Accuracy of Tetrode Spike Separation as Determined by Simultaneous Intracellular and Extracellular Measurements

KENNETH D. HARRIS, DARRELL A. HENZE, JOZSEF CSICSVARI, HAJIME HIRASE, AND GYÖRGY BUZSÁKI

Center for Molecular and Behavioral Neuroscience, Rutgers, The State University of New Jersey, Newark, New Jersey 07102

Harris, Kenneth D., Darrell A. Henze, Jozsef Csicsvari, Hajime Hirase, and György Buzsáki. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J Neurophysiol* 84: 401–414, 2000. Simultaneous recording from large numbers of neurons is a prerequisite for understanding their cooperative behavior. Various recording techniques and spike separation methods are being used toward this goal. However, the error rates involved in spike separation have not yet been quantified. We studied the separation reliability of "tetrode" (4-wire electrode)-recorded spikes by monitoring simultaneously from the same cell intracellularly with a glass pipette and extracellularly with a tetrode. With manual spike sorting, we found a trade-off between Type I and Type II errors, with errors typically ranging from 0 to 30% depending on the amplitude and firing pattern of the cell, the similarity of the waveshapes of neighboring neurons, and the experience of the operator. Performance using only a single wire was markedly lower, indicating the advantages of multiple-site monitoring techniques over single-wire recordings. For tetrode recordings, error rates were increased by burst activity and during periods of cellular synchrony. The lowest possible separation error rates were estimated by a search for the best ellipsoidal cluster shape. Human operator performance was significantly below the estimated optimum. Investigation of error distributions indicated that suboptimal performance was caused by inability of the operators to mark cluster boundaries accurately in a high-dimensional feature space. We therefore hypothesized that automatic spike-sorting algorithms have the potential to significantly lower error rates. Implementation of a semi-automatic classification system confirms this suggestion, reducing errors close to the estimated optimum, in the range 0–8%.

INTRODUCTION

Most knowledge about the physiological function of the brain is based on sequential analysis of single-site recordings. Although it has long been recognized that the computational power of complex neuronal networks cannot fully be recognized by studying the properties of single cells or the activity of a few selected sites, experimental access to the emergent properties of cooperating neurons has largely been impossible until quite recently. Direct investigation of the temporal dynamics of neuronal populations can only be based on simultaneous observation of large neuronal aggregates (Buzsaki et al. 1992; Wilson and McNaughton 1993). The two critical requirements for achieving this goal are placing a large number of recording electrodes in a small amount of tissue without significant tissue damage and efficient isolation of action potentials emanating from individual neurons. Action potentials generated by neurons ("unit" or "spike" activity) can be monitored by extracellular glass pipettes, single etched (sharp) electrodes, or multiple-site probes. The closer the electrode to the soma of a neuron, the larger the size of the extracellularly recorded spikes. Whereas glass pipettes and high-impedance sharp metal electrodes can be used to monitor the activity of a single cell (Evarts 1968; Georgopoulos et al. 1993), these electrodes are not always practical in freely moving animals because small movements of the electrode can damage the neuron and because isolation of large numbers of neurons with independently moving drives is difficult (Kruger and Aiple 1988; Llinas and Sasaki 1989; McNaughton et al. 1996). Larger size wires can record the activity of multiple neurons and provide better mechanical stability because the electrode tip is not placed directly against the cell membrane. The use of multiple recording channels from the same neuron(s) provides improved methods for single-unit sorting, based on the temporal coherence of spikes across channels (Drake et al. 1988; Gray et al. 1995; McNaughton et al. 1983; Nadasdy et al. 1998; Recce and O'Keefe 1989; Wilson and McNaughton 1993). The most widely used multiple-site probe, the wire "tetrode" (Recce and O'Keefe 1989; Wilson and McNaughton 1993), separates neurons based on their spatial location, assuming that neurons are point sources of action potential-associated currents.

There are three stages between the recording of extracellular unit activity and the identification of spikes representing the activity of a single neuron. The first stage is spike detection, in which the electrical activity measured on the electrodes is used to derive the times corresponding to extracellular spikes. This is usually achieved by high-pass filtering followed by thresholding and may be done by hardware or software. The second stage is feature extraction. During this stage, a feature vector (i.e., an array of quantitative parameters) is calculated for every spike. In the simplest cases, the feature vector represents the amplitude of the spikes recorded by the four tetrode sites. More advanced methods quantify additional information about the spikes, such as waveshape and discharge pattern. Waveshapes may be quantified by measuring parameters such as spike

---

width, or by principal component analysis (Abeles and Goldstein 1977). The third stage is "clustering" of spikes. In this stage spikes with similar feature vectors are grouped into clusters, assumed to represent the spikes of a single neuron. In most laboratories, the clustering stage is done manually, with a graphical user interface that displays scatter plots in feature space and allows the operator to separate the clusters by drawing polygons or straight lines around them (Gray et al. 1995; Rebrick et al. 1997; Skaggs and McNaughton 1996; Wilson and McNaughton 1993; Wood et al. 1999). This process is time consuming, and may be affected by subjective factors. It has been proposed that automatic methods may significantly speed up the clustering process, and reduce the effect of subjective factors (Fee et al. 1996; Lewicki 1998; Sahani 1999; Wright et al. 1998).

A major problem with all currently used spike separation methods is that their reliability cannot be verified with independent methods, and thus errors inherent in subjective clustering by the human operator cannot be measured quantitatively. Errors can occur because spikes belonging to different neurons are grouped together (false positive, Type I, or commission error) or because not all spikes emitted by a single neuron are grouped together (false negative, Type II, or omission error). In this paper, we quantified the error rates of spike separation methods. This was achieved by simultaneous recording from the same neuron intracellularly with a glass pipette and extracellularly with a tetrode (Henze et al. 2000; Wehr et al. 1998). This approach allowed us to reveal the major causes of error in the spike separation process.

METHODS

Simultaneous intracellular-extracellular recording in anesthetized rats

Description of the surgical and implantation methods are described in the accompanying paper (Henze et al. 2000).

Spike detection

The continuously recorded wideband signals were digitally high-pass filtered (Hamming window-based finite impulse response filter, cutoff 800 Hz, filter order 50). Spike detection was achieved by a previously described system (Csicsvari et al. 1998). Briefly, the root-mean-square power of the filtered signal was computed using a sliding window. The mean and standard deviation of the power were computed, and spikes were extracted when the power exceeded a threshold derived from the standard deviation from the baseline mean. The spikes were then upsampled using the sampling theorem, and the waveforms were realigned by peak position.

Feature vector extraction

Feature vectors were extracted from the spike waveforms in two ways. In the first method, the feature vector consisted of the peak-to-peak (i.e., maximum-to-minimum) amplitudes of the resampled filtered waveforms. In the second method, we employed a principal component analysis (PCA) to create feature vectors (Abeles and Goldstein 1977; Csicsvari et al. 1998). For each channel, the resampled waveforms of all spikes were pooled and the first three principal components of the waveform set were found. A 12-dimensional feature vector was then created for each spike by projecting the waveform on each channel onto each of the three principal components.

Manual cluster cutting

The feature vectors were clustered using a previously developed custom graphical interface program (gclust). This program allows the operator to select two feature vector components to produce a two-dimensional scatter plot for the chosen components. The operator then draws polygons in the two-dimensional space to assign spikes to individual clusters. By iteratively viewing different projections, the operator can refine the boundaries of clusters. The program also displays superimposed waveforms for each cluster, autocorrelograms for each cluster, and cross-correlograms for all pairs of clusters.

Correspondence between intracellular and extracellular traces

Intracellular action potentials were detected by the method described in the accompanying paper. As a function of the site of the impalement of the neuron (somatic vs. dendritic), the delay between the peak of the intracellular action potential and the negative peak of the wideband recorded extracellular spike varied as much as 1 ms (Henze et al. 2000). However, the delay was constant for a given set of intracellular and extracellular recordings. Some jitter of the extracellular peak was caused by the background activity. Therefore extracellular spikes corresponding to intracellular action potentials were identified if they occurred within 0.25 ms on either side of the peak of the extracellular spike, as predicted by the intracellular spike. The false positive error rate was determined as the percentage of spikes clustered by an operator that were not correlated by the simultaneous presence of an intracellular action potential. The false negative error rate was determined as the percentage of the intracellular spikes for which the operator's cluster did not contain a corresponding extracellular spike.

Best ellipsoid error rates

To estimate the optimal clustering performance for a given set of feature vectors, we defined a measure called the "best ellipsoid error rate" (BEER). This measure was designed to estimate the optimal clustering performance for a given set of feature vectors by searching over all possible ellipsoidal cluster boundaries. In the present context, "optimal" may be defined in several different ways since it is possible to give different weights to Type I and Type II errors. We therefore introduced a "conservatism" parameter and defined the optimal ellipsoid to be the one that minimized a cost function equal to the weighted averaged Type I and Type II errors, with weights specified by the conservatism parameter. By systematically varying this parameter, the BEER method produced a curve, illustrating the trade-off between Type I and Type II errors. The minimization of the cost function over the space of ellipsoidal cluster boundaries was performed by a two-layer neural network. The first or "input" layer contained one node for every element of the feature vector plus one node for each pairwise product of feature vector elements (so for an n-dimensional feature vector, this layer would have $n + n(n + 1)/2$ nodes). The second or "output" layer had a single logistic node which was trained to be 1 if the spike belonged to the identified cell and 0 otherwise. The network was trained by the "quickprop" algorithm (Fahlman 1988), modified to allow for differential weighting of Type I and Type II errors. After convergence, the weights of the network specified an optimal quadratic criterion for spike identification, i.e., an ellipsoidal cluster boundary. To reduce the training time of the network and to ensure that it did not converge to an incorrect local maximum, the initial values of the weights were specified by the ellipsoid whose center and axes were derived from the mean vector and covariance matrix of the feature vectors corresponding to intracellularly identified spikes.

Error rates were evaluated with a cross-validation method. The data were split into two halves consisting of even and odd spike numbers. The network was trained twice, once on each half of the data, and

tested on the other half. This method produced two sets of error percentages for each value of the conservatism parameter, i.e., two error curves. The network was deemed to have converged, and training was terminated when the percentage errors on the test set had not changed more than 0.1% within the last 15 training epochs. The final percentage error value was then taken to be the mean error value over these last 15 epochs. The computations were performed on an IBM SP2 station (IBM, Armonk, NY) with a custom written C++ program.

