PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

A quantitative method for the detection of edges in noisy time–series

D. A. Smith

Phil. Trans. R. Soc. Lond. B 1998 **353**, 1969-1981
doi: 10.1098/rstb.1998.0348

THE ROYAL SOCIETY

A quantitative method for the detection of edges in noisy time-series

D. A. Smith

Randall Institute, King's College, 26–29 Drury Lane, London WC2B 5RL, UK

A modification of the edge detector of Chung & Kennedy is proposed in which the output provides confidence limits for the presence or absence of sharp edges (steps) in the input waveform. Their switching method with forward and backward averaging windows is retained, but the output approximates an ideal output function equal to the difference in these averages divided by the standard deviation of the noise. Steps are associated with peak output above a pre-set threshold. Formulae for the efficiency and reliability of this ideal detector are derived for input waveforms with Gaussian white noise and sharp edges, and serve as benchmarks for the switching edge detector. Efficiency is kept high if the threshold is a fixed fraction of the step size of interest relative to noise, and reliability is improved by increasing the window width W to reduce false output. For different steps sizes D, the window width for fixed efficiency and reliability scales as $1/D^2$. Versions with weighted averaging (flat, ramp, triangular) or median averaging but the same window width perform similarly. Binned above-threshold output is used to predict the locations and signs of detected steps, and simulations show that efficiency and reliability are close to ideal. Location times are accurate to order $\sqrt{W}$. Short pulses generate reduced output if the number of data points in the pulse is less than W. They are optimally detected by choosing W as above and collecting data at a rate such that the pulse contains approximately W data points. A Fortran program is supplied.

Keywords: edge detector; predictors; efficiency; reliability; switching method

1. INTRODUCTION

The detection of 'edge' signals in a noisy time-series is required for many purposes, including the study of single ion-channel currents through cell membranes (Sakmann & Neher 1983), and force and displacement steps generated by single motor proteins using optical tweezers (Svoboda *et al.* 1993; Finer *et al.* 1994). Low-pass filtering techniques have traditionally been used to extract signals out of a noisy time-series (Oppenheim & Schafer 1989), but more sophisticated methods based on analysis of variance (Mosteller & Tukey 1977), Fourier or wavelet transforms (Press *et al.* 1992), or neural networks (Peretto 1992), are required to detect abrupt jumps in the signal, where each jump may occur on a time-scale comparable to or less than the correlation time of the noise. In many applications the focus is initially on the detection of individual 'edge' events, rather than the statistical description of the underlying processes. To avoid making subjective judgements about each event it is essential to use statistical predictors for the efficiency and reliability of the detector (Block & Svoboda 1995). Although robust predictors are available (Kassam & Poor 1985), it is natural to use a superposition of sharp edges (steps) and Gaussian white noise as a standard input to which performance criteria can be referred. Variance-based methods are traditionally preferred because quantitative predictors of performance are available by adapting the Student's t-test for the significance of the difference in the means of two populations (Weatherburn 1968).

A variance-based method for edge detection uses two moving averaging windows of width W, providing forward and backward local averages of each data point. A Student's t-statistic at each point can then be constructed by dividing the difference in the forward and backward means by the standard error in both windows. However, this quantity can be conventionally applied to edge detection only if the edge is located between the two windows, which is generally not the case. When the edge occurs within one window, the t-test overestimates the confidence limit (the probability that the difference is due to noise only). This behaviour arises because the variance in the window with the edge contains a large contribution from the edge itself. The contamination of one variance by an edge was turned to advantage by Chung & Kennedy (1991) to construct from the input $x(t)$ at time t a filtered output function $X_{\mathrm{out}}(t)$, which ideally is the mean input in the window with no edge. To approach this result, they proposed that the ratio of the variances of $x(t)$ in each window be used to switch the output to the average in the window with smaller variance. If the switching function is sharp enough, it follows that $X_{\mathrm{out}}(t)$ will show an unbroadened edge in response to a step in the input, while the noise component of the input is reduced by averaging. Similar results can be obtained by median-averaging (Gallagher & Wise 1981) but the latter is computationally more expensive.

In this paper, Chung & Kennedy's switching technique is used to define a dedicated edge detector. For each time, t, in the input waveform, the output $Y(t)$ is chosen to approximate an ideal output function equal to the

Received 8 April 1997 *Accepted* 30 June 1997

difference of the forward and backward averages divided by the standard deviation of the noise component only. This quantity is proportional to the Student's *t*-statistic for the populations in each window, and can be associated with confidence limits for the presence and absence of an edge at that time. Thus, the efficiency of the detector is measured by the probability $P_T(D)$ that an edge of size D in the input waveform is detected. Its reliability is inversely indicated by the rate of generating false output signals or the probability P_F that an edge is detected when none is present in the input. The threshold Y_c for the absolute value of Y should then be set for maximum efficiency and reliability, that is

$$P_F \ll P_T(D) \approx 1. \tag{1}$$

It is convenient to define Y so that the expected output from a step of size D is equal to D/σ where σ is the standard (root-mean-square) noise level in the input. The standard noise output in the absence of edges is a function of the width W of each averaging window, and is of the order of $1/\sqrt{W}$. Hence equation (1) should be obeyed if the threshold satisfies

$$\frac{1}{\sqrt{W}} \ll Y_c < \frac{|D|}{\sigma}, \tag{2}$$

which is possible for steps of magnitude comparable to noise by using wide windows. This will cause some loss of time resolution, but it can be shown that the standard error in the time of location of a step is proportional to $\sqrt{W}$, which for large windows is much less than the window width. In practice, choosing $Y_c = 2D/3\sigma$ to detect steps of size D or greater keeps the detection efficiency close to unity, while W is determined by setting an acceptable level P_F of false output. Figure 1 illustrates the operation of both forms of the switching edge detector on a time-series of current interest.

In the next section, Chung & Kennedy's detector is redefined with the new output function Y and the statistics of its point output are compared with those of the ideal detector, for which the distribution is known exactly. Comparisons are made first in the presence or absence of a single step. They are easily extended to short-lived pulses (positive plus negative steps or vice versa) when the duration of the pulse is less than W. In § 3 it is considered how output should be collated to avoid multiple detections of the same edge without losing efficiency or location accuracy. Binned output from numerical simulations (§ 4) is used to calculate efficiency, reliability and location errors as a function of step size and window width. The binned output cannot easily be compared with the ideal case, but is used to assess different methods of window-averaging. In § 5 it is suggested how the detector might best be used. Readers not concerned with details or numerical testing should read § 2 and § 5 and proceed to figure 4, which gives the performance curves required for best use.

2. THE SWITCHING EDGE DETECTOR AND ITS IDEAL

Let the input time-series x_i be defined at time points $t_i = ih$ $(i = 1, \ldots, N_T)$ spaced by h. Unless stated otherwise, this series is assumed to consist of a Gaussian random

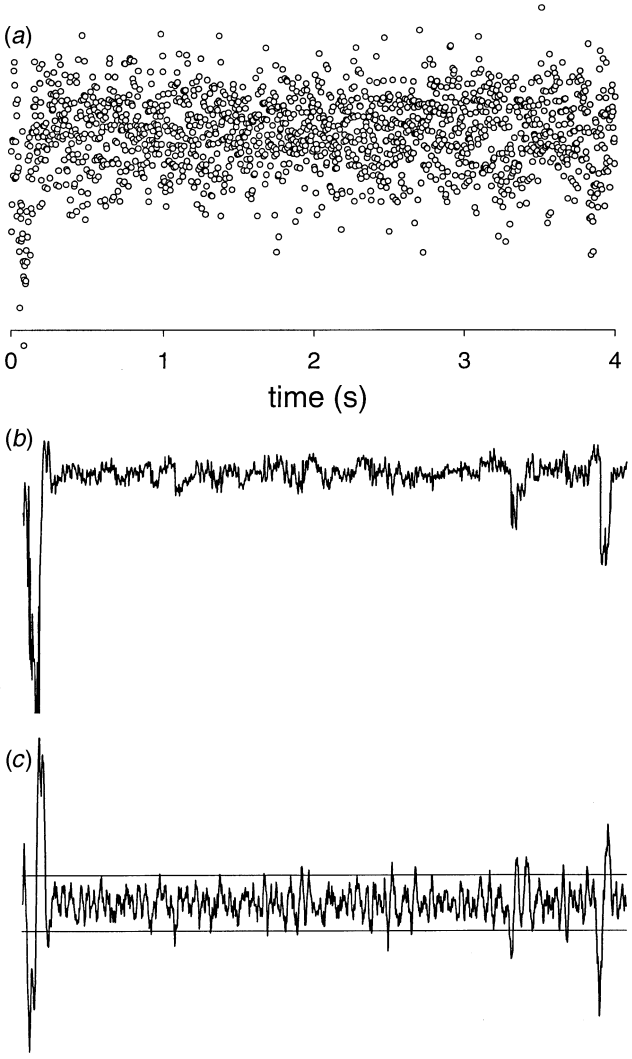

Figure 1. (*a*) Time-series of the displacement of an actin filament tethered by optically trapped glass beads, containing interaction events with myosin molecules tethered to a coverslide (A. Trombetta, unpublished data). (*b*) The corresponding filtered output function X_{out} from the switching detector of Chung & Kennedy (1991), by using $W = 100$. (*c*) Y-output from the modified switching edge detector. The threshold lines $Y = \pm Y_c$ are drawn for $Y_c = 0.333$, appropriate for a minimum step size of half the standard noise deviation in the input waveform.

