

## Toolbox

**Editor's Note:** Toolboxes are intended to describe and evaluate methods that are becoming widely relevant to the neuroscience community or to provide a critical analysis of established techniques. For more information, see [http://www.jneurosci.org/misc/ifa\\_minireviews.dtl](http://www.jneurosci.org/misc/ifa_minireviews.dtl).

# Quality Metrics to Accompany Spike Sorting of Extracellular Signals

Daniel N. Hill,<sup>1</sup> Samar B. Mehta,<sup>3</sup> and David Kleinfeld<sup>2</sup>

<sup>1</sup>Institute of Neuroscience, Technical University Munich, 80802 Munich, Germany, <sup>2</sup>Departments of Physics and Neurobiology, University of California, San Diego, La Jolla, California 92093-0374, and <sup>3</sup>College of Medicine, SUNY Downstate, Brooklyn, New York 11203

## Introduction

Studies of neuronal activity throughout the brain predominantly rely on extracellular recordings of spiking activity. These records often contain action potentials from more than one neuron, so it is imperative to discriminate the spikes that originate from separate neurons. The process of discrimination is based on differences in each neuron's extracellular waveform that originate from cell-specific biophysical properties. We briefly review the electronics of extracellular recording, the detection of spike events, and the clustering of similar spike waveforms as a means to sort spikes according to their putative underlying sources. When spike sorting succeeds, it transforms a fundamental weakness of extracellular recording, namely the inability to isolate single neurons, into one of its greatest strengths,

the simultaneous measurement from multiple cells.

Special emphasis is placed on visualization schemes for spike data as well as on a set of metrics to estimate the number of incorrectly categorized spikes. False-negative contributions to a cluster lead to a suppression of inferred spike rates, while false-positive contributions lead to a distortion in the inferred receptive field for the cell. Both errors reduce the estimated information carried by the cell. A matrix of values for these metrics allows readers to assess claims, e.g., the size and reliability of multiple peaks in a receptive field, relative to the level of contamination of the data.

## Analog processing

Extracellular action potential waveforms appear roughly biphasic in the vicinity of a cell soma and dendrite, as the capacitive current flow across the cell membrane dominates and the voltage approximates the derivative of the intracellular waveform. The amplitude of the extracellular signals is small compared with that of the intracellular potential, on the order of hundreds of microvolts, but still significant in terms of signal-to-noise ratio (Henze et al., 2000) (Fig. 1A). The resistive nature of the extracellular space ensures that currents summate without preferential attenuation of their high-frequency components (Logothetis et al., 2007), so that the signals from different cells may be recorded with a single or multiwire electrode at a fixed location (Buzsáki, 2004).

The most important factors for successful spike sorting are the choice of appropriate microelectrodes, the precision of the placement of the electrodes, and quality of the acquisition electronics. When spiking activity is synchronous between neighboring cells, there is considerable overlap of the spike waveforms, and sorting becomes very difficult. In this case, it appears best to use a fine individual electrode to isolate a single cell. These electrodes typically have 1- to 5- $\mu\text{m}$ -diameter tips and an impedance whose magnitude is typically 1–20 M $\Omega$  measured at a frequency of 1 kHz. Multiwire electrodes, such as stereotrodes (McNaughton et al., 1983) and tetrodes (Gray et al., 1995), are appropriate for simultaneous recording from two or more units when neuronal activity is largely asynchronous (Ainsworth and O'Keefe, 1977; Krüger and Bach, 1981; Reitboeck, 1983; Venkatachalam et al., 1999; Fee and Leonardo, 2001; Yamamoto and Wilson, 2008; Battaglia et al., 2009). These electrodes are made of 12–25  $\mu\text{m}$  microwire with exposed ends that have impedances of  $\sim$ 100 k $\Omega$ . While the extracellular waveform of two neurons may appear similar on a single microwire, they are unlikely to be similar on multiple electrodes at different albeit nearby locations. Last, the impedance of electrodes exhibits equal reactive and resistive components (Humphrey and Schmidt, 1991); the magnitude of these components decreases as a power law with increasing frequency (Blum et al., 1991) and the measured noise decreases with in-

Received Feb. 23, 2011; revised April 11, 2011; accepted April 19, 2011.

This work was supported by the National Institutes of Health (Grant NS051177 to D.K., Grant FNS054393A to D.N.H.) and the US-Israeli Binational Science Foundation (Grant 2007121). We thank the two reviewers for insightful comments, Ehud Ahissar, Elad Assa, John Curtis, Michale Fee, Kenneth Harris, Stephen Lisberger, and Jeffrey Moore for useful discussions, Gyorgy Buzsáki and Darrel Henze for supplying data for Figure 1A, and E.J. Chichilnisky, Vicky Lam, Jill Leutgeb, and Emily Mankin for supplying test data for analysis during the development of our sorting criteria. The material in this article has been presented as part of the NIH-sponsored school on "Neuroinformatics" held at the Marine Biological Laboratories. We apologize to all engineers and scientists whose work could not be cited as a consequence of space constraints.

Correspondence should be addressed to David Kleinfeld, University of California 0374, 9500 Gilman Drive, La Jolla, CA 92093-0374. E-mail: dk@physics.ucsd.edu.