The BEER measure provides an estimate of the optimal performance of any clustering system, automatic or manual. It determines the location of the optimal cluster based on a knowledge of which spikes correspond to the identified cell (i.e., it performs supervised learning). This is only possible because of the simultaneous intracellular recording. Manual or automatic spike sorting systems will not have this knowledge (i.e., they must perform unsupervised learning). The BEER method therefore estimates an upper bound on the performance achievable by any unsupervised system. This upper bound will be approached by an automatic system based on an accurate probabilistic model in the limit of a large number of data points (Anderson 1984). It should be emphasized that the ellipsoidal cluster boundary used by the BEER measure is not necessarily optimal. Even for perfectly ellipsoidal clusters, the optimal cluster boundary will be given not by a single quadratic but by an intersection of quadratics. A search over all such boundaries is possible in theory but would require excessive computer time. Therefore the BEER measure may underestimate optimal possible performance.

Prediction of cluster covariance matrix from background activity

To compare the cluster shape to that predicted by a fixed spike shape superimposed with background noise, we extracted periods of background activity containing no extracellular spikes. These were determined by removing 32 samples around every extracellularly detected spike and retaining any remaining intervals of 100 samples or longer. The auto- and cross-covariance functions of the background activity were determined from these periods. These functions were used to create Toeplitz matrices for the time-domain covariance matrix between each pair of channels. The covariance between two principal components for the selected channels was then obtained by pre- and post-multiplying by the time-domain projection template used to generate the principal components. Analysis was performed in MATLAB (The MathWorks, Natick, MA).

Detection of sharp waves

Sharp waves were detected based on the occurrence of their associated high-frequency ripples. Broadband extracellular signals were filtered between 100 and 250 Hz using a digital finite impulse response filter, rectified, and smoothed using a median smoothing algorithm (LabView, National Instruments, Austin, TX). The occurrence of a ripple was determined based on a deflection of the processed extracellular signal >7 SDs from the baseline. The start and end of the ripple were then determined from the closest zero values preceding and following the threshold crossing.

Automatic cluster cutting

To gauge the practicability of automatic cluster cutting, we tested the program AutoClass (Cheeseman and Stutz 1996, http://ic-www. arc.nasa.goc/ic/projects/bayes-group/autoclass). The program was used with the multivariate normal mode, with two parameters changed from their default values (max_n_tries = 50, rel_delta_range = 0.01), to reduce running time. As AutoClass tended to overcluster the data, the output of the program was further examined by a human operator, using the same program used for manual cluster cutting. Nearby clusters with visible refractory period in the cross-correlogram were judged to correspond to a single cell, and merged. In one case, a cluster produced by the program showed a bimodal shape, and was judged to correspond to two cells. This was confirmed by the cross-correlogram of the two subclusters, which showed no refractoriness or burst shoulders.

RESULTS

Errors of human operators

We recorded from 33 neurons in 30 rats, both intracellularly and extracellularly (Henze et al. 2000). Of these, we selected six data sets from three neurons, which contained sufficient action potentials for statistically significant error analysis, and which represented various levels of difficulties in clustering related to amplitude and "burstiness."

Nine human operators were asked to cluster the feature vector sets for these six sessions. The error rates of the operators are shown in Table 1 and Fig. 1. There was a spread of error rates with some operators being more "conservative" (making more false negative errors but less false positive errors) and others being more "liberal" (making less false negative errors but more false positive errors). The same operators fell in similar portions of the scale on several data sets, indicating some individual bias. In general, neurons with lower amplitude spikes, burst patterns, or several simultaneously active neighbors were more difficult to cluster. The potential causes of the human errors are considered in detail in the following text.

CAUSES OF ERROR. *Similarity of spikes of different cells.* Figure 2, *A* and *B*, *insets,* illustrates the mean extracellular waveforms of an identified cell (*cell 2* in Table 1) together with the mean waveforms of spikes emitted by another cell recorded by the same tetrode. The latter neuron was not intracellularly recorded, and spikes of this cell were determined by our standard clustering method. Based on our cell classification criteria (Csicsvari et al. 1999), this second cell was also a pyramidal cell. The amplitude profile of the confounding cell across the four recording sites was very similar to the identified cell. This similarity is also visible on the cluster diagram of the first principal component (Fig. 2*A*). There is considerable overlap between the spikes of the identified neuron (red) and confounding unit (green). However, the wave shape of the confounding cell was different from the identified neuron. The most prominent difference was a larger "initial positivity" and a smaller late positive wave after the large negative deflection. These differences in waveform were revealed by the second principal component (Fig. 2*B*).

We assumed that the difference in waveshape between the two pyramidal cells was due to electrode placement relative to the somatodendritic axis. To test this assumption, we recorded units with a linear six-site silicon probe ("hexatrode") (Wise and Najafi 1991). Figure 2*C* shows the mean waveform for each of the six sites along the probe for a presumed pyramidal neuron, clustered in the usual manner. The waveform of the spike varied as a function of distance from the soma in the somatodendritic axis, supporting the hypothesis that the waveshape produced by a single cell can be very different at different recording sites. These features of hexatrode-recorded unit activity may be exploited for the improvement of clustering methods in the future.

TABLE 1. *Operator error rates*

| | Data Set | | | | | | | | | | | |
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
	FP	FN	FP	FN	FP	FN	FP	FN	FP	FN	FP	FN
Operator												
K	10.0	14.3	7.4	16.6	0.2	11.7	1.2	17.0	1.1	22.9	3.8	25.1
H	3.8	26.7	6.0	11.5	1.4	1.3	1.3	16.2	0.5	21.3	3.1	20.1
D	9.4	16.1	0.7	19.7	0.4	3.3	1.6	10.1	1.1	9.1	1.5	21.2
C			4.2	5.6	0.9	3.2	2.3*	67.3*	0.8	12.5		
J	6.4	16.4	6.1	14.7	0.2	8.8	15.9	20.3	0.8	19.2	11.6	23.5
G	7.2	14.4	23.9	6.3	1.1	5.9	43.4	14.2	0.5	16.5	X	X
R	7.3	27.0	20.3	8.5								
X	31.2	1.9	5.4	9.6	0.4	23.8	8.8	25.1	0.5	20.9	2.2	19.4
2					0.7	2.9	10.0	11.1				
Mean	10.8 ± 92	16.7 ± 8.5	9.2 ± 8.2	11.5 ± 5.0	0.7 ± 0.4	7.6 ± 7.4	11.7 ± 15.0	16.2 ± 5.2	0.8 ± 0.3	17.5 ± 5.1	4.4 ± 4.1	21.9 ± 2.4
Automatic system	8.0	5.9	0.2	5.3	0.0	3.5	1.4	5.9	0.2	4.0	4.3	7.5

Amplitudes (μV) and number of spikes for data sets, respectively, were 1: 50 and 1,952; 2: 76 and 2,462; 3: 117 and 690; 4 (bursting): 68 and 1,682; 5: 71 and 1,420; and 6: 59 and 1,010. Mean ± SD. FP, false positive, Type I, or commission error; FN, false negative, Type II, or omission error; X, a session where the operator did not identify the intracellular unit. *, a session where the operator divided the spikes of a bursting cell into two clusters corresponding to early and late spikes within the burst (excluded from mean ± SD calculation).

Variability of extracellular wave shapes from a single cell. The "clouds" identifying individual clusters (i.e., hypothesized neurons) were variable in shape and size. What determines the spread of points in a cluster? In other words, what is the probability distribution for the feature vectors of a cell in cluster space? A first assumption is that the feature vectors follow a multivariate normal distribution. One way to test this is by calculating the Mahalanobis distance of points from the cluster center (Gnanadesikan 1977; Johnson and Wichern 1992). Under the normal distribution assumption, the Mahalanobis distance of points from the cluster center will have a χ^2 distribution. The quantile-quantile plot in Fig. 3*A* compares the predicted and empirical distribution. Approximately 98% of the data points lie along the diagonal, confirming that these spikes are distributed normally. The remaining 2% of the

points, though, start to diverge, and at least eight points can be considered definite outliers. What is the cause of the variability of these outlying spikes? The waveshapes of the three most extreme outliers are shown in Fig. 3*B*, superimposed on the mean waveform for the whole cluster. A closer inspection of the individual spike waveforms suggests that the altered shape was caused by the presence of another cell's spike overlapping with the spike of the intracellularly identified neuron. The example illustrates that for "nonbursting" cells the clusters are approximately normal, with only a minority of the spikes as outliers. We could not divide the spikes into two groups (overlapping vs. nonoverlapping), most likely because the secondary cell(s) causing the spike variability may be any distance from the identified cell. Therefore the magnitude of distortion of any given spike may vary continuously from large values to
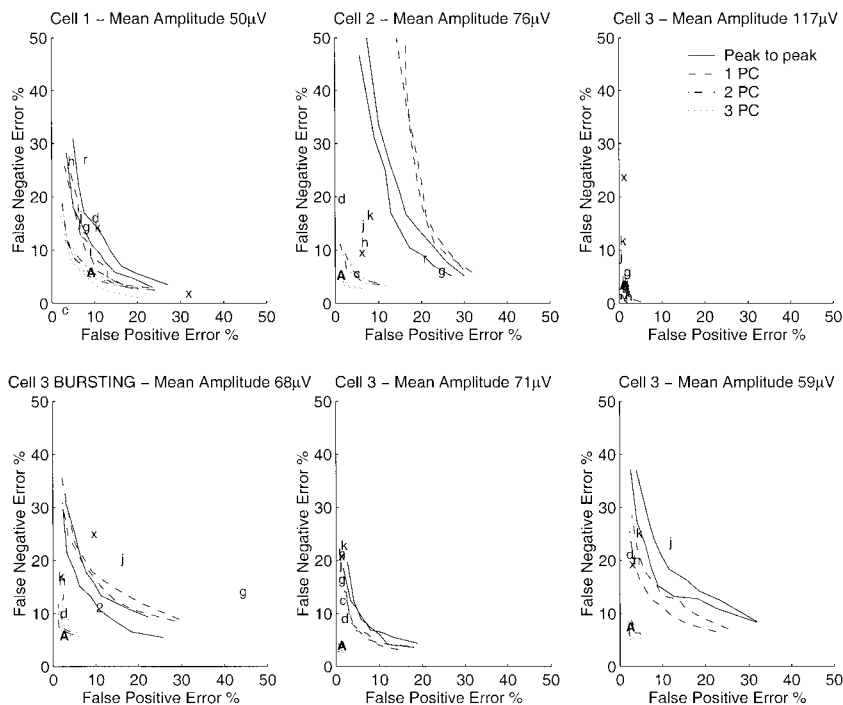