process and abrupt step changes. Construct moving averages $X_{i\pm}$ and the variances $S^2_{i\pm}$ from these averages over two windows of W points, forward and backward of the ith time point:

$$X_{i\pm} = \frac{1}{W}\sum_{k=1}^{W} x_{i\pm k}, \qquad s^2_{i\pm} = \frac{1}{W}\sum_{k=1}^{W} (x_{i\pm k} - X_{i\pm})^2. \tag{3}$$

The ideal output function Y_i is defined as

$$Y_i = \frac{X_{i+} - X_{i-}}{S_i}, \tag{4}$$

where S_i^2 is the variance of the noise alone as estimated from one window. For the moment it is assumed that steps are sparse enough for one step at most to lie within the windows; this restriction is relaxed later. Then one window will always be free of a step and its variance can be used for this purpose. The statistical motivation for

equation (3) is that if the step is at time t_i then the x-values in each window differ by noise alone and Y_i is proportional to the Student's t-statistic (Weatherburn 1968) measuring the difference in their means, at least if the noise variance is the same in both windows.

In practice, the locations of edges are not known in advance and so Chung & Kennedy (1991) proposed that the variances defined in equation (3) be used to define switching factors

$$g_{i+} = \frac{S_{i-}^{2r}}{S_{i+}^{2r} + S_{i-}^{2r}}, \qquad g_{i-} = \frac{S_{i+}^{2r}}{S_{i+}^{2r} + S_{i-}^{2r}}, \tag{5}$$

and hence their output function $X_{\mathrm{out}}(t_i) = g_{i+}X_{i+} + g_{i-}X_{i-}$. The positive exponent r is arbitrary but $r \gg 1$ is required to produce abrupt switching between 0 and 1 as a function of the ratio S_{i+}/S_{i-}. However, the same principle can be used to generate the derivative-like output function in equation (4) if the 'noise-only' variance is approximated by

$$S_i^2 \approx g_{i+}S_{i+}^2 + g_{i-}S_{i-}^2. \tag{6}$$

Thus equations (3–6) define a practical switching detector, while the first two equations contain the ideal case.

The nature of the output waveform can be understood from analysing the signal and noise contributions to Y_i that arise from step signals and Gaussian random noise respectively in the input. In the ideal case, the denominator in equation (4) is due to noise alone, but the numerator is the sum of signal and noise terms. Let $\Delta(t)$ be the signal component of the difference of the two means, so that the signal component of the output is $\Delta(t)$ multipled by the time-averaged inverse standard deviation $(1/S) \approx 1/\sigma$. It responds as shown in figure 2*b* to a step of size D, with a triangular waveform of width $2W$ and a peak value D/σ at the time of the step. For large edges, the peak output measures the size of the step directly in units of σ, and for numerical comparisons σ is set to unity. For smaller steps, noise fluctuations must be taken into account. In the presence of noise only, Y is symmetrically distributed about zero with a variance of $2/(W-3)$ (Appendix A, equation (A9)). With an edge present at, say, time 0 the distribution of Y at time $t \neq 0$ is shifted according to the value of $\Delta(t)$, and its most probable value is close to $\Delta(t)/\sigma$. Hence the output is most likely to peak at the time of the edge itself (figure 2*c*), where $\Delta(t)$ has its maximum value D (for $D > 0$). However, noise components in the numerator and denominator conspire to shift the location and the size of the maximum in the output waveform.

The virtue of using the Y function as output is that its probability distribution $p(Y|\Delta)$ for a given value of Δ can be calculated from first principles (Appendix A, equation (A5)). The probability of above-threshold positive output ($Y > Y_c$) in the presence of a step defined by Δ is given by the integrated distribution

$$P(Y_c|\Delta) = \int_{Y_c}^{\infty} p(Y|\Delta)\,\mathrm{d}Y. \tag{7}$$

Hence,

$$P_{\mathrm{T}}(D) = P(Y_c|D), \qquad P_{\mathrm{F}} = P(Y_c|0), \tag{8}$$

are the probabilities of true output for a positive step of size D and of false output (output for no step). These quantities are defined for each point in the time-series, and the averaging effect of the two windows should not be forgotten. If the output is measured at a different time t to that of the step, then equation (8) gives the probability of true output if D is replaced by $\Delta(t)$, and of false output if both windows are step-free. They give input to the quantitative criteria (equation (1)) by which edge detectors should be judged, and will be used for comparing the performance of the switching edge detector with the ideal case, and with versions using different averaging methods.

Figure 3 shows the probability function $P(Y|D)$ as calculated from equation (A5) for the ideal edge detector, and as sampled from the switching edge detector with $r = 50$ and synthetic input containing Gaussian random noise and steps of fixed size D. The steps were repeated at intervals of $4W$ for windows of size W to avoid overlaps; further details are given in the legend. For all step sizes, the sampled distribution from the detector is close to but slightly higher than the ideal distribution. This appears to be an effect of finite sampling which generates slightly different numbers of positive and negative output events from Gaussian input, reflected in an estimate for $P(Y_c = 0|\Delta = 0)$ (the probability of positive output from noise) not equal to 0.5. To make P_{T} close to 1 and P_{F} close to 0 (equation (1)), the knees of the distributions for $D > 0$ and $D = 0$ should be widely separated and the threshold Y_c set between them. The choice

$$Y_c = \frac{2D}{3\sigma}, \tag{9}$$

is sufficiently below the most probable value D/σ of Y to give detection efficiencies near unity in almost all cases of interest (above 94% for $D/\sigma > 0.5$ and $W \geqslant 20$). However, for small steps the distribution of signal output tends to overlap the noise distribution, and the probability of false output can be kept below 5% only if $D/\sigma > 0.84$ for $W = 20$ or 0.35 for $W = 100$. These and other values can be obtained from table 1.

For short pulses, the distribution of output can be obtained without further calculation, because the value of Y at any time t is controlled by the difference $\Delta(t)$ of mean values arising from the signal component of the input. For a step $+D$ followed by a step $-D$ after time τ, the peak value of $|\Delta(t)|$ is reduced from D to $D\tau/W$ if $\tau < W$ (figure 2*e*). Therefore, in this case each step is detected with confidence limits equal to those for an isolated step of reduced amplitude $D_{\mathrm{eq}} \equiv D\tau/W$.

The output function $Y(t)$ can predict the locations and signs (+ or −) of detected steps, and this type of output requires much less storage. Although each value of $Y(t)$ above Y_c detects a step up and each value below $-Y_c$ a step down, there will usually be far more above-threshold values than input steps. Some of this redundancy is owing to multiple detections of the same step, because the difference of the two window averages generates a triangular pulse of width $2W$ from a sharp edge (figure 2*b*). Hence the output function should be collated into time bins whose width is comparable with W. Binning is a necessary preliminary to generating logical output.