DOI:10.1523/JNEUROSCI.0971-11.2011

Copyright © 2011 the authors 0270-6474/11/318699-07\$15.00/0

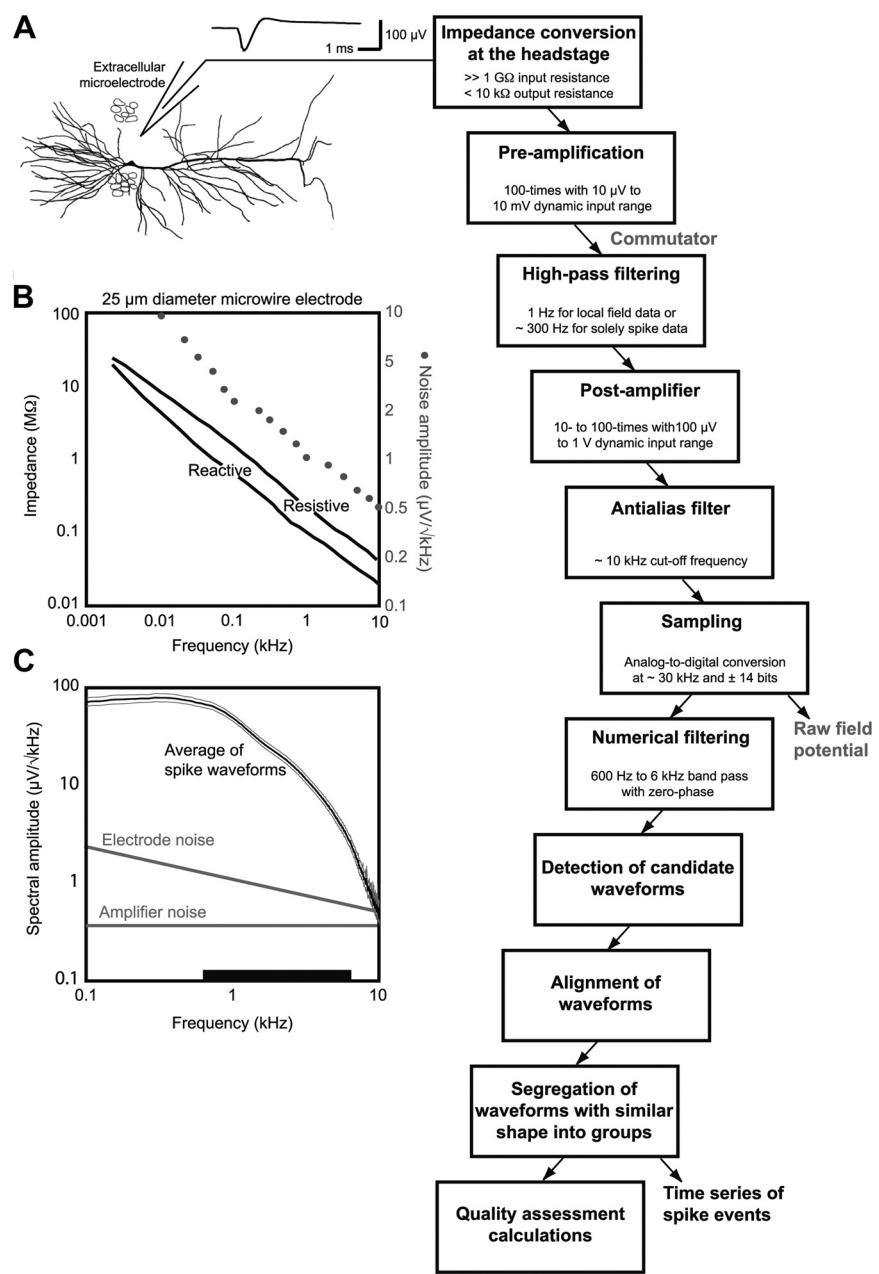
creasing frequency, in a manner consistent with thermal Johnson noise (Fig. 1*B*).

The electrical signal recorded by the electrode must be conditioned for acquisition (Fig. 1*A*). In general, the signal is impedance buffered as close to the electrode tip as possible to minimize contamination by noise from environmental sources. Spectral filtering is performed to isolate frequency components relevant for spike waveforms. The signal from the amplifier is high-pass filtered near 300 Hz solely for spike recording or at frequencies down to 0.1 Hz when field potential data are simultaneously acquired. The signal is typically low-pass filtered at 10 kHz, which is where the amplifier noise and the spectral power of the spike are comparable (Fig. 1*C*). The signal is oversampled, and the entire digitized data stream is stored for analysis. Acquisition schemes where only putative spike events are stored do not allow for the characterization of background noise or for different spike detection criteria to be applied *post hoc*. Before offline sorting, the digitized signals are additionally bandpass filtered to specifically isolate those frequencies present in the extracellular waveform that exceed the noise (Fig. 1*C*).