FIG. 1. Error rates for clustering. *x* axis: percentage of operator-marked spikes that do not correspond to the identified cell (Type I or commission error). *y* axis: percentage of identified spikes that were not marked by the operator (Type II or omission error). Each plot corresponds to a separate recording session. The last 4 sessions are the same cell recorded under different conditions of burst mode and amplitude. Individual human operators are identified by letters. The performance of the AutoClass program is indicated by A. The lines indicate theoretically optimal performance, determined by a computer search for the ellipsoidal cluster shape that minimizes a weighted average of false positive and false negative errors. PC, principal component.
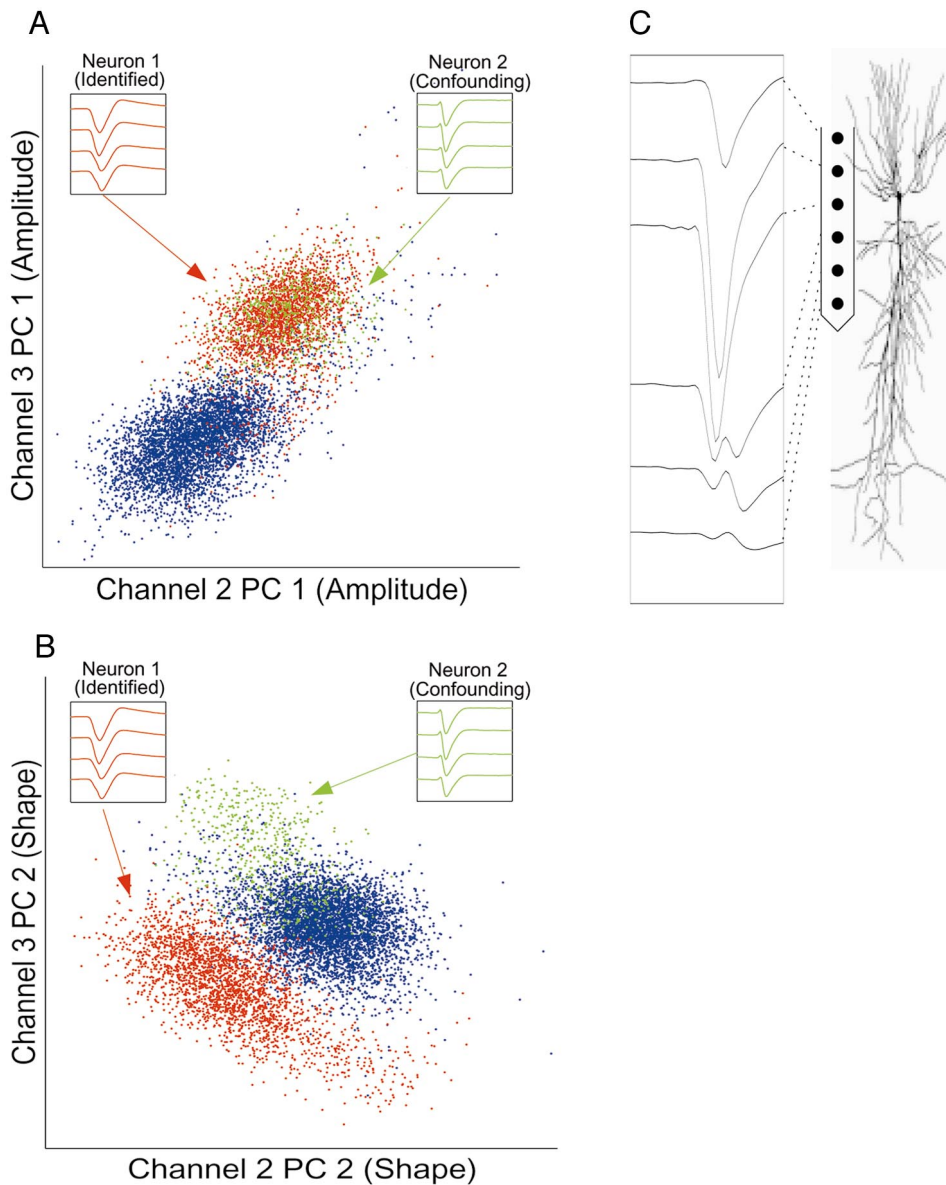
FIG. 2. Wave shape information is necessary for accurate spike discrimination. *A* and *B*: cluster plots showing the spikes of the identified cell (red) and a 2nd confounding cell with a very similar amplitude profile but different wave shape (green). All other extracellularly detected spikes shown in blue. When using amplitude information alone, the 2 cells cannot be separated. *C, left*: mean waveforms recorded by a linear silicon hexatrode from a pyramidal cell in a separate extracellular recording session. There is substantial variation of wave shape along the somatodendritic axis, indicating that electrode location is an essential determinant of wave shape. *Right*: hypothesized position of the hexatrode along the somatodendritic axis.

zero. This hypothesis would explain the slow trend away from the straight line (predicted values), instead of a distinct group of "overlapping spikes" (Fig. 3*A*).

In addition to overlapping spikes, what else determines the shape of the clusters? Possible variables include the effect of background field and unit activity and/or variability of the amplitude and shape of the intracellular action potential (Henze et al. 2000; Nadasdy et al. 1998). For a normal distribution, cluster shape is characterized by the covariance matrix. Figure 3*C* shows a pseudocolor image of the covariance matrix for a cluster of intracellularly identified spikes (after outlier removal). This can be compared with the predicted covariance matrix (Fig. 3*D*) under the hypothesis that the variability of spike shapes is caused purely by background field and unit activity (see METHODS). The two matrices are strikingly similar, indicating that for this nonbursting cell, the shape of the cluster is determined mainly by superimposed background field and unit activity.

EEG field and cellular synchrony. During sleep and awake immobility, the hippocampus from time to time expresses a

form of population activity called *sharp wave burst.* Sharp waves in the stratum radiatum are associated with a coherent 200-Hz field oscillation in the pyramidal cell layer ("ripples") and phase-locked discharges of pyramidal cells and interneurons (Buzsaki et al. 1992). In the anesthetized preparation, as used here, the ripple frequency is lower (120 Hz on average) and the incidence of population bursts is lower (Ylinen et al. 1995). We studied the effects of sharp wave-associated ripples on the error rates of unit clustering. Figure 4*A* shows the performance of human operators and the BEER measures (see following text) during epochs when the spikes occurred in association with field ripples. For this cell, the error rate increased dramatically during ripples up to 50% for Type I and Type II errors. The majority of the error increase was for Type I (false positive) errors, indicating the false inclusion of other units with the "target" unit. The minimum achievable error, as estimated by the BEER measure, also increased significantly. Even when all three principal components were considered, the combined errors could exceed 20–30%, reflecting a four- to fivefold decrease of spike isolation reliability for this cell
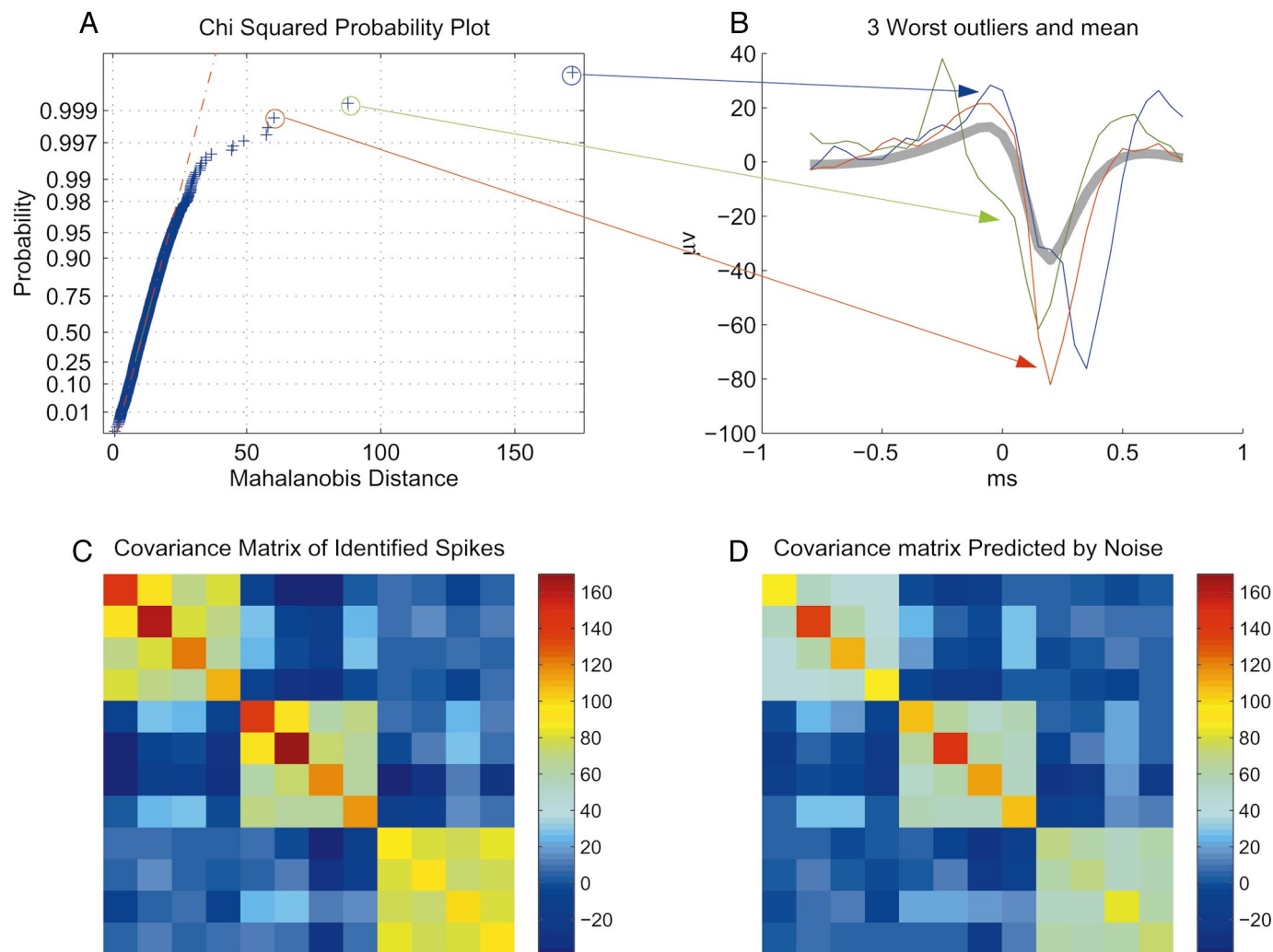
FIG. 3. Analysis of cluster shape for a nonbursting cell. *A*: χ^2 plot for the feature vectors for a single cell in 12-dimensional space (*data set 1* of Table 1). If the data are normally distributed, the Mahalanobis distance will follow a χ^2_{12} distribution (indicated by the red dot-dashed line). The inner 98% of the data fits this well. However the final 2% diverges, and there are 8 extreme outliers. *B*: waveforms of the 3 most extreme outliers, superimposed on the mean spike waveform for this cell (gray). The outliers appear to be caused by the overlap of a spike from the identified cell with another cell. *C*: pseudocolor plot of the 12×12 covariance matrix of the cluster of identified spikes. Column and row order index by principal component number (1–3) and channel number (1–4). *D*: covariance matrix predicted by the power spectrum of the noise. This is very similar to that seen in *C*, indicating that for this cell, the spread of the cluster is caused mainly by superimposed background noise. Color scales for *C* and *D*: arbitrary units.

during ripple-associated population synchrony of CA1 neurons. Examination of errors for all operators and data sets indicated that in the presence of field ripples, Type I errors increased more substantially than Type II errors (Fig. 4, *C* and *D*).