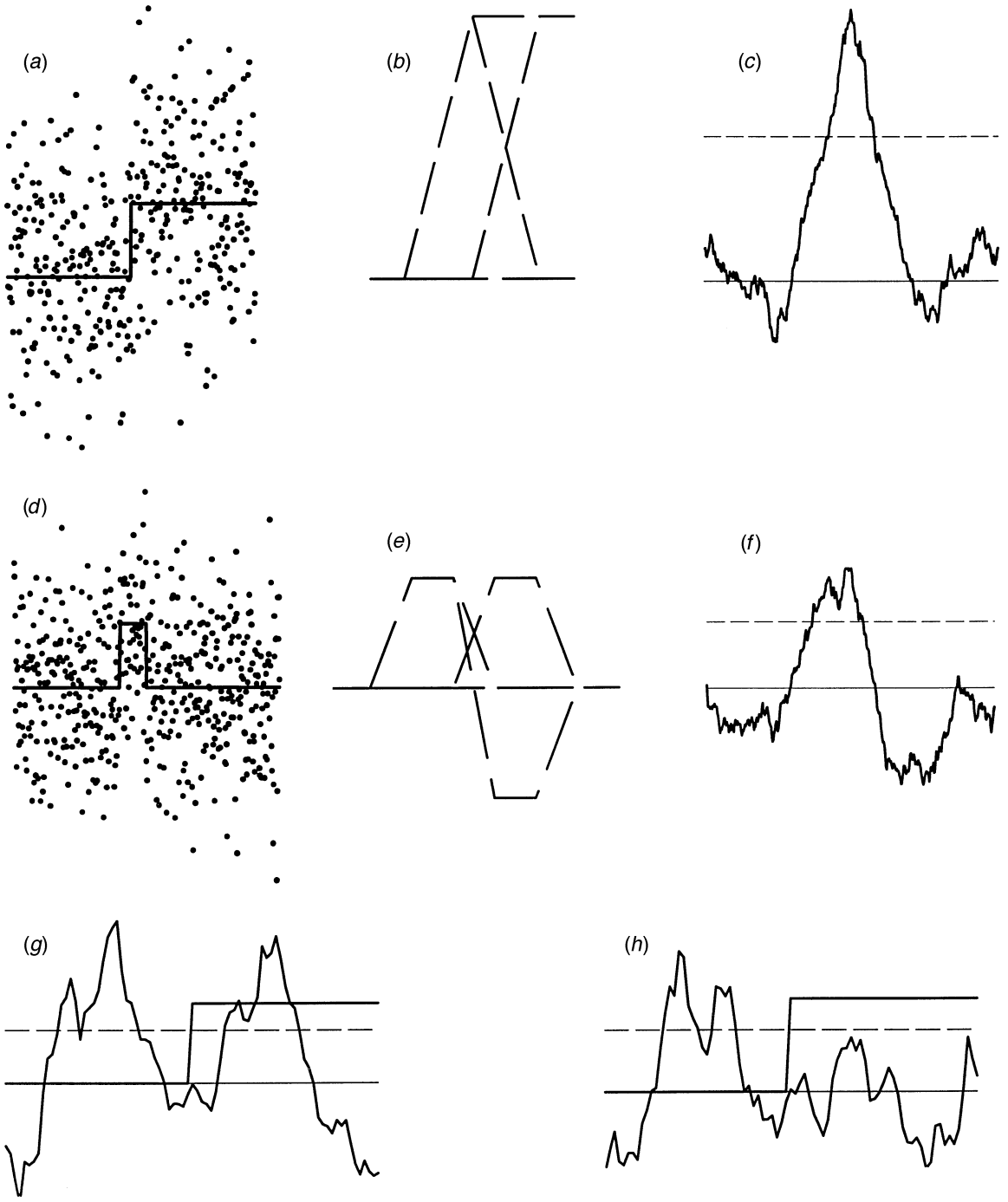

Figure 2. Examples of input and Y-output waveforms from the switching edge detector with uniform averaging. (*a*) Gaussian noise input (unit variance) with a single step $D=1$ noise unit. (*b*) The signal component $\Delta(t)=(X_+(t) - X_-(t))/\sigma$ of the output with $W=100$, giving a triangular response (solid line) as the forward and backward window averages (dashed line) rise through the step. (*c*) The full output $Y(t)$ (actual simulation) showing a response peaking within a few data points of the step. (*d*) Input waveform with a short pulse of 50 data points, (*e*) corresponding signal output and (*f*) total output, also with $W=100$. (*g*) Multiple binned detection of a single edge with $E = W$ (simulation, $D=0.5$ and $W=20$). (*h*) True edge output (within the signal triangle) deleted under binning by a false edge using a bin width W (simulated as in (*g*)). Dashed horizontal lines in (*c*, *f–h*) mark the upper threshold, and bars the detection window of width $2W$.

3. RULES FOR BINNED LOGICAL OUTPUT

To produce a list of the times and signatures of detected steps, the raw output $Y(t)$ must be examined for above-threshold peaks. Multiple detections of the same edge will arise from noise fluctuations which add to the triangular response function $\Delta(t)$, while opposing fluctuations may cause edges to be missed. Multiple output events can be reduced or avoided altogether by deleting peak events that are too close to other peaks. Consider the following binning procedure for detecting positive steps, which operates by deleting apparent steps within E data points of each other: (i) for each time point t_i with above-threshold positive output ($Y>Y_c$), test whether Y_i is a local maximum; (ii) if so, search over the previously recorded times t_j of positive steps, if any, in the range

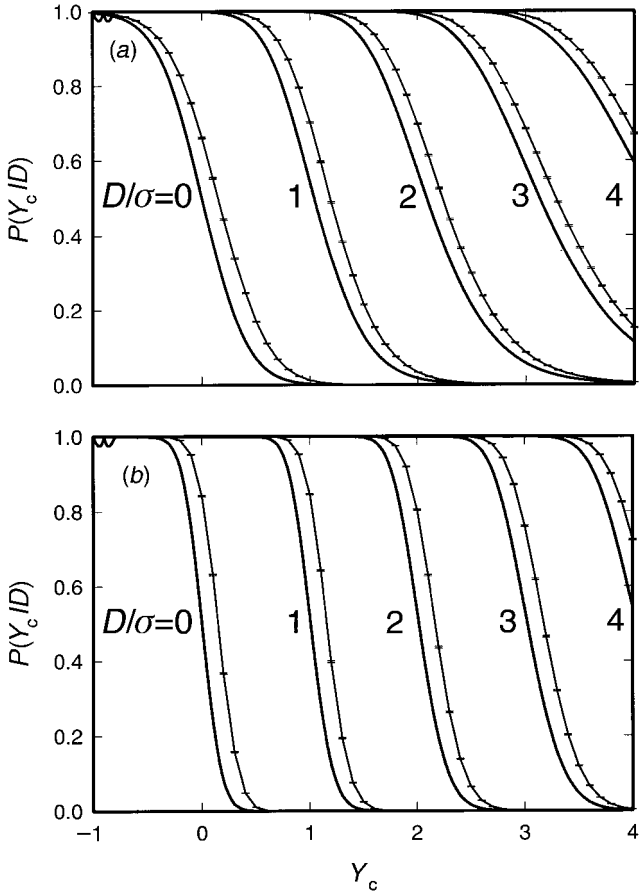

Figure 3. The probability of above-threshold output as a function of the threshold level Y_c for various values of D/σ (step size relative to standard noise deviation), using window widths (*a*) $W=20$, and (*b*) $W=100$. Full lines show output from the ideal edge detector using equations (6 and 7), and error bars with lines are from the switching edge detector using synthetic input with $r=50$. Input waveforms of 20 000 points were obtained from a Gaussian random number generator (Press *et al*. 1992) of unit variance plus steps of size +10 at times $4W$, $8W$, ... and so on. As the detector responds at time t to the quantity $\Delta(t) = X_+(t) - X_-(t)$ of figure 2*b* which passes linearly from 10 to 0 in the time-intervals $(4W,5W)$, $(8W,9W)$,..., responses equivalent to those from steps as above were recorded as $\Delta(t)$ passed through the values 5,4,3,...0. The number of above-threshold responses was recorded in each case. This process was iterated 200 or 1000 times, respectively, with different random-number-generator calls to generate 50 000 steps of each size. Error bars represent the standard deviation expected from a binomial distribution by using the sample mean.

$i-E \leqslant j < i$ for an additional maximum; and (iii) keep in the record only the time with the highest maximum value of Y (delete subsidiary maxima over the bin of width E).

These operations produce a list of the times of detected positive steps, all separated by more than E data points. An equivalent procedure with negative output ($Y < -Y_c$) and local minima does the same for negative steps. As positive and negative steps are binned independently, the detection of a short pulse will be affected by deletions in the same way as single edges. However, repeated steps in the same direction with spacings $\leqslant E$ will be removed.

What are the effects of this binning procedure on true and false output events, and how should E be chosen relative to W? Apart from reducing multiple detection events (figure 2*g*), there are three other effects to consider because false steps arising from noise can delete each other, a false event can be deleted by a true one and a true event by a false one. The last effect (figure 2*h*) lowers the efficiency of detection and should be avoided. If $E=2W$, all multiple detections are eliminated because output events from a single input step are confined to a time-interval $2W$ centred on the step time. However, the deletion of true output by false output is significant and so smaller values of E should be considered. With $E = W$, simulations show that the efficiency of detection is higher and the number of multiple events remains small. Errors in the times of located steps are also reduced with the smaller bin width. The effect on the incidence of false events is not large if one remembers that all events are now separated by E time points or more. Hence the probability of a false step should be defined over a time-interval E.

A better appreciation of the occurrence of multiple detection events can be obtained from the distribution of raw output for times within one window width of an input step. Again, this is done by using the function $\Delta(t)$, which drops from D to zero on moving away from the edge, as the height of an equivalent edge. If $E=2W$ then two detections of the edge can arise only at the outer ends of the double window and the probability of both is $P(Y_c|0)^2$ where each event is equivalent to noise. When $E=W$, two detections can arise in a number of ways, the most probable being from times $\pm W/2$ away from the edge. These events are not statistically independent as their averaging windows overlap, but their probability must lie between $P(Y_c|D/2)$ and its square. Hence with $W=20$ and $D/\sigma=1$, the probability of a multiple detection is $(0.026)^2=6.7\times10^{-4}$ if $E=2W$ and lies between 0.336 and $(0.336)^2=0.113$ if $E=W$ (table 1). These estimates take no account of deletions produced by binning, which lowers the incidence of multiple events.