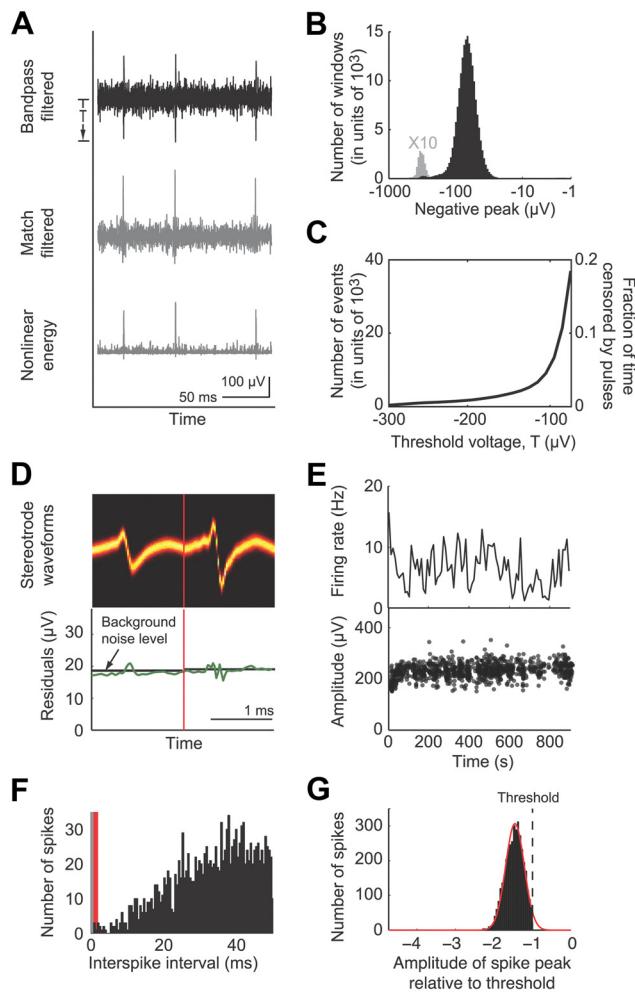
### Algorithmic spike sorting

We summarize the key issues of spike sorting and refer the reader to the work by Lewicki (1998) for a thorough review. The first step in automated spike sorting is the extraction of spike waveforms from extracellular data, and the second is the clustering of these waveforms into groups that represent the activity of single neurons. The simplest method for detecting a spike is to select a threshold value of voltage for the extracellular signal on each microwire (Fig. 2*B*). This approach considers the voltage only at a particular time point, whereas extracellular waveforms have characteristic shapes that extend over many data samples. If the extracellular waveform is known *a priori*, the matched filter provides an optimal way of linearly filtering the recording to use the full shape of the waveform in detection (Fig. 2*A*). As the waveform is not typically known *a priori*, a nonlinear filter can increase the signal-to-noise ratio by emphasizing voltage deflections that are both large in amplitude and high in frequency content (Kim and Kim, 2000) (Fig. 2*A*).

The choice of threshold for detection of a putative spike waveform is a trade-off between false-positive and false-negative



**Figure 1.** Signals and signal conditioning steps in the acquisition of extracellular data. **A**, The extracellular signal from one microwire of a tetrode, positioned near a hippocampal CA1 neuron *in vivo* (Henze et al., 2000), followed by a schematic of the signal flow. Each box refers to a specific process that is implemented by an electronic or computational unit; see Ganguly and Kleinfeld (2004), their supplemental material, for details of the circuit. All extracellular signals are referenced to an unrelated region of cortex and impedance-buffered at the head of the animal. We use N-channel field effect transistors (SST4118; Vishay Siliconix) with matched transconductances that were configured as common source transconductance amplifiers; the input noise of these transistors does not exceed the noise generated by the microelectrode. The transistors provide noise immunity as well as current to drive flexible cables that connect the head stage to precision load resistors, followed by 1 Hz high-pass single-pole filters. We measured the difference between each filtered channel and the filtered reference signals with a field effect transistor-input instrumentation amplifier with a fixed gain of 100 (*v/v*) (INA101, Burr Brown, Texas Instruments). The amplified signal then passes through a commutator, if applicable, and is high-pass filtered (4-pole Bessel filter configured around a no. UAF42, Burr Brown), further amplified (no. OPA101, Burr Brown), anti-alias filtered (8-pole constant-phase low-pass filter; no. D68L8D—10.0 kHz, Frequency Devices), and finally digitized (16 bit digitizer, NI-6251, National Instruments). All numerical operations were performed with programs written in MATLAB (The MathWorks). **B**, The complex impedance of a 25- $\mu\text{m}$ -diameter microwire electrode, cut at 45° and beveled on a diamond grinding wheel, as a function of frequency, together with the measured electrode noise. The measured curves have a phenomenological falloff as  $|\text{Impedance}| \propto f^{-0.78}$ , where  $f$  is the frequency. The measured noise is close to the expected value for thermal Johnson noise, which varies as  $\sqrt{4k_B T Z_{\text{Resistive}}}$  for a 1 Hz bandwidth, where  $k_B T \approx 100 \text{ meV}$  is the thermal energy and  $Z_{\text{Resistive}}$  is the resistive part of the impedance. **C**, The spectral composition of the spike signal averaged over 151 waveforms from the same single unit. The bands have 0.95 confidence limits. Part of the decrement near 10 kHz is caused by the low-pass filter. Also shown are the fit to the electrode noise and the specified amplifier noise.



**Figure 2.** Threshold selection for spike detection and the stability and self-consistency of a single cluster of spikes. **A**, Example from rat motor cortex containing three obvious extracellular spikes in a single-electrode record. Top trace, Data were low-pass filtered at 6 kHz and high-pass filtered at 600 Hz. Middle trace, The data were convolved with a matched filter expressed by  $\lambda \bar{\Sigma}^{-1}\mathbf{v}$ , where  $\mathbf{v}$  is a vector containing the mean waveform of a spike cluster found on a previous iteration of spike sorting,  $\bar{\Sigma}$  is the covariance matrix of the background noise sampled at random time intervals, and  $\lambda$  is a scalar that normalizes the filter gain to one (Duda and Hart, 1973). Bottom trace, The data were filtered by the nonlinear energy operator,  $E_{NL}(n) = x^2(n) - x(n+1)x(n-1)$ , where  $x(n)$  is the value of the signal at the  $n$ th sample (Kim and Kim, 2000). Note that the scale bar does not apply to the output of the  $E_{NL}$ . **B**, Histogram of values in entire extracellular signal from single electrode data as in **A**, after bandpass filtering. Note the presence of a second peak of large negative amplitude that corresponds to data points from spike events. There are 100 bins spaced out on a logarithmic scale between  $-1$  and  $-1000$  µV. **C**, Number of detected events as a function of detection threshold for larger portion of same dataset as in **A**. Left y-axis gives the total number of events, while the right y-axis gives the proportion of the experiment that is censored by this many events. The censored period was 0.75 ms in a dataset of 200 s. **D**, Two-dimensional histogram of waveforms (top) and standard deviation of waveform for each time point in waveform (bottom) for stereotrode data; black line is the standard deviation of the background activity of that channel. The waveform acquisition time is 2 ms of which a 1.5 ms interval is retained after centering of the waveform on their negative peak. **E**, Firing rate (top) and waveform amplitude (bottom) as a function of time during the experiment; bin width is 10 s. **F**, Interspike interval histogram that illustrates a relative lack of spike events occurring during the refractory period (red stripe) of the previous spike; bin width is 0.25 ms. **G**, Histogram of spike waveform amplitudes relative to the threshold used for detection. Red line is a Gaussian fit that can be used to estimate the number of waveforms from this putative unit that did not cross threshold.

errors. On the one hand, it is favorable to err on the side of a low threshold since noise events that are erroneously included as spikes can be eliminated when waveforms are clustered. On the other hand, each detected event is typically followed by a censored period that momentarily prevents another spike from being detected. This censor period is included to prevent a single spike waveform from trig-

gering multiple events. If a spike detection threshold is set too low, noise events may cause a high fraction of the data to be censored and true events will go undetected. The number of expected noise events for a particular threshold can be calculated from the background noise and used to set detection criteria (Fig. 2C).