Figure 4*B* illustrates the mean wave shapes of the intracellularly identified cell. Although spikes detected during the ripple were phase-locked to a highly rhythmic "ripple" EEG background (not shown), the amplitude and waveshape of the average filtered spikes in the presence and absence of ripples were quite similar (Fig. 4*B*). However, the standard deviation was twice as high in the case of ripples compared with their absence. This finding indicates that digital filtering (>800 Hz) successfully removed the ripple field effect. We therefore conclude that the major factor contributing to errors during sharp wave-associated ripples is the superimposition of other synchronously firing neurons.

Complex spike bursts of single pyramidal cells. Single hippocampal pyramidal cells exhibit "complex spike" bursts (Ranck 1973), which cause variability of amplitude and waveform. To investigate the effects of complex spike activity on the extracellular clusters, we simulated regular bursting discharges by injection of short, strong depolarizing current steps (0.5 nA for 40 mS; intra-burst firing frequency ~200 Hz). Figure 5*A* shows the cluster diagram containing the bursting cell (red; *cell 4* in Table 1) and the averaged waveforms of the successive extracellular and intrasomatic spikes during the burst (Fig. 5*B*). The dominant effect on the cluster shape is the change in amplitude. The cluster of the identified cell is "stretched" compared with the clusters of other cells. The change in cluster shape was also reflected in the covariance matrix. Figure 5, *C* and *D*, shows a similar analysis as illustrated for the nonbursting cell in Fig. 3. For the bursting neuron, the cluster shape is no longer similar to that predicted
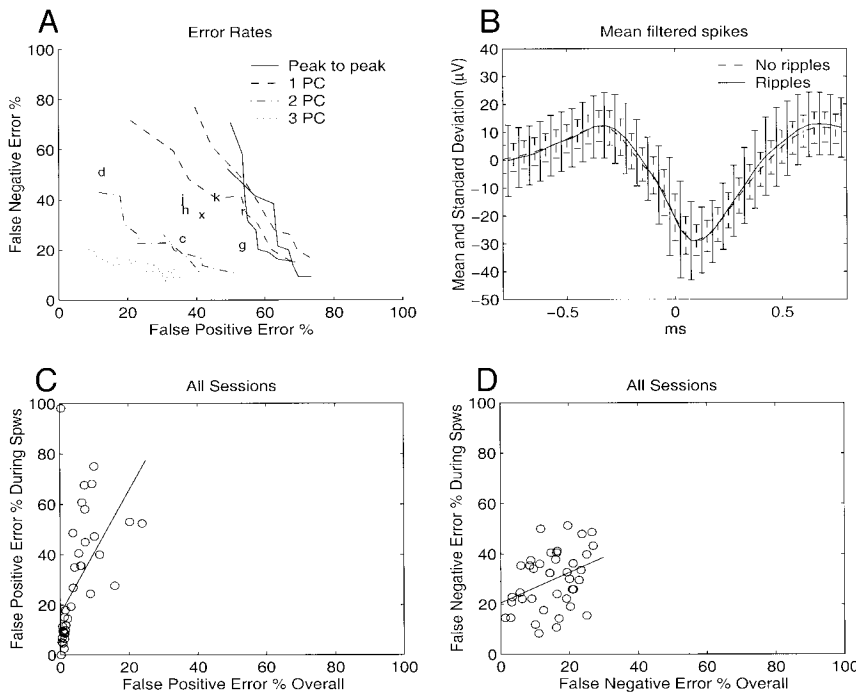
FIG. 4. Effect of cellular synchrony. *A*: operator error rates and theoretical optima for *data set 1,* restricted to hippocampal sharp wave burst (ripple) periods. Both operator error rates and theoretical optima increase dramatically relative to epochs not containing population bursts. *B*: mean and standard deviation of filtered waveforms of identified spikes during and outside of ripple periods. The mean waveform and amplitude are very similar but standard deviation increases. This indicates that the degraded performance is not due to contaminating synchronous field activity. *C* and *D*: scatter plots of overall error rates vs. error rates restricted to sharp wave periods. False positive errors are more seriously affected, suggesting that the main cause of error is the firing of otherwise silent cells.

by the background activity. Instead we see increased variance in amplitude, and a strong correlation between the amplitudes on all recording sites.

Postclustering tools: autocorrelograms and correlograms

It has been proposed that the neuron's spike refractory period, as reflected by the absence of spikes at short intervals in the autocorrelogram, can be used as a reliable indicator of whether the recording was made from a single cell or from multiple neurons. Conversely, cross-correlation of two clusters with a common refractory period is taken as an indication that the clusters actually represent the same cell (Fee et al. 1996; Johnston and Wu 1995).

Auto- and cross-correlograms provide valuable help during the clustering process. However, they should be interpreted with caution, due to the possibility of factors causing two cells not to co-fire (such as spatially separated place fields) or the presence of effective monosynaptic connections (Csicsvari et al. 1998). Comparison of autocorrelograms of clustered spikes and intracellular action potentials further helped to clarify the situations where correlograms may be safely used. Figure 6, *A* and *B,* illustrates the autocorrelogram of a poorly isolated cluster (Type I error: 43.4%). The presence of spikes in the refractory period (<2 ms) indicated the contamination of other units. However, the height of the autocorrelogram in the central (refractory) portion was quite small relative the large "shoulders" surrounding it. Does this mean that Type I (false positive) error was negligible? To get an insight why even in the presence of "confounding" spikes the autocorrelograms look "normal," we must consider the mathematical properties of correlograms in more detail.

To simplify the discussion, we scaled the *y* axis of the cross-correlogram to represent the "cross-intensity function" (Moore et al. 1970). This is achieved by dividing the number of counts in each bin by the product of the bin length and the total number of spikes of *cell 1*. The *y* axis can then be read in hertz, and the value at time *t* interpreted as the probability of *cell 2* firing per unit time, given that *cell 1* fired a spike a time *t* ago (Moore et al. 1970).

In the case of two completely unrelated spike trains, the cross-correlogram is expected to be flat, with value in hertz equal to the firing rate of *cell 2*. Even if the cells show an interaction (e.g., if *cell 1* drives *cell 2*), we expect the asymptotic value of the cross-correlogram at long times to be equal to the firing rate of *cell 2*. For an autocorrelogram of a single isolated cell, we expect that the asymptotic value will be equal to the firing rate of the cell. For an autocorrelogram of a poorly isolated cell, the central bins in the autocorrelogram (i.e., the refractory period) reflect the frequency of contaminating spikes.

The heights of the autocorrelogram "shoulders," if they exist, are set by a different time scale. The shoulder height reflects the probability per unit time that a spike will occur given that another spike occurred one intra-burst interspike interval ago. The shoulder height is therefore of the order of the probability that a given spike participates in a burst, divided by the variability of the intra-burst interspike intervals. Because this variability is <1 ms, the height of the autocorrelogram shoulders will be orders of magnitude larger than the central portion of the autocorrelogram, even for a poorly isolated cell.