4. NUMERICAL TESTS, SCALING LAWS AND AVERAGING METHODS

The statistics of binned logical output of the edge detector cannot be analysed as simply as the raw output $Y(t)$, which is why raw output was used for comparisons with the ideal case. However, the probabilities of true and false output can be estimated by sampling large populations of output events, and the results provide a set of design curves which permit the detector to be used with maximum benefit in specified situations. The detection of rectangular pulses whose width is comparable to, or shorter than, W has also been investigated, and statistics for the errors in the detection times of true steps are presented. Scaling laws exist which connect the above probabilities with different step sizes and window widths and allow information to be displayed on a single curve. As suggested in §2, another scaling law connects the detection efficiency of a short pulse with that of a long pulse of reduced height. These laws combine to give an optimum method of detecting short pulses. Finally, different methods of taking averages in the two windows are investigated numerically.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 1. *The distribution $P(Y|D)$ from equations (8 and 9), Y values within table*

	D								
Y	0.0	0.1	0.2	0.3	0.4	0.5	1.0	1.5	2.0
(*a*) $W=20$									
0.00	0.5000	0.6241	0.736	0.8286	0.8970	0.9431	0.9992	1.0000	1.0000
0.05	0.4396	0.5652	0.6844	0.7871	0.8670	0.9234	0.9987	1.0000	1.0000
0.10	0.3806	0.5048	0.6285	0.7401	0.8313	0.8989	0.9978	1.0000	1.0000
0.15	0.3246	0.4445	0.5697	0.6883	0.7900	0.8690	0.9965	1.0000	1.0000
0.20	0.2725	0.3856	0.5096	0.6326	0.7432	0.8335	0.9945	1.0000	1.0000
0.25	0.2252	0.3297	0.4495	0.5741	0.6917	0.7923	0.9914	1.0000	1.0000
0.30	0.1834	0.2777	0.3909	0.5143	0.6363	0.7458	0.9869	0.9999	1.0000
0.35	0.1471	0.2305	0.3352	0.4546	0.5782	0.6946	0.9806	0.9999	1.0000
0.40	0.1163	0.1886	0.2833	0.3964	0.5189	0.6397	0.9718	0.9997	1.0000
0.45	0.0908	0.1522	0.2362	0.3410	0.4597	0.5822	0.9599	0.9995	1.0000
0.50	0.0699	0.1211	0.1943	0.2894	0.4020	0.5234	0.9443	0.9991	1.0000
0.55	0.0532	0.0952	0.1577	0.2424	0.3471	0.4649	0.9245	0.9984	1.0000
0.60	0.0400	0.0740	0.1265	0.2005	0.2958	0.4078	0.8999	0.9974	1.0000
0.65	0.0298	0.0568	0.1002	0.1639	0.2491	0.3534	0.8702	0.9957	1.0000
0.70	0.0220	0.0432	0.0786	0.1323	0.2072	0.3026	0.8354	0.9931	1.0000
0.75	0.0161	0.0326	0.0609	0.1057	0.1704	0.2561	0.7955	0.9893	0.9999
0.80	0.0117	0.0243	0.0468	0.0836	0.1387	0.2143	0.7511	0.9839	0.9998
0.85	0.0084	0.0180	0.0357	0.0655	0.1117	0.1774	0.7028	0.9763	0.9997
0.90	0.0060	0.0133	0.0270	0.0509	0.0891	0.1455	0.6514	0.9661	0.9994
0.95	0.0043	0.0097	0.0203	0.0392	0.0705	0.1181	0.5981	0.9528	0.9990
1.00	0.0031	0.0071	0.0151	0.0300	0.0553	0.0951	0.5439	0.9358	0.9983
1.10	0.0015	0.0037	0.0083	0.0172	0.0334	0.0602	0.4371	0.8896	0.9953
1.20	0.0008	0.0019	0.0045	0.0097	0.0196	0.0371	0.3387	0.8262	0.9887
1.30	0.0004	0.0010	0.0024	0.0054	0.0114	0.0223	0.2537	0.7470	0.9760
1.40	0.0002	0.0005	0.0013	0.0030	0.0065	0.0132	0.1843	0.6563	0.9541
1.50	0.0001	0.0003	0.0007	0.0016	0.0037	0.0077	0.1304	0.5598	0.9203
1.60	0.0000	0.0001	0.0004	0.0009	0.0021	0.0045	0.0901	0.4641	0.8726
1.70	0.0000	0.0001	0.0002	0.0005	0.0011	0.0026	0.0611	0.3745	0.8113
1.80	0.0000	0.0000	0.0001	0.0003	0.0006	0.0015	0.0408	0.2947	0.7382
1.90	0.0000	0.0000	0.0001	0.0001	0.0004	0.0009	0.026	0.2268	0.6568
2.00	0.0000	0.0000	0.0000	0.0001	0.0002	0.0005	0.0175	0.1711	0.5715
2.10	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0114	0.1269	0.4867
2.20	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0073	0.0927	0.4061
2.30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0047	0.0669	0.3327
2.40	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0030	0.0478	0.2679
2.50	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0019	0.0339	0.2125
(*b*) $W=100$									
0.00	0.5000	0.7602	0.9213	0.9830	0.9977	0.9998	1.0000	1.0000	1.0000
0.05	0.3629	0.6391	0.8560	0.9616	0.9934	0.9993	1.0000	1.0000	1.0000
0.10	0.2418	0.5021	0.7615	0.9218	0.9831	0.9977	1.0000	1.0000	1.0000
0.15	0.1470	0.3653	0.6407	0.8566	0.9616	0.9933	1.0000	1.0000	1.0000
0.20	0.0813	0.2442	0.5042	0.7624	0.9218	0.9830	1.0000	1.0000	1.0000
0.25	0.0409	0.1494	0.3679	0.6420	0.8566	0.9613	1.0000	1.0000	1.0000
0.30	0.0187	0.0834	0.2472	0.5063	0.7626	0.9213	1.0000	1.0000	1.0000
0.35	0.0078	0.0425	0.1524	0.3708	0.6430	0.8560	1.0000	1.0000	1.0000
0.40	0.0030	0.0198	0.0860	0.2507	0.5084	0.7624	1.0000	1.0000	1.0000
0.45	0.0010	0.0084	0.0445	0.1560	0.3740	0.6437	0.9999	1.0000	1.0000
0.50	0.0003	0.0033	0.0211	0.0892	0.2547	0.5104	0.9997	1.0000	1.0000
0.55	0.0001	0.0012	0.0092	0.0468	0.1601	0.3774	0.9990	1.0000	1.0000
0.60	0.0000	0.0004	0.0037	0.0227	0.0928	0.2590	0.9969	1.0000	1.0000
0.65	0.0000	0.0001	0.0014	0.0101	0.0496	0.1646	0.9914	1.0000	1.0000
0.70	0.0000	0.0000	0.0005	0.0042	0.0245	0.0969	0.9791	1.0000	1.0000
0.75	0.0000	0.0000	0.0002	0.0016	0.0113	0.0528	0.9546	1.0000	1.0000
0.80	0.0000	0.0000	0.0000	0.0006	0.0048	0.0267	0.9118	1.0000	1.0000
0.85	0.0000	0.0000	0.0000	0.0002	0.0019	0.0126	0.8456	1.0000	1.0000
0.90	0.0000	0.0000	0.0000	0.0001	0.0007	0.0055	0.7545	1.0000	1.0000
0.95	0.0000	0.0000	0.0000	0.0000	0.0003	0.0023	0.6427	0.9998	1.0000
1.00	0.0000	0.0000	0.0000	0.0000	0.0001	0.0009	0.5194	0.9993	1.0000
1.10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.2852	0.9942	1.0000

(*Cont.*)

Table 1(*b*) (*Cont.*)

1.20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1233	0.9692	1.0000
1.30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0423	0.8930	1.0000
1.40	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0117	0.7395	0.9998
1.50	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0027	0.5265	0.9980
1.60	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.3143	0.9882
1.70	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.1565	0.9535
1.80	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0654	0.8690
1.90	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0233	0.7220
2.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0072	0.5317
2.10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0020	0.3420
2.20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.1917
2.30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0941
2.40	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0408
2.50	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0159