The set of waveforms that cross threshold must be aligned to facilitate their com-

parison. Waveforms are typically aligned on either their peak or center-of-mass, as either of these points is less susceptible to noise than the threshold crossing. Further, the true peak of the waveform may occur in between samples, so splines (de Boor, 2001) can be used to interpolate the waveform about its true peak. With multiwire electrodes, the alignment can be performed on the data channel with the largest event, with the resulting alignment applied to the other channels.

A range of different approaches are used to separate the set of waveforms into groups that are based on similarity of their shape and possible other factors, such as the state of the animal and the statistics of neuronal spiking (Abeles and Goldstein, 1977; Lewicki, 1994; Fee et al., 1996b; Nguyen et al., 2003; Quiroga et al., 2004; Delescluse and Pouzat, 2006). Since the clustering of waveforms into common sources involves statistical inference, a given cluster of waveforms is referred to as a unit rather than a neuron.

Multiwire electrode signals are sorted as one composite waveform whose dimensionality is the number of samples times the number of channels. For example, stereotrode voltages sampled at 30 kHz for a 1.5 ms window have a dimension of 90. As it is difficult to visualize data in greater than even three dimensions, sorting relies on automated algorithms to cluster data. A useful parametric method is to assume that the data consist of multiple Gaussian distributions, where each distribution corresponds to the spikes of a single neuron (Pouzat et al., 2002). While extracellular noise typically follows a Gaussian distribution, non-Gaussian variability in waveforms can occur during bursts of action potentials or as a result of electrode drift (Fee et al., 1996a; Harris et al., 2000; Shoham et al., 2003; Delescluse and Pouzat, 2006). It is therefore advisable to use nonparametric clustering algorithms, which typically involve hierarchical aggregation, where waveforms that are the most similar are progressively grouped until some criteria are reached (Fee et al., 1996b).

A final issue concerns on-line sorting, such as by matched filtering with members of a library of spike waveforms (Thakur et al., 2007; Calabrese and Paninski, 2011). On-line sorting is unavoidable for certain engineering applications, such as neuroprosthetic devices (Tillery and Taylor, 2004), and may provide useful feedback on the placement of electrodes in the course of experimentation. However, on-line sorting is at best approxi-

mate and cannot supplant *post hoc* sorting based on the entire collection of spike waveforms and the statistics of the underlying noise (Fee et al., 1996a,b).

### Visual inspection

We consider a series of visual tests on the output of automated spike sorting routines that address whether a single cluster of waveforms is self-consistent with a single neuron.

#### Inspect the waveforms

The clustered waveforms should be inspected for a nonphysiological shape or obvious contamination by multiple neurons. This can be done by overlaying all clustered waveforms together, which has the advantage of clearly indicating any outlier events, or with a heat map of time-voltage values (Fig. 2D), which has the advantage of indicating the variability of the waveform.

Specialized visualization tools may be required for atypical cases. For example, a systematic change in the shape of the recorded waveforms may occur during epochs of synchronous spiking and may be resolved only by the use of electrodes with a smaller exposed tip.

#### Inspect for stationarity

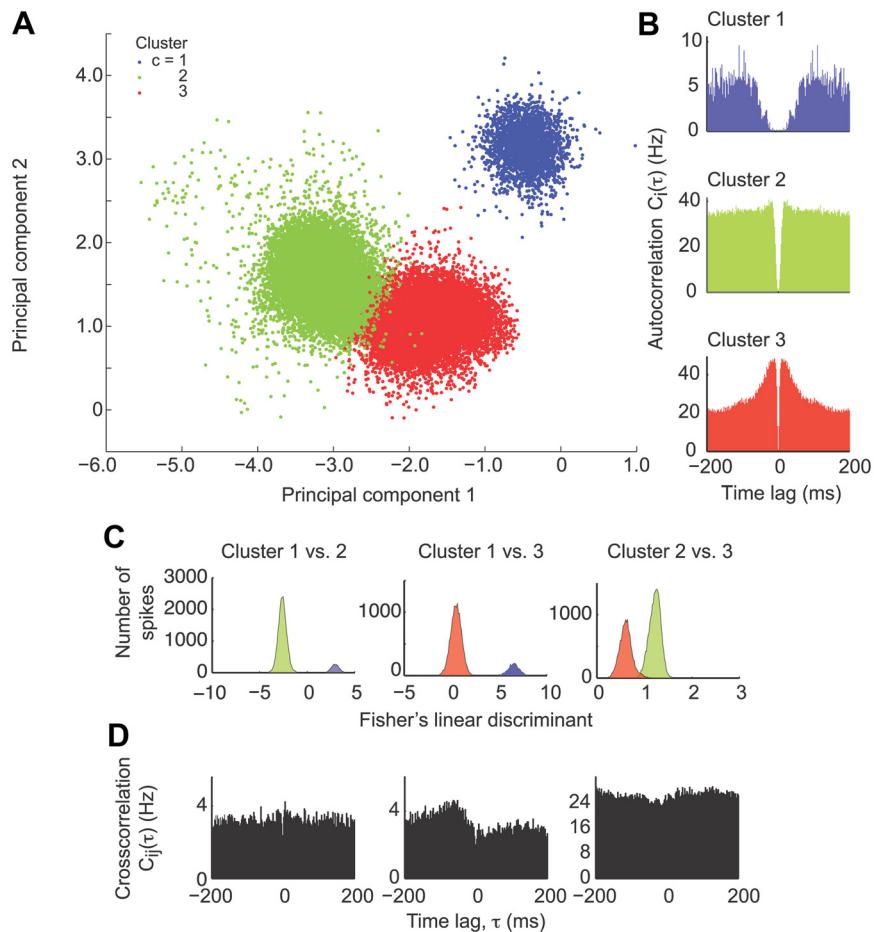
The temporal stability of the waveforms in each cluster can be determined by plotting both the instantaneous spike rate and the amplitude of the waveforms as a function of time (Fig. 2E). The electrodes may have physically drifted or the state of the neuron may have changed if either of these values systematically varies over the course of an experiment.