The central portion of the autocorrelogram should therefore not be compared with the shoulders but with the asymptotic height at long times. For illustration purposes, we considered simulated spike trains. Figure 6*C* shows an autocorrelogram created from spikes of a mixture of two simulated bursting cells with equal firing rates (2.5 Hz) and 3-ms refractory periods. The refractory period (central part) of the correlogram looks reasonable when compared with the height of the shoulders. However, when the number of spikes in the refractory period is compared with the asymptotic portion of the autocorrelogram (Fig. 6*D*), it becomes clear that the spike train did
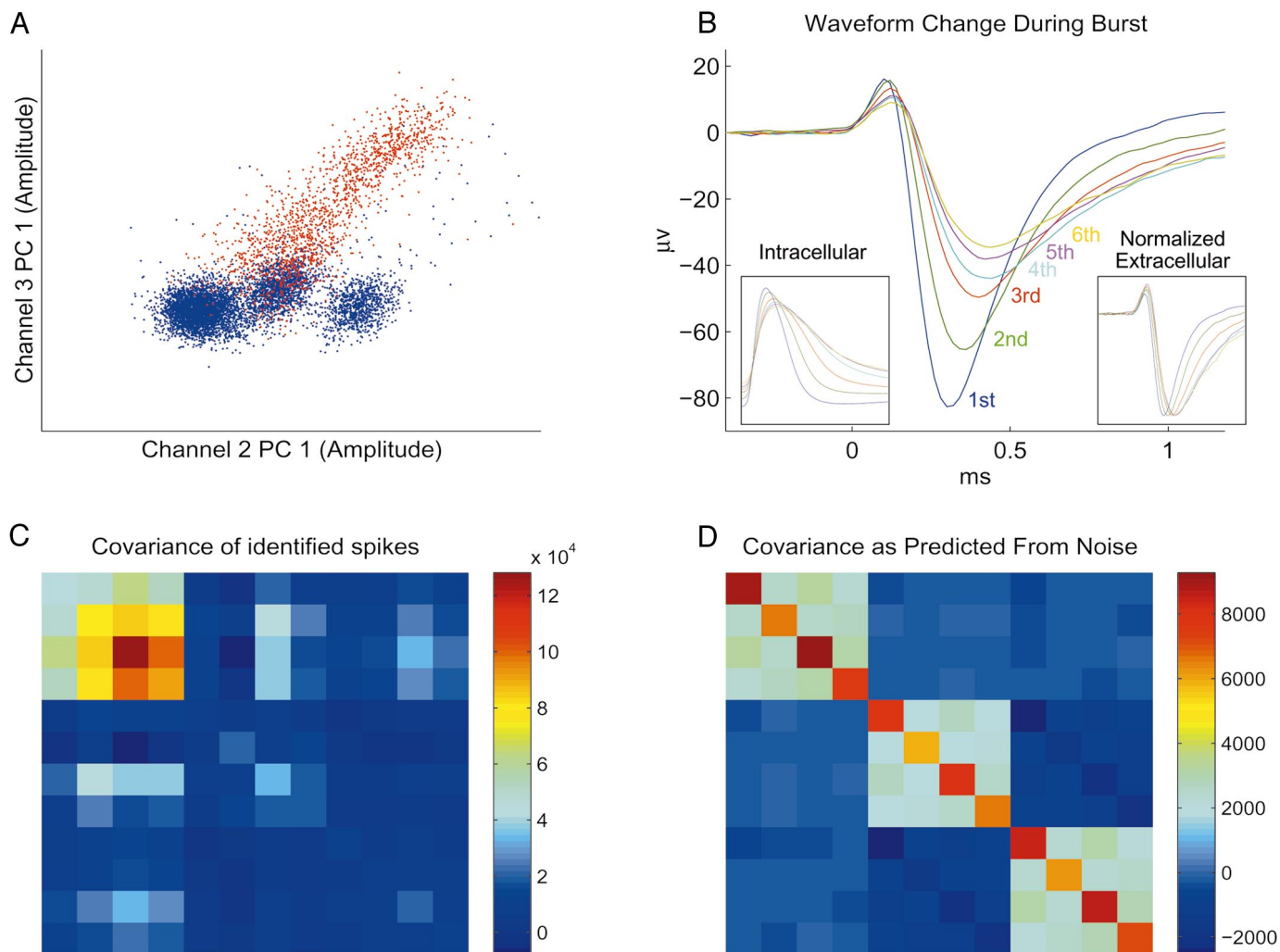
FIG. 5. Analysis of cluster shape for a bursting cell. *A*: cluster diagram for a data set containing an intracellularly identified bursting cell (red; *session 4* in Table 1). *B*: superimposed mean waveforms for the 1st to 6th spikes of a burst, showing a clear amplitude decrement. *Left inset*: corresponding mean intracellular waveforms. *Right inset*: extracellular waveforms normalized by amplitude, showing wave shape change during burst. *C*: covariance matrix for the bursting cell, showing a high correlation in amplitude across the 4 channels. *D*: covariance matrix as predicted by the power spectrum of the noise. For this bursting cell, the cluster shape cannot be predicted from the noise alone, indicating that spike shape variabilility during the burst causes significant change in cluster shape.

not derive from a single neuron. In the case of slow discharging cells or short recordings, there may not be enough spikes in the central portion of the autocorrelogram to allow accurate comparison to the asymptotic region. In these cases, the autocorrelogram cannot be used to determine the successful elimination of Type I errors.

The cross-correlogram can also be useful for evaluating the success of cell isolation (Fee et al. 1996). In particular, shoulders of the cross-correlogram for two clusters may indicate that both clusters contain spikes produced by the same bursting neuron (Fig. 6, *E* and *F*). Asymmetric peaks, in particular, indicate that the decreasing amplitude spikes of a burst have been put into separate clusters. Asymmetric cross-correlogram peaks may also be caused by direct coupling of two cells, but those observed in rat hippocampus may be distinguished from misclassified bursting cells on the basis of a characteristic narrow single cross-correlogram peak (Csicsvari et al. 1998). Conversely, the absence of shoulders in a cross-correlogram

may not be taken as a guarantee that the clusters do not both contain spikes from a common nonbursting cell.

Theoretically optimal performance

Errors are caused by overlapping clouds in cluster space. Is it possible to improve error rates by determining more appropriate boundaries between clusters? We addressed this question by using an estimate of maximum possible performance, termed the BEER (see METHODS). This measure gives an estimate of the minimum error rates for an optimally positioned ellipsoidal cluster, given the positions of the correct and incorrect spikes. Furthermore the BEER measure allows for a differential weighing of Type I and Type II errors, parameterized by a "conservatism" parameter. Therefore the BEER measure specifies a curve showing the tradeoff between Type I and Type II errors as this parameter varies from 0 to 1.

Figure 1 shows the estimated minimum possible errors, superimposed on the performance of the human operators.
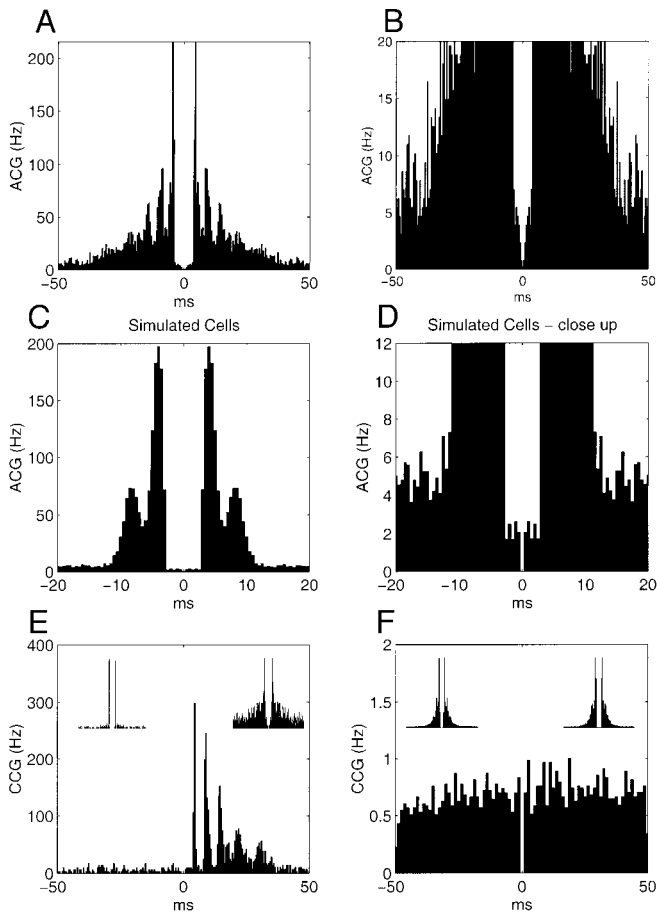
FIG. 6. Auto- and cross-correlograms. *A*: autocorrelogram for a poorly separated cell (*operator G, data set 4;* Type I error 43.4%). *B*: close-up of *A*, showing some contamination of the central refractory portion. *C*: autocorrelogram for 1-h-long simulated spike train. Two cells were simulated with an average firing rate of 2.5 Hz and a bursting pattern. Although the combined spike train consists of 2 independent cells, it appears to have a clear refractory period, when scaled by the burst shoulder height. *D*: close-up of *C*, showing that the autocorrelogram height during the refractory period is approximately half the asymptotic height. *E*: cross-correlogram showing large asymmetric peaks, indicating that 2 clusters both contain spikes from a single bursting cell (*operator C, data set 4*). *F*: cross-correlogram for of 2 well-isolated clusters with no systematic interaction between the 2 neurons. *E* and *F, insets*: autocorrelograms of the individual clusters.

Curves were generated using the peak-to-peak spike height only, the first principal component only, the first and second principal components, or the first three principal components. For each calculation, two BEER curves were generated by a cross-validation method, in which the ellipsoid position was determined from half of the data (even or odd numbered spikes) and evaluated on the other half. Examination of the two sets of curves provides an estimate of the minimum achievable error, whereas the difference between them is an indication of the accuracy of this estimate.

As was the case with the human operators, the BEER curves indicate a trade-off between false-positive and -negative errors. For all data sets, there was a strong distinction between the errors generated by the use of the peak-to-peak measures or the first principal components only versus the use of two- and three-principal components. This was particularly striking for the data set in which the intracellularly identified unit had a confounding unit with a similar amplitude distribution (Fig. 1,

cell 2, and Fig. 2). There was also a slight further improvement by using three over two principal components but this difference was much smaller than the difference between the use of one versus two principal components. In general, the performance of human operators was similar to the BEER performance when only the peak-to-peak amplitudes or one-principal component was used. It must be noted here that the human operators had all the principal components to work with and could get additional information from examination of spike waveforms and auto/cross-correlograms. Importantly, all human operators did worse than the theoretical best achieved using all three principal components. The difference between human operators and BEER measures was least when the extracellular amplitude of the impaled cell was relatively large and did not burst.

Analysis of human operator errors

To assess how performance could be improved, we analyzed the distribution of operator errors on the most difficult neuron, (*cell 1* of Fig. 1). Figure 7 displays the correctly identified spikes along with the false positive and negative errors made by an operator in a single two-dimensional projection. When the first principal component projections were displayed in the cluster space, errors occurred in a shell surrounding the cluster of correctly identified spikes. Furthermore more errors occurred at the border with small-amplitude spikes (Fig. 7A). Examining Fig. 7A, one could form the impression that the clustering performance cannot be substantially improved: decreasing Type I errors can be achieved only by increasing Type II errors. This impression was supported by calculating the Mahalanobis distance in the two-dimensional space for the correctly identified spikes, the false negative errors and false positive errors (Fig. 7B). The correctly identified spikes occurred closest to the center of the cluster, the false positive errors closer to the periphery, and the false negative errors were the furthest out of all (Fig. 7B). However, a different picture arose when the third-principal components were displayed. These projections are rarely used by human operators because the spikes of different cells do not form well-separated clouds. However, displaying the third-principal components revealed that the false positive errors occurred actually further out than the false negative errors (Fig. 7C). Similarly when the Mahalanobis distances were calculated in the full 12-dimensional space, the false positive errors occurred farther from the center of the cluster than the false negative errors (Fig. 7D). This conclusion was borne out in 46 of the 48 clustering sessions across all operators (Fig. 7E).

