inverse* probability of 0.6, meaning that if you had D_2, the probability that your particular set of symptoms would have been observed is 0.6.

The Likelihood Function

Inverse probability is more appropriate for the scientist who may have N measurements of open channel durations and wants to know which time constant best describes their exponential distribution. For a particular time constant, τ_1, one can calculate the *direct* probability of getting those N observed durations by first calculating the probability of observing each duration and then multiplying these individual probabilities together. The resulting number is called the *likelihood* of the time constant τ_1. One could calculate the likelihoods for many such values of τ and plot the logarithm of these likelihoods versus their respective τ values (Figure 3). If N is sufficiently large, this curve will usually be

Gaussian. The value τ^* at the peak of the function is called the *maximum likelihood value* of τ. The root-mean-square spread of the function about τ is known as the *standard deviation of* τ^*, though this is not the same as a standard deviation of a set of numbers in the direct probability case.

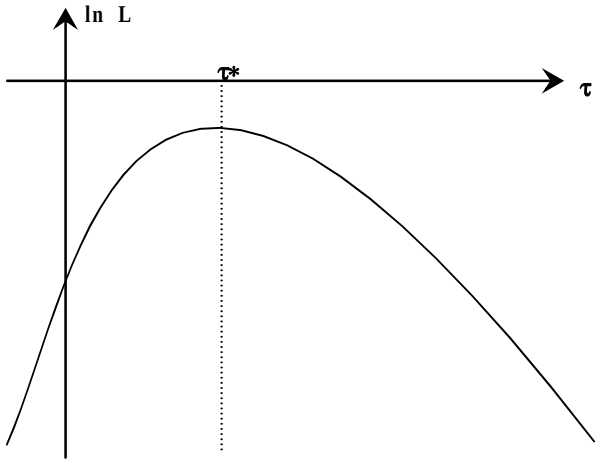

Figure 10-3. The Likelihood Function

The logarithm of the likelihood presented as a function of variable τ; the maximum likelihood of the function occurs at τ^*. For large numbers of samples, the shape of this curve approaches a Gaussian.

It turns out that τ^* reliably converges to the true time constant if the number of events N is sufficiently large. For this reason it is important to either collect a large enough number of data points or repeat the experiment several times so as to reduce the variation of the parameters obtained over the data sets to an acceptable level. If the noise is large or there are sources of significant variability in the signal, the data may be useless except in a qualitative way because of large variations of the best fit parameters between the runs. If one long run is taken instead of several smaller ones, the run can be broken up into segments and the analysis results of the segments compared with each other to assure that convergence is near.

Although the maximum likelihood method is the most reliable, the time requested for the calculations may be prohibitively long. The chi-square method, described below, is an alternative that requires less time.

The Chi-Square Function (Least-Squares Method)

Suppose that a set of p measurements are made at the times $x_1, x_2, ..., x_p$, and that the values measured are $y_1, ..., y_p$. If each y_i is measured with a measurement error distributed as a Gaussian with standard deviation $\sigma_1, ..., \sigma_p$, the maximum likelihood method is equivalent to minimizing the *chi-square function*

$$\chi^2 = \sum_{i=1}^{p} \frac{(y_i - y_i^*)^2}{\sigma_i^2} \qquad (2)$$

where y^*_i is the fit value corresponding to y_i. Minimizing chi-square is also called the *least-squares* method. If the fit is made to a data sweep, each y_i is the value measured at the time x_i, and each σ_i is the standard deviation or uncertainty of that value. In the typical case when all the σ_i's are equal (*i.e.*, the uncertainty in the data does not depend on the time), the σ_i's can be ignored while performing the search for the best fit parameters, but must be specified if the *goodness of fit* is to be calculated. If the fit is made to a histogram, each y_i is the numbers of events N_i in bin i, and each σ_i is $\sqrt{N_i}$.

It is much easier to maximize the chi-square function than to minimize the likelihood function, whether for data or for many mathematical functions used as models. Since the use of the chi-square function is equivalent to the use of the likelihood function only if the uncertainties in the data (*e.g.*, noise) are distributed as Gaussians, the correctness of a least-squares fit can depend on the characteristics of this uncertainty.