#### Inspect for refractory violations

A histogram of the interspike interval (ISI) for the spike times of each waveform in a cluster is expected to show a refractory period, i.e., a dearth of spikes that occur within milliseconds of each other (Fig. 2F). An alternate graph is a histogram of the autocorrelation of the spike times, which preferentially highlights patterns of spikes. Note that the value of the ISI asymptotes to zero at large time lags, while that of the autocorrelation asymptotes to the mean firing rate.

#### Verification of waveform threshold

Independent of what metric was used to detect spikes, e.g., the peak voltage, a histogram of the value of that metric for each waveform can be inspected for a sharp cutoff at the threshold level (Fig. 2G). In general, the vast majority of waveforms in



**Figure 3.** Plots that demonstrate the separation and independence of 3 clusters recorded from the same stereotrode. **A**, Scatter plot of spike waveforms from all clusters using the first two principal components. Cluster membership is indicated by color. For clarity, noise events were not plotted. **B**, Autocorrelation of each cluster; bin width is 2 ms. The autocorrelation as calculated as  $C_i(\tau) = \sum_{i \neq j} \delta[\tau - (t_i - t_j)]$ , where  $t_i$  is a spike time. **C**, Pairwise comparison of clusters via histograms of the projection of each cluster onto Fisher's linear discriminant between the given pair of clusters. The linear discriminant, denoted  $\mathbf{H}$ , between the waveforms of clusters labeled  $n$  and  $m$  is given by  $\mathbf{H} = (\bar{\Sigma}_n + \bar{\Sigma}_m)^{-1}(\mu_n - \mu_m)$  where  $\mu_n$  and  $\mu_m$  are the mean vectors and  $\bar{\Sigma}_n$  and  $\bar{\Sigma}_m$  are the covariance matrices of the waveforms in clusters  $n$  and  $m$ , respectively. The projection of a waveform  $\mathbf{v}$  onto the linear discriminant is given by  $\mathbf{H}\mathbf{v}$ . **D**, Pairwise comparison of the three pairs of clusters via the cross-correlation between the spike times in each pair of clusters; bin width is 2 ms. The cross-correlation is calculated as  $C_{ij}(\tau) = \sum_{i,j} \delta[\tau - (t_i^m - t_j^n)]$ , where  $t_i^m$  and  $t_j^n$  are the spike times of clusters  $m$  and  $n$ . For both the autocorrelation and cross-correlation, values were normalized into terms of firing rate.

a cluster must lie well above the threshold for detection.

#### Inspect residuals

The residuals of a cluster, i.e., the standard deviation at each sample of the waveform, can indicate whether multiple neurons contributed to a cluster (Pouzat et al., 2002) (Fig. 2D). Ideally, the variability of a cluster arises only from background noise and matches the statistics of background noise, which appears to be stationary over the duration of the waveform (Fee et al., 1996a). The presence of strong temporal structure in the background noise may indicate contamination from other neurons.

We now turn to comparisons that involve pairs of potential single-unit clus-

ters as a means to identify cross-contamination (Fig. 3A,B).

#### Projection onto Fisher's linear discriminant

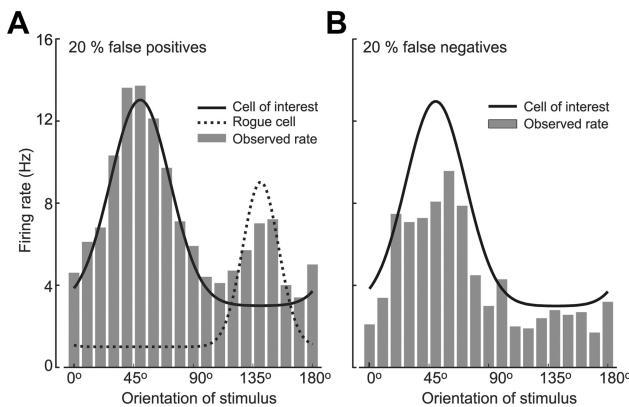
The optimal one-dimensional projection to separate two known Gaussian distributions, given by Fisher's linear discriminant (Fisher, 1936), may be calculated from the mean and covariance matrix of two clusters, even if the data are not strictly Gaussian. As noted by Pouzat et al. (2002), the projection of both clusters onto the discriminant provides a rapid means to visualize potential overlap in their distribution of waveforms (Fig. 3C). A caveat is that an apparent overlap in one dimension may hide a separation in higher dimensions.

**Table 1.** Summary of cluster statistics for three “single” units in Figure 3: false-positive events

	Single cluster error	Multiple cluster error				Composite
		Overlap of clusters				
	ISI violations, $f_1^P$	1–2	2–3	3–1	Sum, $f_2^P$	Computed as Maximum ( $f_1^P, f_2^P$ )
Unit 1	0	0		0	0	0
Unit 2	0.01	0.001	0.005		0.006	0.01
Unit 3	0.12		0.026	0	0.026	0.12

**Table 2.** Summary of cluster statistics for three “single” units in Figure 3: false-negative events

	Single cluster error	Multiple cluster errors				Composite	
		Overlap of clusters					
	Undetected spikes, $f_1^n$	1–2	2–3	3–1	Sum, $f_2^n$	Censored spikes, $f_3^n$	Computed as $[1 - (1 - f_1^n)(1 - f_3^n)] + f_2^n$
Unit 1	0	0		0	0	0.028	0.03
Unit 2	0	0	0.017		0.017	0.016	0.03
Unit 3	0.001		0.008	0	0.008	0.021	0.03

**Figure 4.** Consequences of false-positive and false-negative errors in poorly sorted units. **A**, Tuning histogram of a poorly sorted unit that contains all spikes simulated from the neuron of interest (solid theoretical tuning curve) and contaminating spikes from a simulated rogue neuron (dashed theoretical tuning curve) with a 20% false-positive error rate. The observed tuning histograms were generated as Poisson-distributed spike counts with 10 s of averaging at each orientation. **B**, Tuning histogram of a poorly sorted unit that contains spikes of simulated neuron (solid curve) but with a 20% false-negative error rate. The observed tuning histogram was generated as a Poisson-distributed spike count with 10 s of averaging at each orientation.