For this purpose, time-series were generated by using Gaussian noise of unit variance ($\sigma=1$) plus edges of size $+D$ and $-D$, repeated every $8W$ time points for detection with windows of width W. The separation τ between each positive edge and the next negative one was variable in the range $(0, 4W)$, but was normally kept at the upper limit, giving a mark–space ratio of 1. Binned logical output was generated by using bin widths $E=W$, $1.5W$ and $2W$, and window widths $W=20$ and 100 as before. The numbers of detected edges of each kind were counted, also the numbers that could be matched with input edges were counted as follows. For each input edge of given sign, a detected edge is counted as true if located within W time points of an input edge. Although detected edges are always separated by E points, this rule permits multiple detections of the same input edge if $E<2W$ (figure 2*g*). If multiple events occur, only the nearer one is counted true. In this way the probabilities of true and multiple output events are sampled as

$$P_{\mathrm{T}}(D)=\frac{\mathcal{T}(D)}{\mathcal{N}}, \qquad P_{\mathrm{M}}(D)=\frac{M(D)}{\mathcal{N}}, \tag{10}$$

for $\mathcal{N}$ edges in the input and $\mathcal{T}$ true, M multiple edge events in the output. Because the binning operations mix true and false output, the probability of a binned false output event has to be obtained from a separate simulation using noise only, taking

$$P_{\mathrm{F}}(Y_{\mathrm{c}})=\frac{L(Y_{\mathrm{c}})}{\mathcal{N}_{\mathrm{T}}/E}, \tag{11}$$

for L binned positive output events ($Y>Y_{\mathrm{c}}$) in a run of $\mathcal{N}_{\mathrm{T}}$ data points and bin width E. This quantity is the rate of false steps per bin. It is the analogue of P_{F} in equation (8) for unbinned output, and should be close to 0.5 when $Y_{\mathrm{c}}=0$. The condition $Y_{\mathrm{c}}=2D/3\sigma$ (equation (9)) is used throughout, and enables P_{F} to be plotted together with P_{T} and P_{M} against D although D has no direct meaning for false output.

Figure 4 shows the results of these simulations for two window widths plotted against the quantity $D\sqrt{W}$. The curves coincide almost to the accuracy of the simulations, suggesting that universal curves for each probability exist at least in the limit of large W. In other words, a function Φ_{T} exists such that the detection efficiency is given by

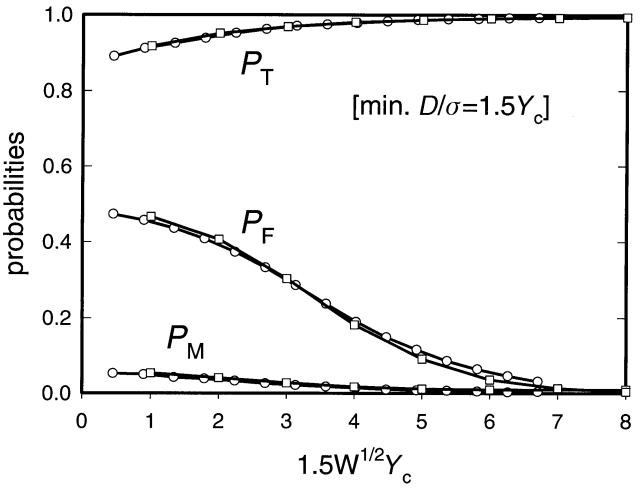

Figure 4. Sampled probabilities of true, false and multiple output events as binned over one window width, for step sizes $D/\sigma=0.1(0.1)1.5$ and $W=20$ (circle) and 100 (square). The threshold is related to step size by $Y_{\mathrm{c}}=2D/3\sigma$. When plotted against $Y_{\mathrm{c}}\sqrt{W}$ the curves superimpose to within sampling error. With $\mathcal{N}$ steps, the estimated standard deviation for each probability P is $(P(1-P)/\mathcal{N})^{1/2}$ which is less than 0.5% for $\mathcal{N}=10\,000$. This number of steps was contained in many iterations (40 and 200 for $W=20$ and 100, respectively) of waveforms of 20 000 data points with different random-number calls.

$$P_{\mathrm{T}}(D,W,\sigma)\approx\Phi_{\mathrm{T}}\left(\frac{D}{\sigma}\sqrt{W}\right) \qquad (W\gg 1). \tag{12}$$

The efficiency remains high at small step size because the threshold Y_{c} is scaled down in proportion. The penalty for doing this is an increased probability of false output, expressed per time bin of width E. However, the situation for large steps can be recovered by a suitable increase in window width. For example, steps of size $D=\sigma$ are detected with 95% efficiency and an error rate of 15% per bin from false steps and 2% from multiple steps if $E=W=20$. The same confidence limits apply to steps of size $D=\sigma/3$ if W is increased to $20\times 3^2=180$. If the bin width E is increased from W to $2W+1$, multiple steps are eliminated but the rate of false steps per bin increases by almost 20% (results not shown). As multiple events are already a very small source of error, it is better to use the smaller bin width. The key characteristics of the edge detector are contained in this figure.

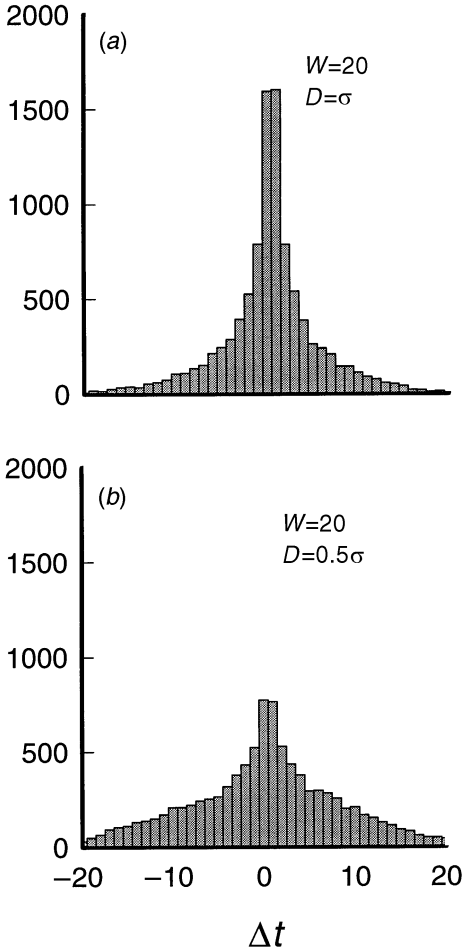

Figure 5. Histograms of the errors Δt in location times of detected steps classified as true output (within one window width of the input step), for $W=20$ and two step sizes generated as for figure 4. Better localization as a fraction of W is achieved with larger windows (results not shown). There are nearly 10 000 steps in each histogram.

Figure 5 gives sample histograms for the location errors of true output events. They show a significant central peak at zero time error, plus wings localized in the range $(-W,W)$. The standard error is of order $\sqrt{W}$, which increases less rapidly than W. As the step size is reduced, the central peak becomes less dominant and the proportion of events in the wings increases. Thus most location errors stay significantly smaller than the bin width, even for small steps.

What changes in performance can be expected for short pulses rather than single edges? The averaging operations intrinsic to this form of edge detection imply that the signal component $\Delta(t)/\sigma$ of the output is reduced when the pulse width τ is less than W (figure 2*e*), so the most probable value of the peak output is reduced as an averaging window overlaps both edges. In this situation, the detection efficiency can be kept high by scaling down the threshold. For figure 6, this was done by choosing

$$Y_c = \frac{\tau}{W}\frac{2D}{3\sigma}, \qquad (\tau < W), \tag{13}$$

which is equivalent to using equation (9) with D replaced by $D_{eq}=(\tau/W)D$ for $\tau<W$ and $D_{eq}=D$ otherwise. Hence

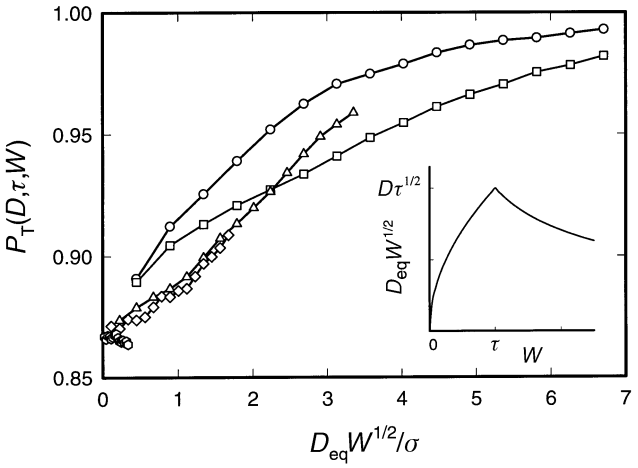

Figure 6. The detection efficiency (probability of true binned output events) for pulse widths $\tau=80$ (circle), 20 (square), 10 (triangle), 5 (diamond) and 1 data point (circle), various step sizes and $W=20$. Results are plotted against $D_{eq}\sqrt{W}$ where the equivalent step size $D_{eq}=\tau D/W$ for $\tau<W$ or D otherwise, and the output threshold was set at $2D_{eq}/3\sigma$ in all cases. The proximity of the curves show that scaling with respect to $D\sqrt{W}$ extends approximately to short pulses when D is replaced by D_{eq} and the threshold lowered accordingly. False steps are minimized by maximizing the abcissa with respect to W, which occurs when $W=\tau$ (see inset). Results for $W=100$ scale similarly.

the scaling relation (12) for the detection efficiency of a single edge generalizes to pulses of width τ

$$P_T(D,W,\sigma,\tau) \approx \Phi_T\left(\frac{D_{eq}}{\sigma}\sqrt{W}\right), \qquad (W \gg 1), \tag{14}$$

giving a universal curve as a function of $D_{eq}\sqrt{W}/\sigma$. This prediction is imperfectly realized because of binning effects (figure 6), but the curves for different pulse widths overlap sufficiently that the approximation of a single universal curve is good enough for practical purposes.