The preceding analysis indicated that the clusters created by the operators were not optimal. A plausible explanation is that human operators are able to use only two dimensions at a time. In any two-dimensional plane, the expected amount of cluster overlap is always more than in the full high-dimensional space. In contrast to the human operator, the BEER estimation can use all dimensions at once. This may explain the significantly lower error rates of the BEER method.

Automatic cluster cutting

The previous analysis has indicated that the performance of human operators was below the theoretical optimum, as esti-
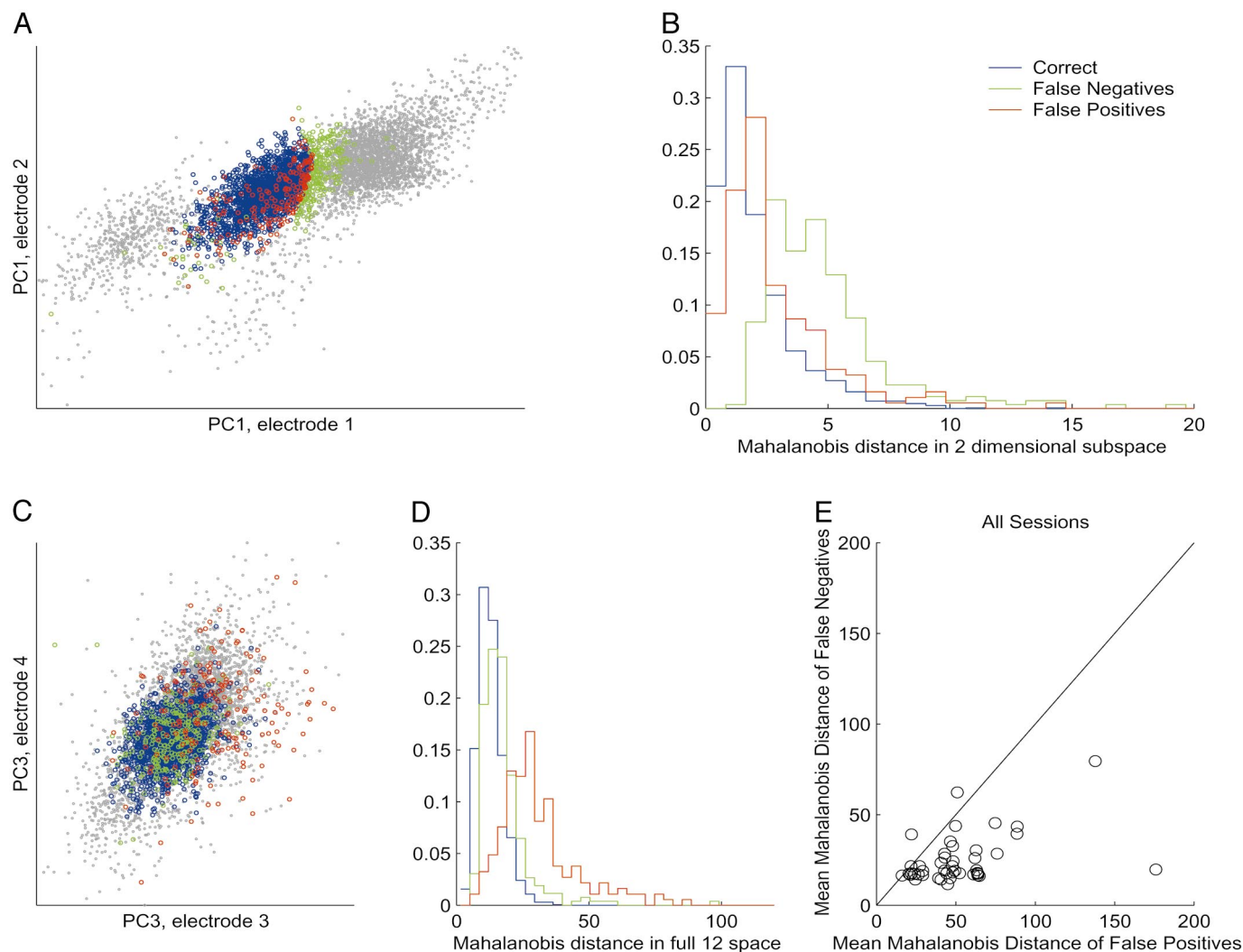
FIG. 7. Analysis of operator errors. *A*: errors made by *operator K* on *cell 1*, shown in one of the projections used by that operator when cluster-cutting. False positive errors (red) are located within the cluster area and false negative errors (green) outside. *B*: histograms of Mahalanobis distance from the cluster center in this 2-dimensional space. As expected, false positive errors are further toward the center than false negative errors. *C*: errors plotted in a projection not used by the operator. Unexpectedly, false positive errors are now further out than false negative errors. *D*: histograms of Mahalanobis distance for the human operator in the full 12-dimensional feature space. False positive errors are again further out than false negative errors. *E*: analysis of all clustering sessions. In every case but 2, the mean Mahalanobis distance in the 12-dimensional space is larger for false positive than for false negative errors. This indicates that the performance of operators is suboptimal because of their inability to visualize the full 12-dimensional space.

mated by the BEER method, and that this may be due to an inability of human operators to visualize the multidimensional cluster space. It is therefore of interest whether automatic clustering systems reduce the error rates made by the human operators.

Several approaches to automatic spike sorting have been proposed (Fee et al. 1996; Lewicki 1998; Sahani 1999; Wright et al. 1998). In our investigation, we examined the applicability of a general-purpose clustering method designed to automatically classify various types of data (Cheeseman and Stutz 1996).

Figure 8*A* shows the raw output produced by the program on *session 4*, in which the intracellular cell was bursting. The program generated nine clusters. However, examination of autocorrelograms indicated that three of these clusters (shown in red, green, and blue) corresponded to spikes from

the same cell (Fig. 8*B*). The clean refractory period in the cross-correlogram for all three clusters, and the asymmetric cross-correlograms between the red cluster and the green and blue clusters, suggested that the latter may correspond to correspond to subsequent spikes of the bursting cell. This is supported by the larger amplitude of the earlier discharging (red) spikes. Examination of other auto- and cross-correlograms (not shown) indicated that no other clusters corresponded to the bursting cell, and that the small-amplitude yellow, mauve, and pink clusters, and the diffuse gray cluster do not correspond to single units. The final clusters are shown in Fig. 8*C*. We refer to the process of automatic cluster determination, followed by manual reassignment as *semi-automatic* clustering.

The tendency of AutoClass to overcluster the data was born out in the other five recording sessions. The output of
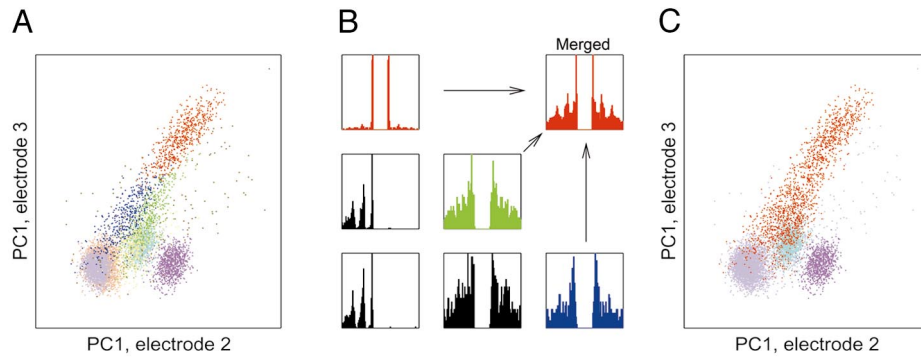
FIG. 8. Semi-automatic cluster cutting. *A*: the output of the AutoClass program (Cheeseman and Stutz 1996). The 9 clusters produced are shown in different colors. *B*: auto- and cross-correlograms for the spikes in the red, green, and blue clusters. The clean refractory periods in all cross-correlograms indicate that the 3 clusters correspond to a single cell. The asymmetric cross-correlograms between the red cluster and the other 2 suggest that it the red cluster contains the earlier spikes in the burst. Similar analysis of the yellow, mauve, pink, and gray clusters indicated that they did not correspond to single units. *C*: revised clusters after manual merging by a human operator.

the program was examined by an operator in each case, and clusters deemed to have belonged to the same cells were merged. In *session 6,* the program also produced a bimodal cluster, which was deemed to contain the spikes of two cells. The cluster was split manually, and the distinctness of the two subclusters was confirmed by a flat cross-correlogram.

The error rates of the semi-automatic system (AutoClass followed by manual cluster merging) are shown in the bottom line of Table 1, and are marked by A in Fig. 1. In all six files, the performance of the semi-automatic system was better than that of any human operator. Furthermore in every case, the performance of the semi-automatic system was comparable with the theoretical optimum as estimated by the BEER measure. We therefore confirm that automatic cluster cutting, if followed by proper postclustering examination and adjustment, can indeed lead to lower error rates than manual spike sorting methods.

Comparison of tetrode to single-wire performance

While tetrodes are often acknowledged to provide more accurate spike separation in hippocampus (Drake et al. 1988; McNaughton et al. 1983; Nadasdy et al. 1998; Recce and O'Keefe 1989; Wilson and McNaughton 1993), and necortex (Gray et al. 1995; Wright et al. 1998), single-wire recordings are still often used in many studies (Deadwyler and Hampson 1995; Hampson et al. 1999; Nicolelis et al. 1997). We characterized the improvement gained from tetrode recordings by comparing optimal possible performance using only one of the tetrode channels relative to optimal performance using all four.

Figure 9 shows the optimal performance estimate for the six recording sessions using each channel individually compared with optimal performance using them all together. In each plot, the four pairs of dotted lines show estimated optimal performance for each of the four sites, and the single pair of solid lines shows estimated optimal performance for the full tetrode.
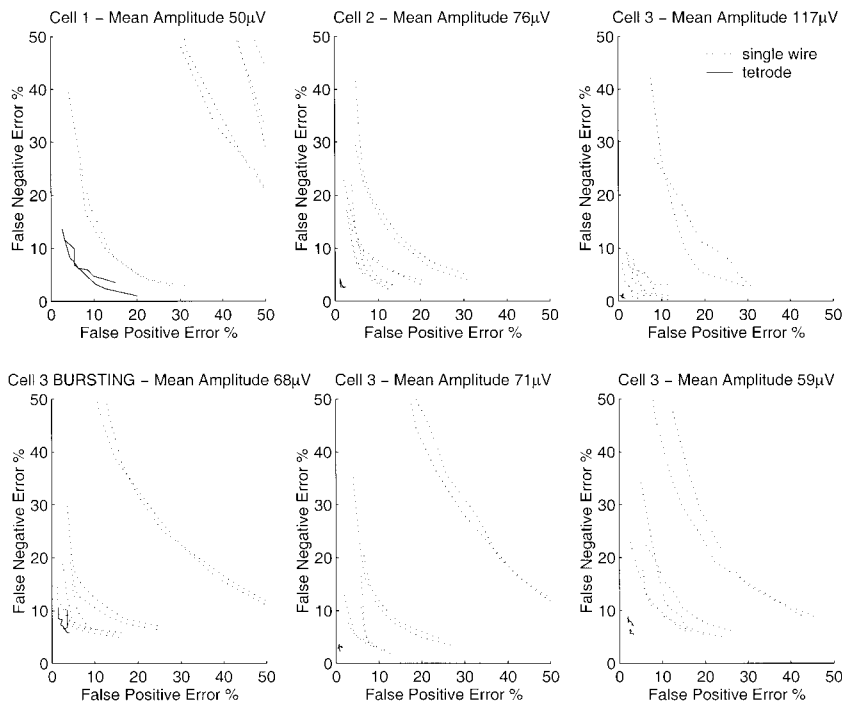

FIG. 9. Comparison of estimated optimal performance with singles wires relative to optimal performance for the full tetrode. ···, best ellipsoid error rate (BEER) estimates for each of the 4 channels used alone. —, the BEER estimate for the full tetrode. In the *bottom-right* plot, only 3 pairs are visible because the errors for the 4th site were off-scale (Type I and Type II errors both >50%). Note significantly poorer performance of even the best single wires relative to the full tetrode.

All three principal components were used (Abeles and Goldstein 1977).

The error estimate for single wires differed greatly from wire to wire with the worst wires showing Type I and Type II errors both exceeding 50% and therefore off the scale of the plots in Fig. 9. However, even the best wires show poorer performance than tetrodes with error rates typically double those for the full tetrode.

DISCUSSION

We used simultaneous intracellular and extracellular recordings to quantify errors of manual spike separation methods. To date, no other methods have been available for an independent and objective evaluation of any on- or off-line unit classification methods. Because of the subjective bias and differing degree of experience of the individual human operators, not only the overall error rates but the types of errors inherent in manual methods are expected to vary across different laboratories. For several types of investigations, multiple-unit recording or contaminated single-unit recording may be adequate. However, the analysis of emerging, cooperative activity of single neurons (Buzsaki et al. 1992; Deadwyler and Hampson 1995; McNaughton et al. 1996; Nadasdy et al. 1999; Riehle et al. 1997; Skaggs and McNaughton 1996; Wilson and McNaughton 1993, 1994) requires very accurate spike sorting. Our major finding is that the causes of separation error can be explained by overlapping clouds in cluster space due to the spread of clusters representing single cells and to the similarity of clusters corresponding to different cells. Specifically, we found that *1*) the Type I and Type II errors of human operators typically ranged from 0 to 30% but could exceed 50% during periods of synchronous population cell discharges, *2*) neurons with similar amplitude profiles across channels may be discriminated when information about the waveforms is also utilized, *3*) auto- and cross-correlograms of spikes may identify poor clusters but cannot always estimate confounding spikes in the clusters, *4*) the most important source of error is human error, caused by the observer's inability to visualize the full high-dimensional cluster space. *5*) Using information from single wires produced substantially more errors than tetrode recordings. *6*) Spike sorting errors may be significantly reduced by semi-automatic clustering.

Synchronous discharge of nearby neurons increases Type I clustering error

The most important error of biological origin occurs due to the near simultaneous discharge of neurons in the vicinity of the recording electrodes. This was demonstrated best by the observation that both Type I and Type II errors increased several-fold when spikes occurred during sharp wave-associated population bursts of CA1 neurons. We examined two potential sources of the increased error. Hippocampal sharp waves are associated with fast field oscillation (ripples, 100–200 Hz) in the CA1 pyramidal layer and an increased synchrony of pyramidal cells and interneurons (Buzsaki et al. 1992). These high-frequency ripple waves are often significantly larger in amplitude than the units to be discriminated, and therefore inadequate filtering may not perfectly separate field events from the extracellularly recorded spike. However,

when filtered traces of spikes collected during the presence and absence of ripples were compared, they were identical in amplitude and shape, indicating that digital filtering successfully eliminated the effects of nonspike related background fluctuations. However, in contrast to the similar mean spike waveforms, the variability of spikes was substantially larger during ripples than in their absence. A likely reason for this increased variability is the random superimposition of spikes tightly coupled to the negative peaks of ripple waves (Csicsvari et al. 1999). If a cell fires during the spike of another neuron, the resulting waveshape will be the combination of the two spikes. If the added spike(s) sufficiently alters the amplitude distribution and shape of the spike of the neuron to be clustered, an omission (false negative) error will occur. Conversely, superimposition of smaller amplitude spikes may result in spike waveforms sufficiently similar to be incorporated into the cluster of the larger amplitude neuron, thus incrementing false positive errors. We hypothesize that a major contributing factor to Type I error is the increased probability of discharge of many nearby neurons in a narrow time window. We have estimated that a single tetrode can monitor >100 clusterable neurons in the CA1 pyramidal layer (Henze et al. 2000). In contrast to this large number, typically <10 of these recordable cells are routinely clustered in the awake behaving animal. The most likely reason for the relatively low number of recovered clusters is that the great majority of pyramidal neurons are silent most of the time. However, during the population synchrony associated with sharp waves (Buzsaki et al. 1992; Csicsvari et al. 1999) even some of the previously silent cells may discharge. Since these silent cells do not have sufficient spike counts to form recognizable cluster clouds, the spikes they emit may be mistakenly incorporated into the clusters of other neurons.

Although population bursts of neurons provided the most robust examples for the demonstration of spike interference effects, the same rules also apply to any state when nearby neurons discharge together within the duration of the action potential. In the hippocampus, such spike-spike interference effects are relatively small because in the absence of sharp wave bursts neighboring pyramidal cells (place cells) only exceptionally represent the same part of the environment and therefore rarely fire together (Recce and O'Keefe 1989). In contrast, neurons in cortical columns are typically coactivated and discharge at a high rate in response to relevant inputs (Hubel and Wiesel 1962; Mountcastle 1957). In addition, neocortical networks can also become synchronized with millisecond precision either spontaneously (Abeles and Gerstein 1988; Buzsaki and Kandel 1998) or in response to sensory inputs (Jones and Barth 1999; Kaneko et al. 1999). Our observations therefore indicate that separation of tetrode-recorded neocortical neurons with the currently available clustering methods are likely to yield much larger Type I error rates than in the hippocampal CA1 region.

Spike waveform variability of single neurons contribute to Type II error

The original motivation of the tetrode technique was that multiple voltage sensors can reliably locate a given neuron in space on the basis of spike-amplitude ratios (Drake et al. 1988; Gray et al. 1995; McNaughton et al. 1983; Recce and O'Keefe

1989). However, research on the biophysical properties of neurons indicates that neurons are not point sources of extracellular currents underlying the action potential. Both the perisomatic region and the dendritic compartments actively support both the generation and propagation of action potentials (Calvin and Hartline 1977; Kamondi et al. 1998; Llinas and Nicholson 1971; Regehr et al. 1993; Stuart et al. 1997). Because spike activity of pyramidal cells can be recorded several hundred micrometers from their somata (Buzsaki and Kandel 1998; Buzsaki et al. 1996; Henze et al. 2000), the extracellular spike recorded by the extracellular recording electrode represents a summed activity of somatic and dendritic currents (Nadasdy et al. 1998). Since the amplitude and waveform of dendritic action potentials vary substantially as a function of the presynaptic network activity (Kamondi et al. 1998), the extracellular waveforms also vary along the somadendritic axis, modulated by ongoing network behavior. Complex-spike bursts of pyramidal cells (Ranck 1973) represent perhaps the most extreme degree of extracellular spike amplitude and waveform variation. Our findings revealed that these factors can often lead to elongated clusters, which also increases the probability of cluster overlap with spikes emitted by other neurons. Because units with spikes of various shapes and sizes may be inadvertently grouped into different clusters, the number of spikes emitted by a single neuron may be substantially underestimated (omission error).

Improved error rates achieved by semi-automatic clustering

Examination of the errors made by human operators indicated that their performance was below the theoretical optimum and that the most likely cause of this sub-optimal performance was an inability to visualize the high-dimensional cluster space. This was confirmed by experiments with a semi-automatic clustering process, which achieved error rates closer to the theoretical optimum.

The semi-automatic process consisted of an automatic classification program, followed by examination and reassignment by a human operator. While the program's parameters needed very little adjustment, it had a tendency to overcluster the data; this tendency was expressed by dividing the cloud corresponding to a single cell into several clusters, which was most noticeable for bursting cells. This problem was rectified by a human operator who manually merged clusters corresponding to the same cell based on cluster proximity, amplitude ratios, and cross-correlograms. In addition, the program would occasionally produce a single cluster combining spikes from two cells. This again may be corrected by manual intervention, including examination for bimodality of clusters and subsequent examination of the cross-correlogram of the divided cluster.

The usual arguments advanced so far in favor of automatic spike sorting have been that it is considerably faster than the manual method and free from the subjective bias and experience level of the operator. We now can add another argument: that it may lower error rates, beyond even those of the most experienced operator.

Superiority of tetrodes over single wires

Spike isolation methods based on multiple-site recordings have been used extensively (McNaughton et al. 1983; Recce and O'Keefe 1989). Nevertheless several investigators use single wires for unit isolation mostly because of the convenience of the fabrication and implantation of single wires. Estimation of optimal spike separation errors for single sites of our tetrode data showed great variability depending on which site was being used. However, even the best channels still showed relatively poor performance with typical error rates approximately twice those of the full tetrode. We therefore confirm that spike separation is significantly more effective when multiple sites are used for unit recording (Drake et al. 1988; McNaughton et al. 1983; Nadasdy et al. 1998; Recce and O'Keefe 1989).

How may spike separation be further improved?

Use of the full tetrode channels and a semi-automatic clustering process improved the performance from the level achieved by human operators to the theoretical estimate of optimum performance provided by the BEER measure. Can spike separation performance be further improved?