### Cross-correlation

The correlation between the spike times of two clusters reveals whether the two clusters are independent (Fig. 3D). A correlation function that appears constant indicates two autonomous spike trains. However, temporal structure in the cross-correlation can arise from many sources, including functional connectivity, receptive field similarity, or cross-contamination of clusters.

### Outlier removal

The final step of spike sorting is the removal of outlier data points from each cluster. An outlier is a waveform that does not resemble any of the mean waveforms. A principled way to remove outliers is to use the  $\chi^2$  probability distribution. For a random variable  $\mathbf{v}$  drawn from a multivariate Gaussian distribution with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ , the distribution of  $(\mathbf{v} - \boldsymbol{\mu})^T \Sigma^{-1}$

$(\mathbf{v} - \boldsymbol{\mu})$  follows a  $\chi^2$  distribution. One could remove from a cluster all waveforms that have a probability of  $<1/N$ , where  $N$  is the number of events in the cluster. This process should be performed last so that the mean waveform is defined for each unit represented in a dataset.

### Quality metrics

Manual inspection provides a qualitative means to evaluate the success of spike sorting. Yet an estimate of false-positive and false-negative errors should be reported as a means to evaluate a given level of modulation of a unit's firing rate relative to the contamination of the cluster. We propose that the following four estimates of false-positive and false-negative errors be included in publications that make use of spike sorting, independent of the algorithm used to sort.

### False positives based on refractory period violations, denoted $f_1^P$

We assume that the majority of spikes in a cluster arise from a single neuron. We consider these the true events and recall that the absolute refractory period of a neuron implies that the ISI distribution should drop to zero below a minimum time. Thus, ISIs that are shorter than the absolute refractory period represent contamination by a rogue unit. One can compute the fractional level of contamination,  $f_1^P$ , with the assumption that refractory period violations arise from a point process that is uncorrelated with the true spikes of the cluster (Meunier et al., 2003).

We denote the number of spike events in a cluster as  $N$ , the width of the refractory period as  $\tau_R$ , the width chosen for the censored period as  $\tau_C$ , and the duration of the experiment as  $T$ . In terms of these definitions, the number of true spikes in the cluster is  $N(1 - f_1^P)$  and the number of rogue spikes is  $Nf_1^P$ . The total duration during which refractory period violations could occur around true spikes is  $2(\tau_R - \tau_C)N(1 - f_1^P)$ ; the factor of two arises since refractory period violations occur whether a rogue spike appears immediately before or after a true spike. The expected number of refractory period violations,  $r$ , is given by the total duration of refractory periods around true spikes multiplied by the rate of rogue spikes, or  $r = 2(\tau_R - \tau_C)N^2(1 - f_1^P)f_1^P/T$ , which can be solved for  $f_1^P$ . For example, under the typical conditions of mean firing rate,  $N/T = 10$  Hz,  $\tau_R = 3$  ms,  $\tau_C = 1$  ms, and  $T = 1000$  s, an observation of  $r = 20$  refractory period violations yields an estimate of  $f_1^P = 0.05$ . These values, particularly the refractory period, are typical for measurements from neocortical pyramidal cells in awake rat (Curtis and Kleinfeld, 2009) and have to be adjusted for other cell types and experimental conditions.

### False negatives from the threshold for detection, denoted $f_1^n$

The histogram of the values of the spike detection metric (Fig. 2G) may be used to estimate the number of spikes whose waveforms were below the threshold for detection. For the case of a voltage threshold, as used here, the histogram may follow a Gaussian distribution. However, as the tail of the data is missing, the estimate of the mean and variance of this distribution may require a special fitting procedure. Once the parameters are fit, the percentage of missing spikes is the value of

the cumulative Gaussian distribution up to the detection threshold.

*False positives and false negatives from the overlap between pairs of clusters of sorted spike waveforms, denoted  $f_2^P$  and  $f_2^N$ , respectively*

We estimate the overlap between a pair of clusters in terms of fits with multivariate Gaussian waveforms to each cluster. Let  $V$  represent the voltage waveform of an extracellular spike that is assigned to either cluster 1, with mean waveform  $\mu_1$ , or cluster 2, with mean waveform  $\mu_2$ . We use a model in which the variability of the waveforms in either cluster are characterized by Gaussian noise with a covariance matrix  $\bar{\Sigma}_1$  for units assigned to cluster 1 or  $\bar{\Sigma}_2$  for those assigned to cluster 2; the dimensionality of the  $\bar{\Sigma}$  values is the number of samples in the composite waveform. All of these parameters are estimated from the data with the expectation-maximization (EM) algorithm (`gmdistribution.fit` in MATLAB) to a mixture model of two multivariate Gaussians; as a technical aside, one must calculate pseudoinverses of the  $\bar{\Sigma}$  values as they are not typically full rank. The relative number of spikes in each cluster is given by the prior distributions, i.e.,  $P(C = 1)$  and  $P(C = 2)$ , where  $C$  labels the cluster, also found by the EM algorithm, where  $P(C = 1) + P(C = 2) = 1$ .