By lowering the threshold for short pulses below that in equation (11), the detection efficiency remains high. The parameter $D_{eq}\sqrt{W}/\sigma$ should then be set as large as possible to minimize the occurrence of false output, as in figure 4. For a pulse of given size and duration, this must be done by changing the window width. The inset of figure 6 shows that this parameter reaches a maximum value of $D_{eq}\sqrt{\tau}/\sigma$ when

$$W = \tau, \tag{15}$$

which is the condition for optimal detection of pulses of width τ. This occurs because windows larger than τ give reduced output from one window overlapping both edges of the pulse, while smaller ones give less filtering and more output noise.

The performance of the detector also varies with the method of taking averages and variances. Equation (1) takes an unweighted average in each window, but it has been suggested that averages weighted away from the central sampling point (Canny 1986), or towards it (Castan *et al.* 1990), are better in some sense. There may also be an advantage in using median averaging which preserves the shape of an edge as the windows pass through (Gallagher & Wise 1981). These alternatives have

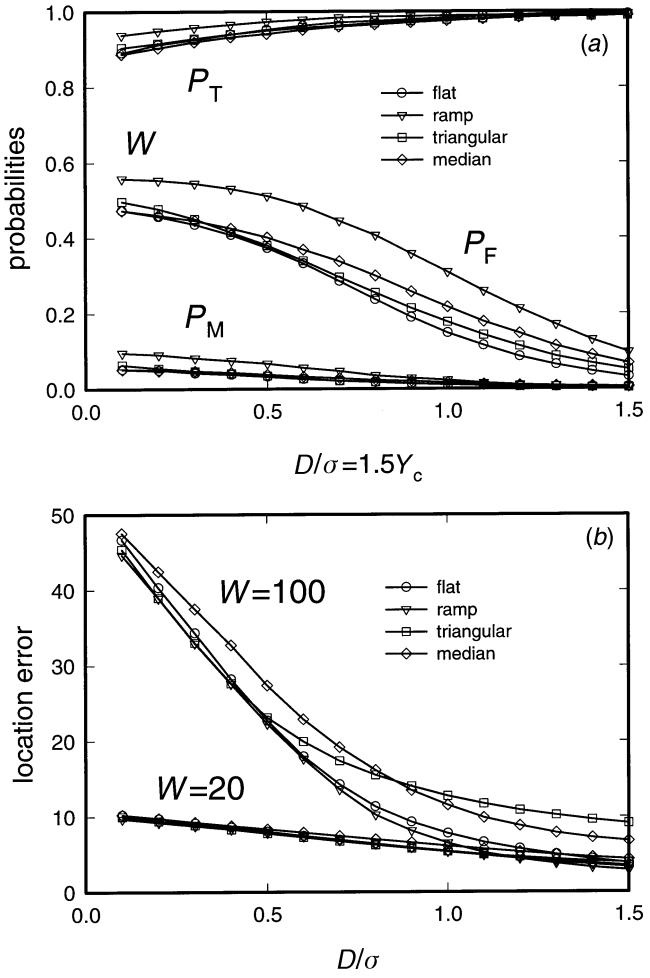

Figure 7. (*a*) Comparative performances for sampled output probabilities and (*b*) the standard location error for true output, by using different window-averaging functions f_k such that

$$X_{i\pm} = \sum_{k=1}^{W} f_k X_{i\pm k},$$

also median averaging. Simulations were made as for figure 5. All plots in (*a*) are for $W=20$. The averaging function f_k equals $1/W$ (flat), $2(W-k)/W^2$ (ramp) or the two-piece function $4k/W$ for $k<\mathrm{W}/2$ and $4(W-k)/W^2$ for $W/2<k<W$ (triangular). These functions are normalized to unity over the window when $W \gg 1$.

been investigated numerically by using ramp and triangular functions for window averaging, and with median averaging for X_{i+} and X_{i-} (but not for the variances). Results for the case $E=W$ are shown in figure 7 and rank as follows. With the same window width, there is little difference between most methods using binned thresholded output. Ramp averaging, which favours output closer to the input step, has the highest detection efficiency (figure 7*a*) and the best location accuracy (figure 7*b*), but also a significantly higher rate of false output than the others and twice as many multiple detections. Its overall performance cannot be improved by changing the bin width; with $E=1.2W$, multiple output is halved, but the rate of false output per bin is even higher. Median averaging in conjunction with the switching method yields no improvement, and can give larger location errors for small steps. It appears that the

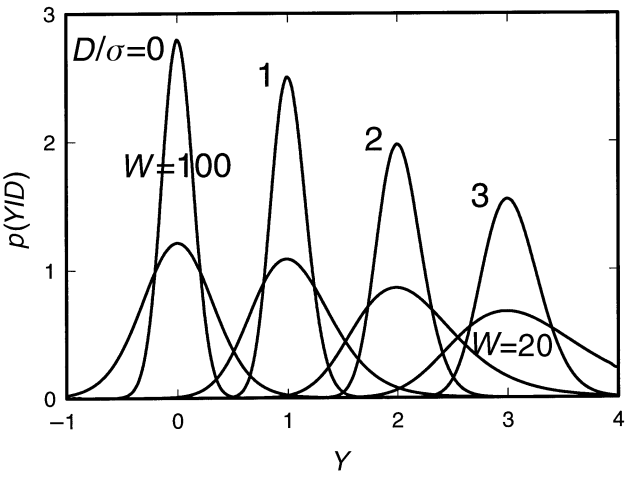

Figure 8. Distributions of the output variable Y as calculated from equations (7 and 8) for the ideal edge detector, for window widths and step sizes relative to noise as shown. The integral in equation (8) was computed recursively from equation (A6). Note that the selectivity of the responses to steps ($D>0$) and noise ($D=0$) increases with W.

best performance is achieved by using uniform averaging, which has the smallest rate of false output.

The switching edge detector of this paper uses window averages X_{i+}, X_{i-} at time t_i in the second equation of (1) to calculate variances in the forward and backward windows. This gives better results than using displaced averages X_{i+k}, X_{i-k}, as was done by Chung & Kennedy, and also allows the distribution of the output function (2) to be calculated from first principles for edges in Gaussian white noise.

5. USAGE

A procedural problem with the predictors of figure 4 is that their values are dependent on the size of steps in the time-series and the durations of any short-lived stepped regions: neither of these may be known in advance, but this information is required to choose optimal settings of the threshold output Y_c and the window width W. For any given Y_c and W, each above-threshold peak in the output is automatically interpreted as a step. If this step is present in the input, figure 8 shows that a peak of height Y should correspond to a step size D approximately equal to σY where σ is the standard noise deviation. If the peak height is 50% higher than the threshold ($Y>1.5Y_c$), the middle curve in figure 4 gives the probability that this peak is false. For bigger peaks the curve is an upper estimate. However, the detector is instructed to count all peaks in the range $Y>Y_c$ as detected edges, and for those with $Y_c<Y<1.5Y_c$ the chance that the peak arises from noise is higher and its interpretation is in doubt.

To proceed further would require estimates for the size of each detected step. This can be done from the output list of step times, for example by averaging the input between adjacent steps if the input signal between adjacent steps is constant. Thus, doubtful peaks could be identified and confidence limits supplied by constructing design curves for values of the ratio $D/\sigma Y_c$ below 1.5.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

A simpler approach is to run the detector with a sequence of increasing window widths to reduce the probability that these peaks are false (figure 4); this method is recommended as a way of confirming the original output list when a range of step sizes is expected. For example, with the optical trap data in figure 1*a*, runs with D/σ set at 1.0 and $W=20(10)100$ gave convergent numbers of detected steps only for $W\geqslant 80$ (four up, four down) with essentially the same location times (differing by 15 time points at most), confirming that steps of this size are sparse. This procedure could then be repeated with a lower threshold to reveal smaller steps. In this case, the output list should not be expected to confirm the original detection of large steps, because the output noise level is higher and false peaks may delete true peaks after binning. Finally, time resolution is limited by the window width; a pulse containing less than W data points gives less output than two well-separated steps of the same magnitude and may not be detected unless the threshold is lowered again. If output lists fail to converge at large W, the detected steps may not be sparse on this time-scale.