Three directions for future spike separation methods are worth emphasizing. First, improvement of the recording hardware may enhance signal-to-noise ratios and separability of units. Silicon technology-based electrodes are especially promising because they can be standardized and optimized for various neuron types without increasing tissue damage (Nadasdy et al. 1998).

Second, although semi-automatic clustering can achieve close to the theoretical optimum performance for a given set of feature vectors, it may be possible to lower error rates further by improving the feature vectors themselves. Two stages are involved in going from raw traces to feature vectors: spike detection and feature extraction. We have seen here that use of waveshape information (in this case by principal component analysis) leads to improved performance over simple peak-to-peak amplitude. However, further improvements in spike detection and feature extraction may lead to more focused clusters and thus lower error rates.

Third, the semi-automatic clustering process described here still requires inspection by a human operator. This takes time, albeit less time than manual clustering, and raises the possibility of subjective bias. Ideally a fully automatic clustering system would produce an output corresponding to single cells without the need for user intervention. The AutoClass software used in this paper was designed for automatic classification in a wide range of domains and made no use of neuroscience knowledge, such as refractory periods and the nature of complex-spike bursts. The incorporation of such knowledge (Fee et al. 1996; Sahani 1999; K. Zhang, personal communication) holds out hope for more reliable, fully automatic spike sorting.

REFERENCES

ABELES M AND GERSTEIN GL. Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J Neurophysiol* 60: 909–924, 1988.

ABELES M AND GOLDSTEIN MH. Multispike train analysis. *Proc IEEE* 65: 762–773, 1977.

ANDERSON TW. *An Introduction to Multivariate Statistical Analysis* (2nd ed.). New York: John Wiley, 1984.

BUZSAKI G, HORVATH Z, URIOSTE R, HETKE J, AND WISE K. High-frequency network oscillation in the hippocampus. *Science* 256: 1025–1027, 1992.

BUZSAKI G AND KANDEL A. Somadendritic backpropagation of action potentials in cortical pyramidal cells of the awake rat. *J Neurophysiol* 79: 1587–1591, 1998.

BUZSAKI G, PENTTONEN M, NADASDY Z, AND BRAGIN A. Pattern and inhibition-dependent invasion of pyramidal cell dendrites by fast spikes in the hippocampus in vivo. *Proc Natl Acad Sci USA* 93: 9921–9925, 1996.

CALVIN WH AND HARTLINE DK. Retrograde invasion of lobster stretch receptor somata in control of firing rate and extra spike patterning. *J Neurophysiol* 40: 106–118, 1977.

CHEESEMAN P AND STUTZ J. Bayesian classification (AutoClass): theory and results. In: *Advances in Knowledge Discovery and Data mining,* edited by Fayyad U, Piatesky-Shapiro G, Smyth P, and Uthurusamy R. Menlo Park, CA: AAAI Press, 1996, p. 153–180. http://ic-www.arc.nasa.goc/ic/projects/bayes-group/autoclass.

CSICSVARI J, HIRASE H, CZURKO A, AND BUZSAKI G. Reliability and state dependence of pyramidal cell-interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. *Neuron* 21: 179–189, 1998.

CSICSVARI J, HIRASE H, CZURKO A, MAMIYA A, AND BUZSAKI G. Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *J Neurosci* 19: 274–287, 1999.

DEADWYLER SA AND HAMPSON RE. Ensemble activity and behavior: what's the code? *Science* 270: 1316–1318, 1995.

DRAKE KL, WISE KD, FARRAYE J, ANDERSON DJ, AND BEMENT SL. Performance of planar multisite microprobes in recording extracellular activity. *IEEE Trans Biomed Eng* 35: 719–732, 1988.

EVARTS EV. A technique for recording activity of subcortical neurons in moving animals. *Electroencephalogr Clin Neurophysiol* 24: 83–86, 1968.

FAHLMAN SE. Fast learning variations on backpropagation: an empirical study. In: *Proc. 1988 Connectionist Models Summer School,* edited by Touretzky D, Hinton G, and Sejnowski T. San Mateo, CA: M. Kaufmann, 1988.

FEE MS, MITRA PP, AND KLEINFELD D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *J Neurosci Methods* 69: 175–188, 1996.

GEORGOPOULOS AP, TAIRA M, AND LUKASHIN A. Cognitive neurophysiology of the motor cortex. *Science* 260: 47–52, 1993.

GNANADESIKAN R. *Methods for Statistical Data Analysis of Multivariate Observations.* New York: Wiley, 1977.

GRAY CM, MALDONADO PE, WILSON M, AND MCNAUGHTON B. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J Neurosci Methods* 63: 43–54, 1995.

HAMPSON RE, SIMERAL JD, AND DEADWYLER SA. Distribution of spatial and nonspatial information in dorsal hippocampus. *Nature* 402: 610–614, 1999.

HENZE DA, BORHEGYI Z, CSICSVARI J, MAMIYA A, HARRIS KD, AND BUZSAKI G. Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J Neurophysiol* 84: 390–400, 2000.

HUBEL DH AND WIESEL TM. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol (Lond)* 160: 106–154, 1962.

JOHNSON RA AND WICHERN DA. *Applied Multivariate Statistical Analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall, 1992.

JOHNSTON D AND WU SM. *Foundations of Cellular Neurophysiology.* Cambridge, MA: MIT Press, 1995.

JONES MS AND BARTH DS. Spatiotemporal organization of fast (>200 Hz) electrical oscillations in rat Vibrissa/Barrel cortex. *J Neurophysiol* 82: 1599–1609, 1999.

KAMONDI A, ACSADY L, AND BUZSAKI G. Dendritic spikes are enhanced by cooperative network activity in the intact hippocampus. *J Neurosci* 18: 3919–3928, 1998.

KANEKO H, SUZUKI SS, OKADA J, AND AKAMATSU M. Multineuronal spike classification based on multisite electrode recording, whole-waveform analysis, and hierarchical clustering. *IEEE Trans Biomed Eng* 46: 280–290, 1999.

KRUGER J AND AIPLE F. Multimicroelectrode investigation of monkey striate cortex: spike train correlations in the infragranular layers. *J Neurophysiol* 60: 798–828, 1988.

LEWICKI MS. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* 9: R53–R78, 1998.

LLINAS R AND NICHOLSON C. Electrophysiological properties of dendrites and somata in alligator Purkinje cells. *J Neurophysiol* 34: 532–551, 1971.

LLINAS R AND SASAKI K. The functional organization of the olivo-cerebellar system as examined by multiple purkinje cell recordings. *Eur J Neurosci* 1: 587–602, 1989.

MCNAUGHTON BL, BARNES CA, GERRARD JL, GOTHARD K, JUNG MW, KNIERIM JJ, KUDRIMOTI H, QIN Y, SKAGGS WE, SUSTER M, AND WEAVER KL. Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *J Exp Biol* 199: 173–185, 1996.

MCNAUGHTON BL, O'KEEFE J, AND BARNES CA. The stereotrode: a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *J Neurosci Methods* 8: 391–397, 1983.

MOORE GP, SEGUNDO JP, PERKEL DH, AND LEVITAN H. Statistical signs of synaptic interaction in neurons. *Biophys J* 10: 876–900, 1970.

MOUNTCASTLE VB. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J Neurophysiol* 20: 408–434, 1957.

NADASDY Z, CSICSVARI J, PENTTONEN M, HETKE J, WISE K, AND BUZSAKI G. Extracellular recording and analysis of neuronal activity: from single cells to ensembles. In: *Neuronal Ensembles: Strategies for Recording and Decoding,* edited by Eichenbaum HB and Davis JL. New York: Wiley-Liss, 1998, p. 17–55.

NADASDY Z, HIRASE H, CZURKO A, CSICSVARI J, AND BUZSAKI G. Replay and time compression of recurring spike sequences in the hippocampus. *J Neurosci* 19: 9497–9507, 1999.

NICOLELIS MA, FANSELOW EE, AND GHAZANFAR AA. Hebb's dream: the resurgence of cell assemblies. *Neuron* 19: 219–221, 1997.

RANCK JBJ. Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. I. Behavioral correlates and firing repertoires. *Exp Neurol* 41: 461–531, 1973.

REBRICK S, TZONEV S, AND MILLER K. Analysis of tetrode recordings in cat visual system. [Online] *http://mccoy.ucsf.edu/Papers/CNS97/CNS97-paper.htm* [1997].

RECCE M AND O'KEEFE J. The tetrode: a new technique for multi-unit extracellular recording. *Soc Neurosci Abstr* 15: 1250, 1989.

REGEHR W, KEHOE JS, ASCHER P, AND ARMSTRONG C. Synaptically triggered action potentials in dendrites. *Neuron* 11: 145–151, 1993.

RIEHLE A, GRUN S, DIESMANN M, AND AERTSEN A. Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278: 1950–1953, 1997.

SAHANI M. *Latent Variable Models for Neural Data Analysis* (PhD thesis). Pasadena, CA: Caltech, 1999.

SKAGGS WE AND MCNAUGHTON BL. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* 271: 1870–1873, 1996.

STUART G, SPRUSTON N, SAKMANN B, AND HÄUSSER M. Action potential initiation and backpropagation in neurons of the mammalian CNS. *Trends Neurosci* 20: 125–131, 1997.

WEHR M, PEZARIS JS, AND SAHANI M. Simultaneous paired intracellular and tetrode recordings for evaluating the performance of spike sorting algorithms. *CNS*. 98.* Santa Barbara, CA, July 26–30, 1998.

WILSON MA AND MCNAUGHTON BL. Dynamics of the hippocampal ensemble code for space. *Science* 261: 1055–1058, 1993.

WILSON MA AND MCNAUGHTON BL. Reactivation of hippocampal ensemble memories during sleep. *Science* 265: 676–679, 1994.

WISE KD AND NAJAFI K. Microfabrication techniques for integrated sensors and microsystems. *Science* 254: 1335–1342, 1991.

WOOD ER, DUDCHENKO PA, AND EICHENBAUM H. The global record of memory in hippocampal neuronal activity. *Nature* 397: 613–616, 1999.

WRIGHT BD, REBRICK S, AND MILLER KD. Spike-sorting of tetrode recordings in cat LGN and visual cortex: cross-channel correlations, thresholds, and automatic methods. [Online] http://mccoy.ucsf.edu/Papers/Sfn98/SFN98 poster.html [1998].

YLINEN A, BRAGIN A, NADASDY Z, JANDO G, SZABO I, SIK A, AND BUZSAKI G. Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms. *J Neurosci* 15: 30–46, 1995.