The probability that a cluster generates a particular waveform,  $v$ , is thus  $P(V = v|C = c) = (2\pi)^{-\frac{nk}{2}} |\bar{\Sigma}_c|^{-\frac{1}{2}} e^{-\frac{1}{2}(v - \mu_c)^T \bar{\Sigma}_c^{-1} (v - \mu_c)}$ , where  $c$  indexes the cluster. The inverse problem of finding that a particular waveform belongs to cluster  $c$ , denoted  $P(C = c|V = v)$ , is found using Bayes' rule, i.e.,  $P(C = c|V = v) = P(V = v|C = c)P(C = c)/P(V = v)$ , where  $P(V = v) = \sum_{i=1,2} P(V = v|C = i)P(C = i)$ . To calculate the fraction of false positives in cluster 1, we calculate the probability that each waveform assigned to cluster 1 was generated by cluster 2, i.e., false positives assigned to cluster 1, as  $f^P(1;2) = (1/N_1) \sum_{v \in \text{cluster } 1} P(C = 2|V = v)$ , where  $N_1$  is the number of spikes in cluster 1. The expressions for the remaining fractions of false negatives and positives are  $f^N(1;2) = (1/N_1) \sum_{v \in \text{cluster } 2} P(C = 1|V = v)$ ,  $f^P(2;1) = (1/N_2) \sum_{v \in \text{cluster } 2} P(C = 1|V = v)$ , and  $f^N(2;1) = (1/N_2) \sum_{v \in \text{cluster } 1} P(C = 2|V = v)$ , where  $N_2$  is the number of spikes in cluster 2. This pairwise analysis can be generalized for  $n$  clusters for which the  $(n - 1)$  fractions of false positives for a given cluster,  $k$ , sum to form  $f_2^P = \sum_{i \neq k} f^P(k; i)$ . A similar summation of the  $(n - 1)$  fractions of false-negative events yields  $f_2^N$  for a given cluster.

*False negatives from censored events, denoted  $f_3^N$*

All detected events outside of a cluster create a brief period of time that potentially censors the detection of true spikes for that cluster. One can estimate the total time that the dataset was censored by multiplying the duration of the censored period, denoted  $\tau_C$ , by the total number of detected events minus the number of events in the cluster of interest, i.e.,  $M_k = \sum_{i \neq k}^{\text{all clusters}} N_i$ , where  $N_i$  denotes the number of spikes in the  $i$ th cluster. The fraction of time that was censored contributes to the estimate of false-negative errors, i.e.,  $f_3^N = M_k \tau_C / T$ .

### Summary matrices

The summary statistics from the above analysis form two matrices, one for the fraction of false-positive events and the other for the fraction of false-negative events. The ISI violations should include false positives from the fractional overlaps of a cluster with all other clusters, i.e., we expect  $f_1^P \geq f_2^P$ . As a hedge against the possibility that ISI violations were underestimated, the composite fraction of false-positive events is taken as the larger of the false-positive estimates based on ISI violations or cluster overlaps (Table 1). The fraction of undetected spikes and the fraction of censored spikes are independent, so these fractions are combined as a product of their complements (Table 2). However, these detection errors are mutually exclusive of false-negative errors that result from the overlap of clusters, so these fractions are added to form a composite fraction of false-negative events (Table 2).

## Discussion

Extracellular recording is the oldest and most common method of recording electrical activity across populations of neurons in awake behaving animals, from invertebrates to human primates. Yet simple criteria for acceptable data, particularly with regard to claims of single unit responses, are largely missing. Such criteria are critical since interpretations of spike trains that are based on inadequately sorted units can lead to erroneous claims on neural coding. For the case of false-positive errors, contamination can result in sorted units that exhibit false multimodal responses (Fig. 4A). This occurs when spike trains from two differently tuned cells are combined into a single spike train. For the case of false-negative errors, the true firing rate is underestimated (Fig. 4B) and thus the signal-to-noise ratio is decreased for statistics that are calculated with the spike train. In ad-

dition, false-negative errors may include temporal structure, such as when spikes are hidden by stimulation artifacts or are misassigned because of a reduction in amplitude and an increase in width for the trailing spikes of a burst.

We suggest that investigators include a matrix of criteria that define the levels of errors for their units (Tables 1, 2) (see <http://physics.ucsd.edu/neurophysics/links.html> for accompanying MATLAB code), or at the minimum report a matrix of upper bounds for these errors across all units. Further, any analysis of neuronal modulation must be made in light of the estimated contamination of the unit; e.g., a feature based on a 20% modulation of the firing rate is not significant for that unit if the measured count contains a 20% false-positive error rate (Fig. 4A). Following Joshua et al. (2007), we note that while calculation of errors involves statistical assumptions about the dataset, even if these assumptions do not hold true, it is more useful to report estimated false-negative and false-positive events for a dataset rather than declare that “only well-isolated units were included in our study.”

Our analysis is practical for even particularly large datasets, such as those that originate from arrays of electrodes (Blum et al., 1991; Churchland et al., 2007). Continuous storage of signals from a large number of channels may be streamed to disk in real time and is compatible with the bandwidth of radio telemetry systems (Tillery and Taylor, 2004); e.g., recordings from 100 stereotrodes leads to a data rate of only 10 Mbytes/s. For well separated electrodes, or sets of stereotrodes or tetrodes, one can cluster the waveform data in parallel. The computational methods we discuss may be scaled up and the results may be subject to rapid visual inspection (Figs. 2, 3). For situations with hundreds to thousands of electrodes in which continuous data storage is currently impractical, we suggest that the threshold for detection be biased low and a time series that is at least twice the period of the spike waveform plus censored period around each event be saved. Overlapping epochs may need to be saved when the spike is high. This procedure allows the threshold to be raised *post hoc* to minimize the number of censored events (Fig. 2C).

## Notes

Supplemental material for this article is available at <http://physics.ucsd.edu/neurophysics/links.html>. This material has not been peer reviewed.