The Goodness of Fit

After a best fit has been obtained, the user might wish to know if the fit was good. If the fit to p data points employed M free parameters, the probability of obtaining a chi-square value greater than that of the best fit, given $p - M$ degrees of freedom, can be read from a table of probabilities of chi-square and compared to a chosen significance level.

Often several models are fitted to a single set of data so as to define the best-fit model. Horn (1987) discussed choosing between certain types of models using their respective chi-square values (with the F test) or the logarithm of the ratio of the likelihoods, which follows the chi-square distribution. These tests can help decide whether the model contains irrelevant parameters, *e.g.*, if a three-exponential function was fitted to a set of data containing only two exponentials.

Confidence Limits for the Parameters

An estimate of the confidence limits for each of the parameters is often useful. For a model with one parameter a, some fit programs will derive the standard deviation σ from the dependence of the likelihood or chi-square on the parameter. One then says that the 68.3% confidence limits are within a distance σ_a from a^*, the chosen value for a. This does *not* mean that 68.3% of the time the true value of a will fall between $a^* + \sigma_a$ and $a^* - \sigma_a$. It *does* mean that we will be right 68.3% of the time if we assert that these limits include the true a. In the limit of a large number of data sets, a^* will tend to converge to the true value of a.

Suppose there are two parameters a and b to be fit. One can make a two-dimensional contour plot of the likelihood or chi-square function, with each axis corresponding to one of the parameters, showing the probability of the limits including the true parameter values (Figure 4). The resultant ellipse is usually inclined at an angle to the parameter axes due to the *correlation* of the two parameters. These correlations tend to increase the confidence limits, which are indicated by the dotted lines which project to the respective axes. The probability that the confidence limits for the best fit a includes the true a is 0.683, but this

does not specify anything about *b*, as indicated by the shaded region. If, in addition, limits $\pm\sigma_b$ are specified for the best fit *b*, the joint probability that both sets of confidence limits include the true *a* and *b* has an upper limit of $(0.683)^2$ or 0.393, *i.e.*, the probability content of the area inside the ellipse. If one has a five exponential fit with an offset, the analogous cumulative joint probability will be $(0.683)^{11}$, or 0.015, which is quite small.

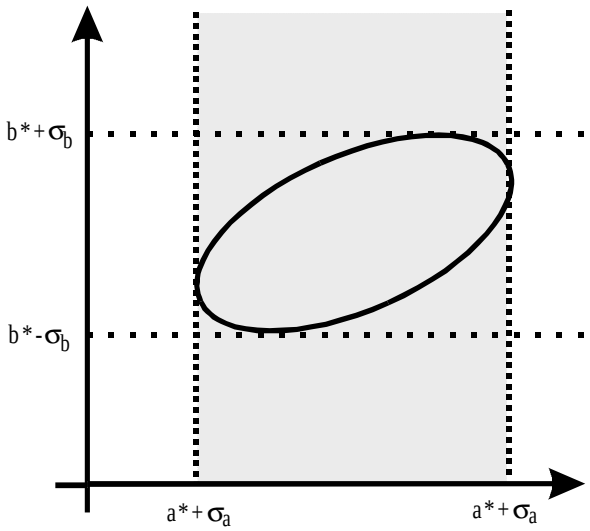

Figure 10-4. Confidence Limits

Two-dimensional contour plot of a likelihood or chi-square function *vs.* two parameters *a* and *b* around the function minimum, with the two sets of dashed lines indicating the respective confidence limits.

Statistical Tests of Significance for Parameters

If one wishes to compare, for example, two time constants from two data sets obtained under different conditions, one must first obtain the standard deviation of the time constant, usually derived from the fitting procedure. Conventional statistical tests, such as the chi-square table, the *F* test or Student's *t* test, can then be applied to determine significance.

Methods of Optimization

Optimization methods are concerned with finding the minimum of a function (*e.g.*, the chi-square) by adjusting the parameters. A *global* minimum, *i.e.*, the absolute minimum, is clearly preferred. Since it is difficult to know whether one has the absolute minimum, most methods settle for a *local* minimum, *i.e.*, the minimum within a neighborhood of parameter values. A number of algorithms have been developed to find such a minimum. For example, to find time constants and coefficients in an exponential fit, the pCLAMP program pSTAT allows the user to choose between the following:

- Minimizing the chi-square using the Levenberg-Marquardt method.
- Minimizing the chi-square using the Simplex method.
- Maximizing the likelihood using the Simplex method.

Of the three methods, the Simplex method is fast and relatively insensitive to shallow local minima. Though it will reliably find the region of the global minimum or maximum, it may not find the precise location of the minimum or maximum if the function is rather flat in that vicinity. The Levenberg-Marquardt method is slower and more easily trapped in local minima of the function, but it can provide better fits than the Simplex because it uses the mathematical characteristics of the function being minimized to find the precise location of the minimum or maximum, within the numerical resolution of the computer. This method also provides statistical information sufficient to find the confidence limits.

These methods are *iterative*, *i.e.,* they continue refining parameter values until the function stops changing within a certain convergence criterion. They also require reasonable starting estimates for the parameters, so that the function to be minimized or maximized is not too far away from its optimum value; a poor starting set can lead some fit programs to a dead end in a shallow local minimum.

Axon Instruments' analysis programs CLAMPFIT for the IBM-PC and AxoGraph for the Apple Macintosh provide a *non-iterative* method, in which the data points and function to be fit are transformed using a set of orthogonal Chebyshev polynomials, and the fit function coefficients are quickly calculated using these transformed numbers in a linear regression. This method is very fast and requires no initial guesses, though the parameters may be slightly different than those found by the methods listed above because the underlying algorithm minimizes a quantity other than the sum of squared differences between fit and data.

References

Colquhoun, D. and Sigworth, F.J. *Fitting and Statistical Analysis of Single-Channel Records*. in **Single-Channel Recording**. Sakmann, B. and Neher, E., Eds. Plenum Press, New York, 1983.

Colquhoun, D. and Hawkes, A.G. *The Principles of the Stochastic Interpretation of Ion-Channel Mechanisms*. in **Single-Channel Recording**. Sakmann, B. and Neher, E., Eds. Plenum Press, New York, 1983.

Dempster, J. **Computer Analysis of Electrophysiological Signals**. Academic Press, London, 1993.

Eadie, W.T., Drijard, D., James, F.E., Roos, M, Sadoulet, B. *Statistical Methods in Experimental Physics*. North-Holland Publishing Co., Amsterdam, 1971.

Horn, R. *Statistical methods for model discrimination*. Biophysical Journal. 51:255-263, 1987.

McManus, O.B., Blatz, A.L. and Magleby, K.L. *Sampling, log binning, fitting, and plotting distributions of open and shut intervals from single channels and the effects of noise*. Pflügers Arch, 410:530-553, 1987.

Ogden, D.C. *Microelectrode electronics*. in **Microelectrode Techniques. The Plymouth Workshop Handbook**. Standen, N.B., Gray, P.T.A. and Whitaker, M.J., eds. The Company of Biologists Ltd., Cambridge, 1987.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. **Numerical Recipes in C**. Cambridge University Press, Cambridge, 1988.

Sachs, F. *Automated Analysis of Single-Channel Records*. in **Single-Channel Recording**. Sakmann, B. and Neher, E., Eds. Plenum Press, New York, 1983.

Sigworth, F.J. and Sine, S.M. *Data transformations for improved display and fitting of single-channel dwell time histograms*. Biophysical Journal, 52:1047-1054, 1987.

Sine, S.M. and Steinbach, J.H. *Activation of acetylcholine receptors on clonal mammalian BC3H-1 cells by low concentrations of agonist*. Journal of Physiology (London), 373:129-162, 1986.

Wonderlin, W.F., French, R.J. and Arispe, N.J. *Recording and analysis of currents from single ion channels*. in **Neurophysiological Techniques. Basic Methods and Concepts**. Boulton, A.A., Baker, G.B. and Vanderwolf, C.H., Eds. Humana Press, Clifton, N.J., 1990.