Repeated averaging is possible by using the filtered output function $X_{\mathrm{out}}=g_+X_+ + g_-X_-$ defined by Chung & Kennedy (1991) as input for a second pass. This has not been studied in detail, but may be a useful way of reducing false output.

To summarize, for sparse edges the threshold should be set at two-thirds the smallest edge size of interest, and the window width for the desired level of reliability. The Fortran program in Appendix B incorporates this rule of thumb by asking for a minimum edge size.

6. DISCUSSION

The switching technique of Chung & Kennedy (1991) has been applied to edge detection by defining an output function of known statistical distribution which peaks near edges in the input waveform, leading to logical output for detected steps and their locations. Statistical predictors for its efficiency, reliability and accuracy in locating edges are generated for sharp edges in Gaussian white noise input. Hence it should be an appropriate tool for detecting force or displacement steps from single motor proteins or single-channel ion currents from membranes, where the noise arises from Brownian motions and the steps from single chemical-kinetic events. Its efficiency and reliability have not been tested with other kinds of noise input. Detectors based on window averaging and classical variance constructs may be sensitive to 'wild' points in the data, which can be avoided by using median averaging (Kirlin & Moghaddamjoo 1986). Although the modified switching edge detector is of this kind, the existence of a threshold makes it insensitive to 'wild' points in the data, which are equivalent to very short pulses and not efficiently detected unless special measures are taken as discussed.

In §5 there is a practical, but somewhat makeshift, approach to confidence limits for assigning steps to output events when the step sizes are not known. A better approach is to use Bayesian methods to estimate the distribution of step sizes from a given peak output. A practical solution lies beyond the scope of this paper, but a start can be made from the formula

$$p(D|Y)=\frac{p(Y|D)p(D)}{\int p(Y|D)p(D)\mathrm{d}D}, \tag{16}$$

in terms of the distribution of equation (A5) and the distribution $p(D)$ of step sizes from an ensemble of time-series. Of course the latter distribution is not known, but recursive methods for generating it from the output should be possible.

Alternative approaches to edge detection exist which avoid estimating the likelihood of individual edge events. They require a known mechanism for generating all edges in the time-series, for example, a Markov process giving kinetic transitions between discrete states. In that case, the parameters of the mechanism (rate constants, for a Markov process) can be sought by maximizing the likelihood of the whole time-series (see, for example, Horn & Lange 1983; Rabiner 1989; Fredkin & Rice 1992). However, in practice the mechanism may not be fully known or competing mechanisms may be at work. For example, the application to displacement steps from single motor proteins is complicated by overlapping events which generate waveforms with ramps as well as steps (Finer *et al.* 1994), and may change the operation of the force mechanism. Slow ramps can be observed in the filtered output function $X_{\mathrm{out}}=g_+X_+ + g_-X_-$ defined by Chung & Kennedy (1991), but not in the function Y. However, a localized ramp between regions of constant signal will appear as an edge at low time resolution and can be detected as such after filtering out the more rapid components.

I thank Dr Jimmy Boyce and Dr Darren Toulson of the imaging group of the Department of Physics, King's College London for a helpful discussion, and Alberto Trombetta of the Randall Institute for permission to display a fragment of his trap data using the muscle proteins. This work was supported by a grant from the Wellcome Trust.

APPENDIX A. THE IDEAL DISTRIBUTION OF *Y*

Equation (4) defines the quantity Y as the difference of two averages of an input with signal and noise divided by the variance of the noise only. It is convenient to separate the numerator into signal and noise components, giving

$$Y=\frac{X_+ - X_- + \Delta}{S}, \tag{A1}$$

where Δ is the signal component as discussed in §2, and $X_\pm$ and S^2 are now the averages and variance of Gaussian white noise over windows of width W. Note that X_+ and X_- are not identical since they are calculated in different windows. The distributions of these noise variables follows by combining Gaussian distributions of variance σ^2 for the individual variates x_i.

The derivation can be summarized in terms of gamma variates as described by Weatherburn (1968). Each variate $\gamma=x^2/2\sigma^2$ has the distribution $(\pi\gamma)^{-1/2}\exp(-\gamma)$. The quantity $x+\Delta$ has a shifted Gaussian distribution and so the quantity $\gamma=(x+\Delta)^2/2\sigma^2$ has a shifted gamma distribution given by

$$\Phi(\gamma|\Delta) = \frac{1}{(\pi\gamma)^{1/2}}\exp(-\gamma - b^2)\cosh(2b\gamma^{1/2}), \left(b = \frac{\Delta}{\sqrt{2}\sigma}\right). \tag{A2}$$

Consider the numerator of (A1). The quantity $X_+ - X_-$ is $1/W$ times the sum or difference of $2W$ independent Gaussian variates of zero mean and unit variance, and therefore has zero mean and variance equal to $2\sigma^2/W$. Hence the quantity

$$\gamma = \frac{W}{4\sigma^2}(X_+ - X_- + \Delta)^2, \qquad \left(b = \frac{\Delta\sqrt{W}}{2\sigma}\right), \tag{A3}$$

is distributed as in equation (A2). This accounts for the square of the numerator in (A1).

The square of the denominator is the variance associated with just one window, and its mean value is just σ^2/W. Hence, by well-known arguments (Weatherburn 1968), the quantity $WS^2/2\sigma^2$ is the sum of $W-1$ unshifted gamma variates and is itself a gamma variate of degree $m=(W-1)/2$ with the distribution $\phi_m(\gamma)=\gamma^{m-1}\exp(-\gamma)/\Gamma(m)$.

The quotient of these two quantities is equal to $Y^2/2$. It can be shown that the numerator and denominator are statistically independent and hence the distribution of the quotient can be calculated from the joint distribution, which is factored. The details follow the analogous calculation for the unshifted case $\Delta=0$, which yields Student's distribution for the quantity $t=((W-1)/2)^{1/2}Y$ with $W-1$ degrees of freedom (Weatherburn 1968). In this way, the distribution function of the quantity $v=Y^2/2$ is found to be

$$\frac{\exp(-b^2)}{\Gamma(1/2)\Gamma(m)}\frac{v^{-1/2}}{(1+v)^{m+1/2}}\int_0^\infty u^{m-1/2}\exp(-u)\cosh\left(2b\sqrt{\frac{uv}{1+v}}\right)\mathrm{d}u. \tag{A4}$$

To convert this result to a two-sided distribution for Y, the distributions of positive and negative values can be separated as for the case $\Delta=0$, giving

$$p(Y|\Delta) = \frac{1}{\sqrt{2\pi}}\frac{\exp(-b^2)}{\Gamma((W-1)/2)}\left(1+\frac{Y^2}{2}\right)^{-W/2}\int_0^\infty u^{W/2-1} \times \exp\left(-u + \frac{2bY\sqrt{u}}{\sqrt{2+Y^2}}\right)\mathrm{d}u, \tag{A5}$$

where $b=W^{1/2}\Delta/2\sigma$. This result is equivalent to the biased t-distribution (Patnaik 1949). With no step, equation (A5) collapses to a scaled version of the Student's t-distribution

$$p(Y|0) = \frac{1}{\sqrt{2\pi}}\frac{\Gamma(W/2)}{\Gamma((W-1)/2)}\left(1+\frac{Y^2}{2}\right)^{-W/2}, (-\infty < Y < \infty), \tag{A6}$$

where $\Gamma(x)$ is the gamma function.

The distribution (A5) is shown in figure 8 for a range of values of Δ and two values of W, setting $\sigma=1$. The most probable value of Y is close to Δ/σ. To discriminate edges of size D from noise, the distributions with $\Delta=0$ and $\Delta=D$ must be well-separated; for $W=20$ this is achieved when $D/\sigma \geqslant 1$ and for $W=100$ when $D/\sigma \geqslant 0.5$. Thus selectivity increases with the size of the averaging window.

The integral in (A5) can be computed recursively after rewriting in the form $\sqrt{\pi}\exp(z^2)F_{W-1}(-z)$ where $z=bY/\sqrt{(2+Y^2)}$ and

$$F_n(z) = \frac{2}{\sqrt{\pi}}\int_z^\infty (w-z)^n\exp(-w^2)\mathrm{d}w, \tag{A7}$$

is a repeated integral of the error function. It can be computed from the recursion relation

$$F_n(z) = (n-1)F_{n-2}(z)/2 - zF_{n-1}(z), \tag{A8}$$

where $F_0(z)=\mathrm{erfc}(z)$ and $F_1(z)=-zF_0(z)+\exp(-z^2)/\sqrt{\pi}$.