## References

- Abeles M, Goldstein MH (1977) Multispike train analysis. *Proc IEEE* 65:762–773.
- Ainsworth A, O'Keefe J (1977) A lightweight microdrive for the simultaneous recording of several units in the awake, freely moving rat. *J Physiol* 269:8–10.
- Battaglia FP, Kalenscher T, Cabral H, Winkel J, Bos J, Manuputy R, van Lieshout T, Pinkse F, Beukers H, Pennartz C (2009) The Lantern: an ultra-light micro-drive for multi-tetrode recordings in mice and other small animals. *J Neurosci Methods* 178:291–300.
- Blum NA, Carkhuff BG, Charles HK Jr, Edwards RL, Meyer RA (1991) Multisite microprobes for neural recordings. *IEEE Trans Biomed Eng* 38:68–74.
- Buzsáki G (2004) Large-scale recording of neuronal ensembles. *Nat Neurosci* 7:446–451.
- Calabrese A, Paninski L (2011) Kalman filter mixture model for spike sorting of non-stationary data. *J Neurosci Methods* 196: 159–169.
- Churchland MM, Yu BM, Sahani M, Shenoy KV (2007) Techniques for extracting single-trial activity patterns from large-scale neural recordings. *Curr Opin Neurobiol* 17:609–618.
- Curtis JC, Kleinfeld D (2009) Phase-to-rate transformations encode touch in cortical neurons of a scanning sensorimotor system. *Nat Neurosci* 12:492–501.
- de Boor CD (2001) A practical guide to splines. Revised edition. New York: Springer.
- Delescluse M, Pouzat C (2006) Efficient spike-sorting of multi-state neurons using interspike intervals information. *J Neurosci Methods* 150:16–29.
- Duda RO, Hart PE (1973) Pattern classification and scene analysis. New York: Wiley-Interscience.
- Fee MS, Leonardo A (2001) Miniature motorized microdrive and commutator system for chronic neural recording in small animals. *J Neurosci Methods* 112:83–94.
- Fee MS, Mitra PP, Kleinfeld D (1996a) Variability of extracellular spike waveforms of cortical neurons. *J Neurophysiol* 76:3823–3833.
- Fee MS, Mitra PP, Kleinfeld D (1996b) Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *J Neurosci Methods* 69:175–188.
- Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Ann Eugenics* 7:179–188.
- Ganguly K, Kleinfeld D (2004) Goal-directed whisking behavior increases phase-locking between vibrissa movement and electrical activity in primary sensory cortex in rat. *Proc Natl Acad Sci U S A* 101:12348–12353.
- Gray CM, Maldonado PE, Wilson M, McNaughton B (1995) Tetrodes markedly improve the reliability and yield of multiple single-unit isolation recordings in cat striate cortex. *J Neurosci Methods* 63:43–54.
- Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsáki G (2000) Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J Neurophysiol* 84:401–414.
- Henze DA, Borhegyi Z, Csicsvari J, Mamiya A, Harris KD, Buzsáki G (2000) Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J Neurophysiol* 84:390–400.
- Humphrey DR, Schmidt EM (1991) Extracellular single-unit recording methods. In: *Neurophysiological techniques: applications to neural systems* (Boulton AA, Baker GB, Vanderwolf CH, eds), pp 1–64. New York: Springer.
- Joshua M, Elias S, Levine O, Bergman H (2007) Quantifying the isolation quality of extracellularly recorded action potentials. *J Neurosci Methods* 163:267–282.
- Kim KH, Kim SJ (2000) Neural spike sorting under nearly 0-dB signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier. *IEEE Trans Biomed Eng* 47:1406–1411.
- Krüger J, Bach M (1981) Simultaneous recording with 30 microelectrodes in monkey visual cortex. *Exp Brain Res* 41:191–194.
- Lewicki MS (1994) Bayesian modeling and classification of neural signals. *Neural Comput* 6:1005–1030.
- Lewicki MS (1998) A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* 9:R53–R78.
- Logothetis NK, Kayser C, Oeltermann A (2007) In vivo measurement of cortical impedance spectrum in monkeys: implications for signal propagation. *Neuron* 55:809–823.
- McNaughton BL, O'Keefe J, Barnes CA (1983) The stereotrode: a new technique for simultaneous isolation of several units in the central nervous system from multiple unit records. *J Neurosci Methods* 8:391–397.
- Meunier N, Marion-Poll F, Lansky P, Rospars JP (2003) Estimation of the individual firing frequencies of two neurons recorded with a single electrode. *Chem Sens* 28:671–679.
- Nguyen DP, Frank LM, Brown EN (2003) An application of reversible-jump Markov chain Monte Carlo to spike classification of multi-unit extracellular recordings. *Network* 14: 61–82.
- Pouzat C, Mazor O, Laurent G (2002) Using noise signature to optimize spike-sorting and to assess neuronal classification quality. *J Neurosci Methods* 122:43–57.
- Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput* 16:1661–1687.
- Reitboeck HJ (1983) A 19-channel matrix drive with individually controllable fiber micro-electrodes for neurophysiological applications. *IEEE Trans Syst Man Cybern* 113: 626–683.
- Shoham S, Fellows MR, Normann RA (2003) Robust, automatic spike sorting using mixtures of multivariate *t*-distributions. *J Neurosci Methods* 127:111–122.
- Thakur PH, Lu H, Hsiao SS, Johnson KO (2007) Automated optimal detection and classification of neural action potentials in extracellular recordings. *J Neurosci Methods* 162: 364–376.
- Tillery SI, Taylor DM (2004) Signal acquisition and analysis for cortical control of neuroprosthetics. *Curr Opin Neurobiol* 14:758–762.
- Venkatachalam S, Fee MS, Kleinfeld D (1999) Ultra-miniature headstage with 6-channel drive and vacuum-assisted micro-wire implantation for chronic recording from neocortex. *J Neurosci Methods* 90:37–46.
- Yamamoto J, Wilson MA (2008) Large-scale chronically implantable precision motorized microdrive array for freely behaving animals. *J Neurophysiol* 100:2430–2440.