The moments of Y can be found more directly from the definition (A1), equivalent to $Y=Y_o+\Delta/S$ where Y_o arises from noise only and is distributed according to equation (A6), with zero mean and variance $2/(W-3)$. As stated above, $S=(2\gamma/W)^{1/2}\sigma$ where the distribution of γ is the function $\phi_m(\gamma)$ with $m=(W-1)/2$. Hence the first and second moments of $1/S$ are

$$\overline{(S^{-1})} = \left(\frac{W}{2\sigma^2}\right)^{1/2}\frac{\Gamma((W-2)/2)}{\Gamma((W-1)/2)}, \qquad \overline{(S^{-2})} = \frac{1}{\sigma^2}\frac{W}{W-3}. \tag{A9}$$

Thus the mean value is given by

$$\overline{Y} = \Delta\overline{(S^{-1})} \approx \frac{\Delta}{\sigma} \quad (W \gg 1), \tag{A10}$$

as stated in the main text, and the variance is

$$\delta Y^2 = \overline{Y^2} - (\overline{Y})^2 = \overline{Y_o^2} + \Delta^2\overline{(S^{-2})} - (\overline{Y})^2 \approx \frac{2}{W} + \frac{1}{2W}\left(\frac{\Delta}{\sigma}\right)^2, \qquad (W \gg 1). \tag{A11}$$

Hence for large steps ($\Delta \gg \sigma$) the corresponding standard deviation is much less than the mean when $W \gg 1$. For small steps ($\Delta/\sigma \ll 2$) the standard deviation is independent of step size and is less than the mean when $\Delta \gg (2/W)^{1/2}\sigma$.

APPENDIX B. COMPUTER PROGRAM

```
PROGRAM sed
C motil/edge
C Switching edge detector for noisy time series (July
   1997)
C- - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
C Reads X input file (ASCII format) for function X(t).
C Requests minimum step size relative to noise, window
   width.
C Output file sed.dat gives filtered output XX(t) and
   statistical output Y(t). Times and signs of edges listed in
   sed.out.
C- - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
   IMPLICIT REAL*8 (A-H,O-Z)
   IMPLICIT INTEGER*4 (I-N)
   INTEGER*4 E,R,W,PSI
   PARAMETER (NP=20001, KP=2000, R=50,
   EOW=1.0)
   CHARACTER*15 FNAME
```

```
      DIMENSION X(-151:NP+151),XP(NP+151),XM(-151:NP),
     XX(NP),Y(NP),K(KP),PSI(KP),KU(0:KP),KD(0:KP)
      WRITE(6,*) 'Enter name of data-file (including .DAT)'
      READ(*,1) FNAME
    1 FORMAT (A)
      OPEN(1,FILE=FNAME,STATUS='UNKNOWN')
      DO J=1,NP
      READ(1,*,END=2) X(J)
      END DO
    2 CLOSE(1)
      N=J-1
      WRITE(6,*) 'N=',N
      WRITE(6,*) 'Enter no. of input lines (<N)'
      READ(6,*) NT
      WRITE(6,*) 'Enter (min. step size)/noise, window
     width'
      READ(6,*) D,W
      E=INT(EOW*W)
      FW=FLOAT(W)
      YC=0.6666667*D
      WRITE(6,*) 'YC=',YC
C- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
C Extend time series by W points at each end.
      DO I=1,W
      X(1-I)=X(I)
      X(N+I)=X(N-I)
      END DO
C- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
C Form pre-and post-averages XP(J), XM(J).
      DO J=1,N
      XAP=0.0
      XAM=0.0
      DO I=1,W
      XAM=XAM+X(J-I)/FW
      XAP=XAP+X(J+I)/FW
      END DO
      XP(J)=XAP
      XM(J)=XAM
C also running variances and switch factors to get XX,Y
      SP=0.0
      SM=0.0
      DO I=1,W
      SM=SM+(X(J-I)-XM(J))**2/FW
      SP=SP+(X(J+I)-XP(J))**2/FW
      END DO
C Form switching functions, etc
      RSP=SP**R
      RSM=SM**R
      GM=RSP/(RSP+RSM)
      GP=RSM/(RSP+RSM)
      XX(J)=GP*XP(J)+GM*XM(J)
      Y(J)=(XP(J)-XM(J))/SQRT(GP*SP+GM*SM)
      END DO
C- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
C Search for maxima/minima in Y and decimate smaller
C extrema with separations less than E.
      LU=0
      LD=0
      DO J=2,NT-1
      YJ=Y(J)
C Maxima first. . . ..
      IF (YJ.GT.YC) THEN
      YP=Y(J+1)
      YM=Y(J-1)
      IF (YJ.GT.YP.AND.YJ.GT.YM) THEN !new
     maximum
C Now test for adjacent maximum, and delete the lower one.
      IF (LU.EQ.0.OR.(LU.GE.1.AND.(J-KU(LU)).GT.E))
     THEN LU=LU+1 !no old maximum within E points
     ELSE !old maximum exists
      IF (YJ.LT.Y(KU(LU))) GO TO 3 !reject new
     maximum
C (if false, overwrite old one by leaving LU unchanged)
      END IF
      KU(LU)=J
    3 CONTINUE
      END IF
      END IF
C Now minima. . . ..
      IF (YJ.LT.-YC) THEN
      YP=Y(J+1)
      YM=Y(J-1)
      IF (YJ.LT.YP.AND.YJ.LT.YM) THEN !new minimum
C Now test for adjacent minimum, and delete the higher
one.
      IF (LD.EQ.0.OR.(LD.GE.1.AND.(J-KD(LD)).GT.E))
     THEN LD=LD+1 !no old minimum within E points
     ELSE !old minimum exists
      IF (YJ.GT.Y(KD(LD))) GO TO 4 !reject new
     minimum
C (if false, overwrite old one by leaving LD unchanged)
      END IF
      KD(LD)=J
    4 CONTINUE
      END IF
      END IF
      END DO
C- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
C Put up/down steps into K(L),PSI(L)
      DO L=1,LU
      K(L)=KU(L)
      PSI(L)=1
      END DO
      DO L=LU+1,LU+LD
      K(L)=KD(L-LU)
      PSI(L)=-1
      END DO
      L=LU+LD
      WRITE(6,*) 'L,LU,LD=',L,LU,LD
C- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
      LL=L
      OPEN(2,FILE='sed.dat',STATUS='UNKNOWN')
      DO J=1,NT
      WRITE(2,5) XX(J),Y(J)
    5 FORMAT(2(1X,E14.6))
      END DO
      CLOSE(2)
      OPEN(2,FILE='sed.out',STATUS='UNKNOWN')
      DO L=1,LL
      WRITE(2,6) K(L),PSI(L)
    6 FORMAT(I7,2X,I5)
      END DO
      CLOSE(2)
C- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
      STOP
      END
```

REFERENCES

Block, S. M. & Svoboda, K. 1995 Analysis of high-resolution recordings of motor movement. *Biophys. J.* **68** (suppl.), 230–241.

Canny, J. 1986 A computational approach to edge detection. *IEEE Trans. Patt. Anal. Mach. Intell.* **8**, 679–698.

Castan, S., Zhao, J. & Shen, J. 1990 Optimal filter for edge detection: methods and results. *Lect. Notes Comput. Sci.* **427**, 13–17.

Chung, S. H. & Kennedy, R. A. 1991 Forward–backward non-linear filtering technique for extracting small biological signals from noise. *J. Neurosci. Meth.* **40**, 71–86.

Finer, X., Simmons, R. M. & Spudich, J. A. 1994 Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368**, 113–119.

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

Gallagher, N. C. & Wise, G. L. 1981 A theoretical analysis of the properties of median filters. *IEEE Trans. Acoust. Speech Signal Process.* **29**, 1136–1141.

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

Kassam, S. A. & Poor, H. V. 1985 Robust techniques for signal processing. *Surv. Proc. IEEE* **73**, 433–481.

Kirlin, R. L. & Moghaddamjoo, A. 1986 A robust running-window detector and estimator for step-signals in contaminated Gaussian noise. *IEEE Trans. Acoust. Speech Signal Process.* **34**, 8160–8823.

Mosteller, F. & Tukey, J. W. 1977 *Data analysis and regression.* Reading, MA: Addison-Wesley.

Oppenheim, A. V. & Schafer, R. W. 1989 *Discrete-time signal processing.* New Jersey: Prentice-Hall.

Patnaik, P. B. 1949 The non-central chi-squared and F-distributions and their applications. *Biometrika* **36**, 202–232.

Peretto, P. 1992 *An introduction to the modelling of neural networks.* Cambridge University Press.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. R. 1992 *Numerical recipes in Fortran*, 2nd edn. Cambridge University Press.

Rabiner, L. R. 1989 A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**, 257–285.

Sakmann, E. & Neher, E. 1983 *Single-channel recording.* New York: Plenum.

Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. 1993 Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727.

Weatherburn, C. E. 1968 *A first course in mathematical statistics.* Cambridge University Press.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES