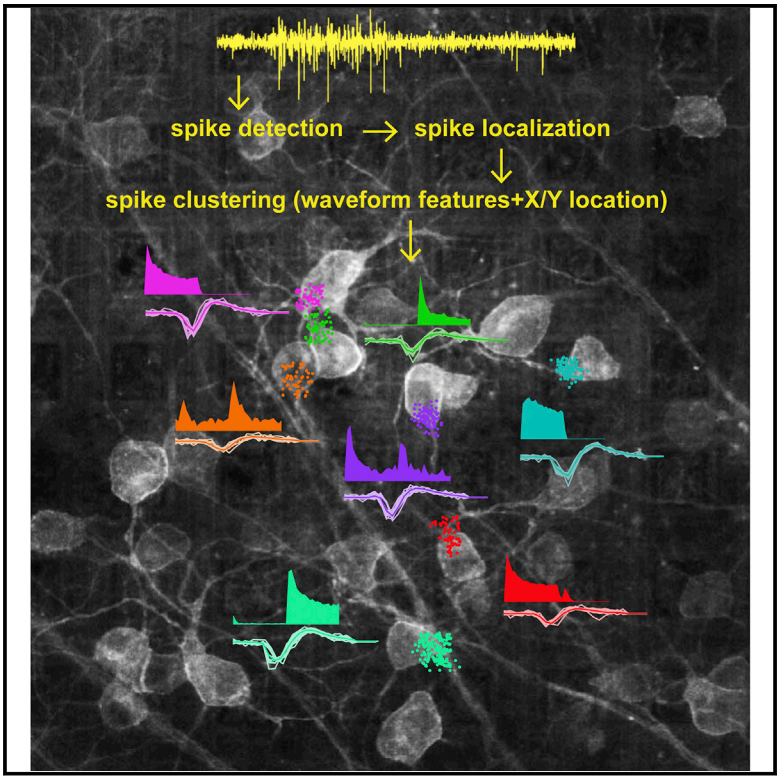
Cell Reports

Resource

Unsupervised Spike Sorting for Large-Scale, High-Density Multielectrode Arrays

Graphical Abstract



Authors

Gerrit Hilgen, Martino Sorbaro, Sahar Pirmoradian, ..., Francesca Cella Zanacchi, Evelyne Sernagor, Matthias Helge Hennig

Correspondence

m.hennig@ed.ac.uk

In Brief

Data volume and complexity make spike sorting for large-scale extracellular recordings computationally extremely challenging. Hilgen et al. introduce a method enabling analysis of recordings with thousands of channels and provide tools for automated quality control and unit selection.

Highlights

- An automated spike sorting method for dense, large-scale recordings is presented
- Efficient data representation enables sorting of thousands of channels
- Automated unit selection through model-based quality control
- Conventional spike sorting frequently fails under non-optimal signal conditions

Hilgen et al., 2017, Cell Reports *18*, 2521–2532
March 7, 2017
http://dx.doi.org/10.1016/j.celrep.2017.02.038

CellPress

Cell Reports
Resource

Unsupervised Spike Sorting for Large-Scale, High-Density Multielectrode Arrays

Gerrit Hilgen,[1,10] Martino Sorbaro,[2,3,10] Sahar Pirmoradian,[2] Jens-Oliver Muthmann,[2,4,5] Ibolya Edit Kepiro,[6,7,11] Simona Ullo,[8] Cesar Juarez Ramirez,[2] Albert Puente Encinas,[2] Alessandro Maccione,[9] Luca Berdondini,[9] Vittorio Murino,[8] Diego Sona,[8] Francesca Cella Zanacchi,[6] Evelyne Sernagor,[1] and Matthias Helge Hennig[2,12,*]

[1]Institute of Neuroscience, Newcastle University, Newcastle NE2 4HH, UK
[2]Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK
[3]Department of Computational Biology, School of Computer Science and Communication, Royal Institute of Technology, Stockholm 100 44, Sweden
[4]Manipal University, Manipal 576104, India
[5]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India
[6]Nanophysics (NAPH), Istituto Italiano di Tecnologia, Genova 16163, Italy
[7]Faculty of Science, Engineering and Computing, Kingston University, Kingston KT1 2EE, UK
[8]Pattern Analysis and Computer Vision (PAVIS), Istituto Italiano di Tecnologia, Genova 16163, Italy
[9]Neuroscience and Brain Technologies (NBT), Istituto Italiano di Tecnologia, Genova 16163, Italy
[10]Co-first author
[11]Present address: National Physical Laboratory (NPL), Teddington TW11 0LW, UK
[12]Lead Contact
*Correspondence: m.hennig@ed.ac.uk
http://dx.doi.org/10.1016/j.celrep.2017.02.038

SUMMARY

We present a method for automated spike sorting for recordings with high-density, large-scale multielectrode arrays. Exploiting the dense sampling of single neurons by multiple electrodes, an efficient, low-dimensional representation of detected spikes consisting of estimated spatial spike locations and dominant spike shape features is exploited for fast and reliable clustering into single units. Millions of events can be sorted in minutes, and the method is parallelized and scales better than quadratically with the number of detected spikes. Performance is demonstrated using recordings with a 4,096-channel array and validated using anatomical imaging, optogenetic stimulation, and model-based quality control. A comparison with semi-automated, shape-based spike sorting exposes significant limitations of conventional methods. Our approach demonstrates that it is feasible to reliably isolate the activity of up to thousands of neurons and that dense, multi-channel probes substantially aid reliable spike sorting.

INTRODUCTION

Large-scale, dense probes and arrays and planar multielectrode arrays (MEAs) enable extracellular recordings of thousands of neurons simultaneously (Ballini et al., 2014; Berdondini et al., 2005; Eversmann et al., 2003; Frey et al., 2010; Hutzler et al., 2006; Maccione et al., 2014; Müller et al., 2015; Obien et al., 2015). Exploiting such data requires the reliable isolation of extracellularly recorded spikes generated by single neurons (spike sorting), a computationally costly task that is difficult to scale up to large numbers of recording channels (Rey et al., 2015). For conventional devices with up to tens of recording channels, a typical workflow consists of initial event detection, followed by semi-automated clustering based on spike waveform differences, followed by manual inspection and refinement. If the recording channels are sufficiently well separated, then there is no or little overlap between their signals, and spike sorting can be performed by clustering a low-dimensional representation of spike shapes (Harris et al., 2000; Lewicki, 1998; Quiroga et al., 2004).

This approach is inappropriate for dense, large-scale recordings. First, on dense MEAs, spike sorting becomes a complex assignment problem because not only multiple neurons contribute to the compound signal recorded on distinct channels, but individual spikes are also recorded by several neighboring channels simultaneously (Prentice et al., 2011; Rossant et al., 2016). Events are thus described by multiple waveforms and their locations, with an exponential number of potential assignments that can only be tackled using approximate algorithms. Second, the size of the datasets makes extensive manual intervention impractical; hence, as much of the process as possible, including quality control, should be automated.

Much of the variability in spike shapes is due to measuring them at different positions relative to the neuron. In conventional recordings, relatively small signals are measured using large electrodes averaging currents originating from different parts of the neuron. High-density MEAs with small electrodes detect primarily strong currents at the axon initial segment (AIS). The mechanism for generating action potentials is thus represented with a higher weight in the measured signals, leading to less variability in measured spike shapes. Existing solutions, demonstrated on data from hundreds of channels, are either template-matching methods (Marre et al., 2012; Prentice et al.,

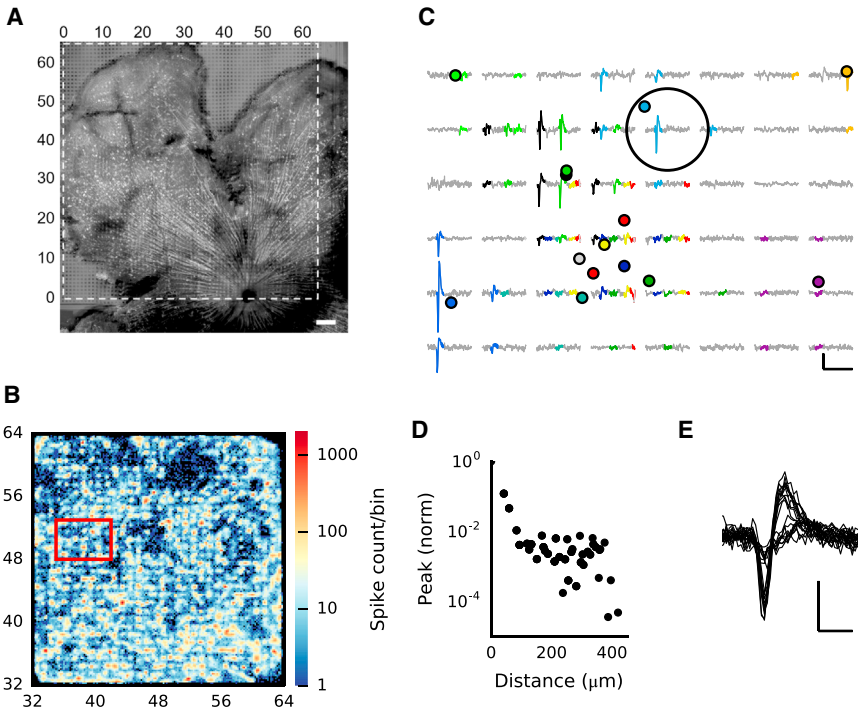


Figure 1. Spatial Event Localization Reveals Isolated Spike Clusters

(A) Confocal image of a Thy1-ChR2-YFP retina expressing yellow fluorescent protein under the Thy1 promoter, placed on the array for recording. Electrodes can be seen as small squares in areas not covered by the retina. The active area of the array is indicated by dashed lines. Scale bar, 200 μm.

(B) Activity map of a quarter of the array after spatial event localization. Spike counts are shown as a density plot, spatially binned with 8.4 μm resolution. Spikes cluster in distinct groups in space, presumably originating from individual neurons.

(C) Several detected events (rectangle in B), shown at their estimated locations (colored circles), and the corresponding episodes in the raw data (colored traces). Scale bars, 5 ms and 200 μV.

(D) Average peak signal decay for detected events as a function of distance. On average, a significant signal is detectable in an area of 100μm around the spike peak location. This plot is based on signal peaks at the spike time ± 2 recording frames, so signals beyond 200 μm reflect primarily noise.

(E) Twenty randomly selected spike shapes for events localized within the area marked by the large circle in (C), indicating the presence of signals from at least two different neurons at this location. Scale bars, 5 ms and 200 μV.

(B–E) The same dataset acquired at 24 kHz on 32 × 32 channels (A shows a different retina).

2011), or the removal of uninformative spike features to make fitting of a mixture model computationally feasible (Rossant et al., 2016).

Here we present a very fast and fully automated method for spike sorting. Dense sampling enabled us to obtain a rough estimate of a source location for each detected event (Muthmann et al., 2015), yielding dense, spatially separated clusters originating from single neurons, as demonstrated using optogenetic stimulation and confocal imaging. Average waveforms are obtained for each event, with noise reduced by signal interpolation. Shape features extracted from this waveform are then combined with spatial locations so that the clustering problem is reduced to finding local density peaks in few dimensions.

We demonstrate this method using light responses in the mouse retina and spontaneous activity in cell cultures recorded with a 4,096-channel MEA. A direct comparison with conventional spike sorting also exposes severe and hard to detect limitations of the latter. A parallelized implementation of this method that is capable of sorting millions of spikes within a few minutes on a fast workstation, as well as a tool for data visualization, can be downloaded at https://github.com/martinosorb/herding-spikes.

RESULTS

Spatial Spike Localization

Figure 1A illustrates a retinal whole-mount placed on the MEA. Spikes are detected using a threshold-based method that exploits dense sampling to improve detection performance and assigns each spike an estimated location based on the barycenter of the spatial signal profile (Muthmann et al., 2015). This procedure yields spatio-temporal event maps, where each event is identified by a time stamp, two spatial coordinates, and a single interpolated waveform. The resulting spatial activity maps provide a higher spatial resolution for spike locations than given by the electrode pitch (Figure 1B). Spikes were found in dense clusters surrounded by areas of low event density. The relationship between recorded signals and spike locations is illustrated in Figure 1C, where estimated spike locations are shown together with corresponding raw data segments from nearby electrodes. The examples show how spike locations relate to the spatial decay of the voltage peaks and that the decay was sufficiently wide to estimate their peak locations (Figure 1D; Pettersen and Einevoll; 2008; Lindén et al., 2011; Mechler et al., 2011). Thus, on dense MEAs, event locations provide a compact summary of the spatial activity footprint for each spike. Inspecting waveforms, however, reveals the presence of multiple units in small areas (Figure 1E), demonstrating that clustering spatial locations alone is insufficient for reliable single-unit isolation (Prentice et al., 2011).

Combined Spatial and Shape-Based Clustering

Next, spikes are clustered using a combination of their estimated locations and dominant waveform features, extracted via principal-component analysis (PCA), which provide a complementary, compact description of the events. The location estimate is an effective way of summarizing the spatial footprint each spike leaves on the array, whereas waveforms enable the separation of spatially overlapping sources, and they remove ambiguities at spatial cluster boundaries.

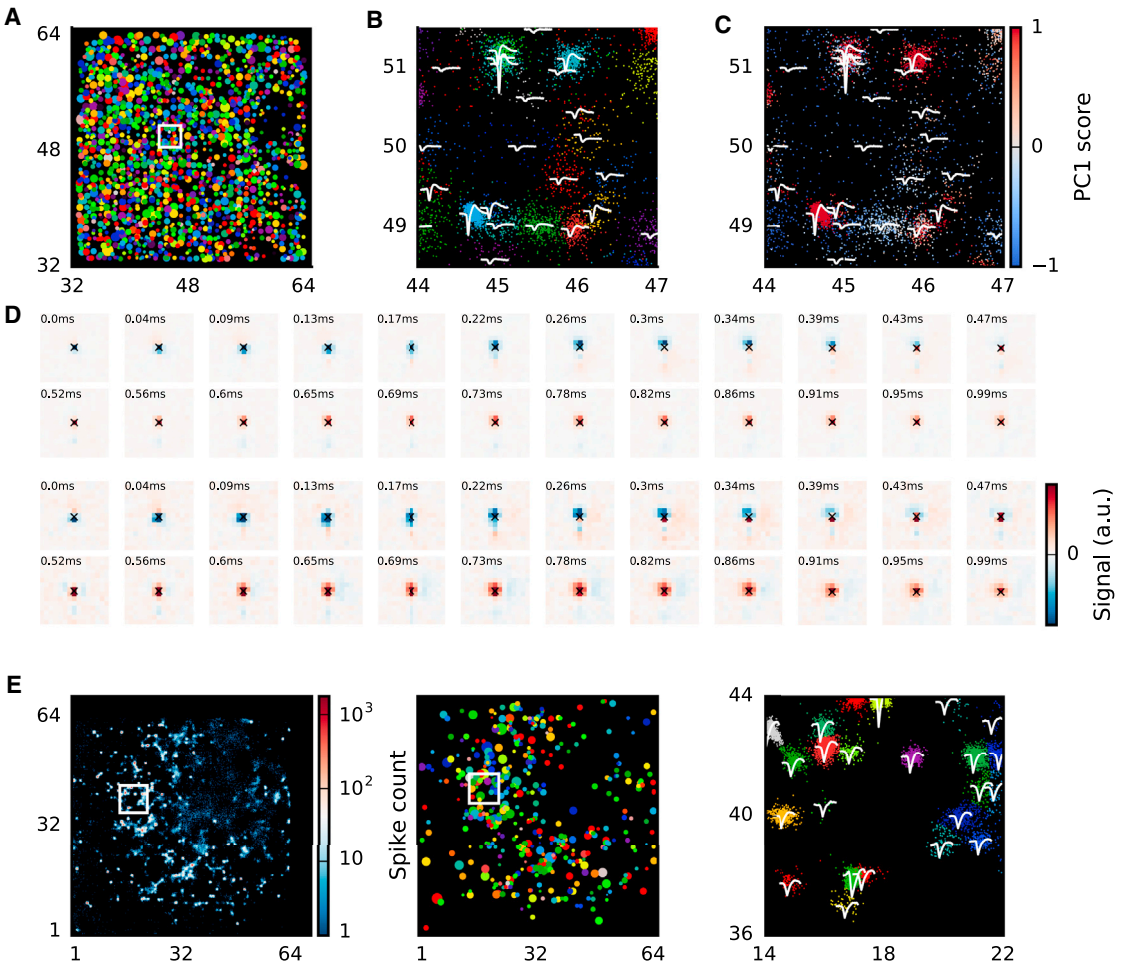


Figure 2. Illustration of Clustered Spike Data

(A) Overview of all single units obtained by clustering a retinal dataset (the same as in Figure 1, acquired at 24 kHz), shown as circles at their estimated locations in array coordinates. Circled areas are proportional to firing rates.

(B) Magnified view of a group of units (the area in the white rectangle in A), showing a subset of spikes at their estimated locations (dots, colored by unit membership; the same colors as in A) and the average waveform associated with each unit.

(C) As (B) but with spike colors encoding the magnitude of the spike waveform projection along the first principal component (PC1 score). Higher scores represent bi-phasic waveforms and low scores weak deflections without a clear bi-phasic shape.

(D) Electrical images for two units. Negative signals relative to baseline are colored in blue and positive signals in red. The cross indicates the centroid of the spike locations. Each square represents one electrode; 15 × 15 (0.63 mm × 0.63 mm) electrodes are shown. Axonal propagation can be seen, moving downward toward the optic disk.

(E) Clustered recording from a hippocampal culture. Shown are raw spike counts (left), all units obtained during the clustering step (center), and a magnified view of a small area of the MEA showing individual spikes and average unit waveforms (right). This recording was acquired with 4,096 channels at a 7-kHz sampling rate, and a waveform classifier was used to remove noise prior to clustering (Figure S1).

The mean shift algorithm was used for clustering, with the number of clusters automatically determined and controlled by a single scale parameter (Comaniciu and Meer, 2002). Clusters are formed by moving spikes along density gradients and augmented by local differences in spike waveforms. Including the first two principal components was sufficient to successfully isolate single units, reducing the high dimensional assignment problem to four-dimensional clustering, which can be performed in minutes for millions of events. In addition to the scale parameter, this method also requires a mixing coefficient for the shape information.

Figures 2A–2C show the result of clustering waveforms acquired at 24 kHz from 1,024 channels, yielding 440,000 spikes separated into 1,600 units. Cluster sizes ranged from tens of spikes to several thousands, corresponding to firing rates ranging from 0.1 to 30 Hz. In a magnified view, Figure 2B shows that units may indeed spatially overlap but are well separated by their waveform features. Overall, units with clearly bi-phasic and large-amplitude waveforms tend to form the more spatially coherent clusters, whereas smaller events are spatially more spread out.

Units with small waveforms originate from neurons with weak signals detected because of low thresholding during the

detection step to avoid false negatives. The first principal component (PC) projection (PC1) for the events is a good indicator of their biphasic character, and, using the convention that positive values always coincide with more biphasic waveforms, this measure may be used to (de)select units for subsequent analysis (Figure 2C). A more precise method, used for all recordings performed at lower sampling rates (<10 kHz), is to train a classifier to pre-select valid spikes prior to clustering based on salient waveform features (Figure S1). This method reliably removes noise because the classifier is well adjusted to the specific recording conditions. Importantly, however, this step is not required for sampling rates of more than 10 kHz.

As a first assessment of the clustered units, we generated electrical images for individual units (Figure 2D). These images provide a spatio-temporal representation of the raw signal recorded around the time of spiking and is generated as a spike-triggered average of the signal on each electrode. Of 406 inspected units with at least 100 spikes, all but one had an estimated location within 40 μm of the electrode that contained the peak signal (median distance, 9.7 μm), indicating that units are indeed well aligned with their spatio-temporal electrical footprint. Furthermore, the recordings were of sufficient detail to isolate axonal propagation (Figure 2D), characterized by a separate, weak positive peak followed by a negative peak traveling downward (toward the optic disk). Because these events peak within less than 100 μs of the main signal, they are not detected as separate events but, instead, introduce a small bias on the location estimates during spike localization.

We also tested our method on activity recorded from cultured hippocampal neurons. Figure 2E illustrates that isolation of single units is also feasible for these preparations, although here the spike localization was less precise than in the retina. We attribute this to a larger effective conductivity in the space above the electrodes, resulting in smaller signal amplitudes, which, in turn, increases the influence of noise on localization (Ness et al., 2015). Such conductivity is likely much lower for the 200- to 300-μm-thick retina, leading to larger and more precisely localizable signals. Ness et al. (2015) show that even small MEA-tissue gaps strongly reduce the signal amplitudes, a likely explanation for the clear, sharp boundaries between areas with and without recorded spikes. Nevertheless, spikes in cultures were typically spatially well clustered, and waveform differences had sufficient detail to allow separation of overlapping units (Figure 2E, right).

Waveform Features Are Essential for Reliable Clustering

To assess the importance of waveform features for sorting and the role of the mixing coefficient α, we compared the correlations between all waveforms within each unit with cross-correlations of waveforms between this unit and its closest neighbor or all nearby spikes within a radius of 42 μm (electrode pitch; Figures 3A–3C). A well sorted unit is expected to have high within-correlations and smaller cross-correlations. Figure 3A shows an example where spatial clustering was sufficient to isolate a unit. Correlations after clustering spatial locations alone ($\alpha = 0$) are very similar to those obtained when waveforms are added ($\alpha = 0.3$), with few spikes re-assigned based on their waveform features. In contrast, Figures 3B and 3C illustrate examples with two clearly distinct units with spatial overlap that could only be separated by waveform features. Increasing α increases self-correlation, with lower cross-correlations for nearby events with sufficiently distinct waveforms in other units (Figure 3B). However, some high cross-correlations can remain for similar but spatially well separated units (Figure 3C).

To quantify the separability of these distributions, we computed the area under the receiver operating characteristic (ROC) curves (AUC), constructed from the distributions of self-correlations and correlations with events in the nearest unit (Figure 3E) or all neighboring events (Figure 3F). The AUC was calculated as the integral of the area spanned by the probability of finding a self-correlation above a sliding threshold, as a function of the probability of finding a cross-correlation above this threshold (true positives versus false positives), so that a value of 1 corresponds to perfectly separated distributions, whereas 0 indicates full overlap.

The median AUC for all units increases with α before plateauing at values about $\alpha \approx 0.4$ (Figure 3D), indicating that the combined features overall improved separation into single units. The AUC distributions show that this effect is substantial (Figures 3E and 3F). Although spatial clustering alone only yielded three (of 788 units with more than 100 spikes) units with AUC > 0.9 compared with events from its closest neighbor, this increases to 130 (of 956 units with more than 100 spikes) for $\alpha = 0.32$. This number rapidly increases when $\alpha \approx 0.25$ and plateaus for larger values, indicating that the precise choice of this parameter is not critical. It is important to note that, although high AUC values indicate well isolated units based on waveform features alone, units with a small AUC should not be rejected because they may still be spatially well isolated.

In summary, waveform features help both to refine existing units found by spatial clustering and to separate spatially overlapping units. Event locations and waveforms provide an effective complementary approach of summarizing the key features of the spatio-temporal footprint left by spikes on the array.

Validation with Optogenetics and Anatomical Imaging

To test whether the detected units indeed correspond to single neurons, we used Thy1-ChR2-YFP retinas (see Experimental Procedures) expressing Channelrhodopsin-2 (ChR2), a light-gated cation channel, under the Thy1 promoter in about half of all retinal ganglion cells (RGCs) (Raymond et al., 2008). This allowed us to stimulate spiking exclusively in a subset of visually identifiable RGCs to clearly establish correlates between single spike-sorted units and individual RGCs.

We first compared the photoreceptor-driven activity recorded during normal light stimulation (irradiance 4 $\mu W/cm^2$, full field flashes at 0.5 Hz) with recordings obtained when these light responses were blocked with 20 μM 6,7-dinitroquinoxaline-2,3-dione (DNQX) and L-AP4, and ChR2-mediated spikes evoked at maximum irradiance (0.87 mW/cm^2; Figure 4A). The activity maps show that only a subset of all RGCs responded to optogenetic stimulation (Figure 4A, top and center). We found 375 units in that dataset with a firing rate of at least 0.5 Hz during photoreceptor-driven light stimulation but only 254 units during direct

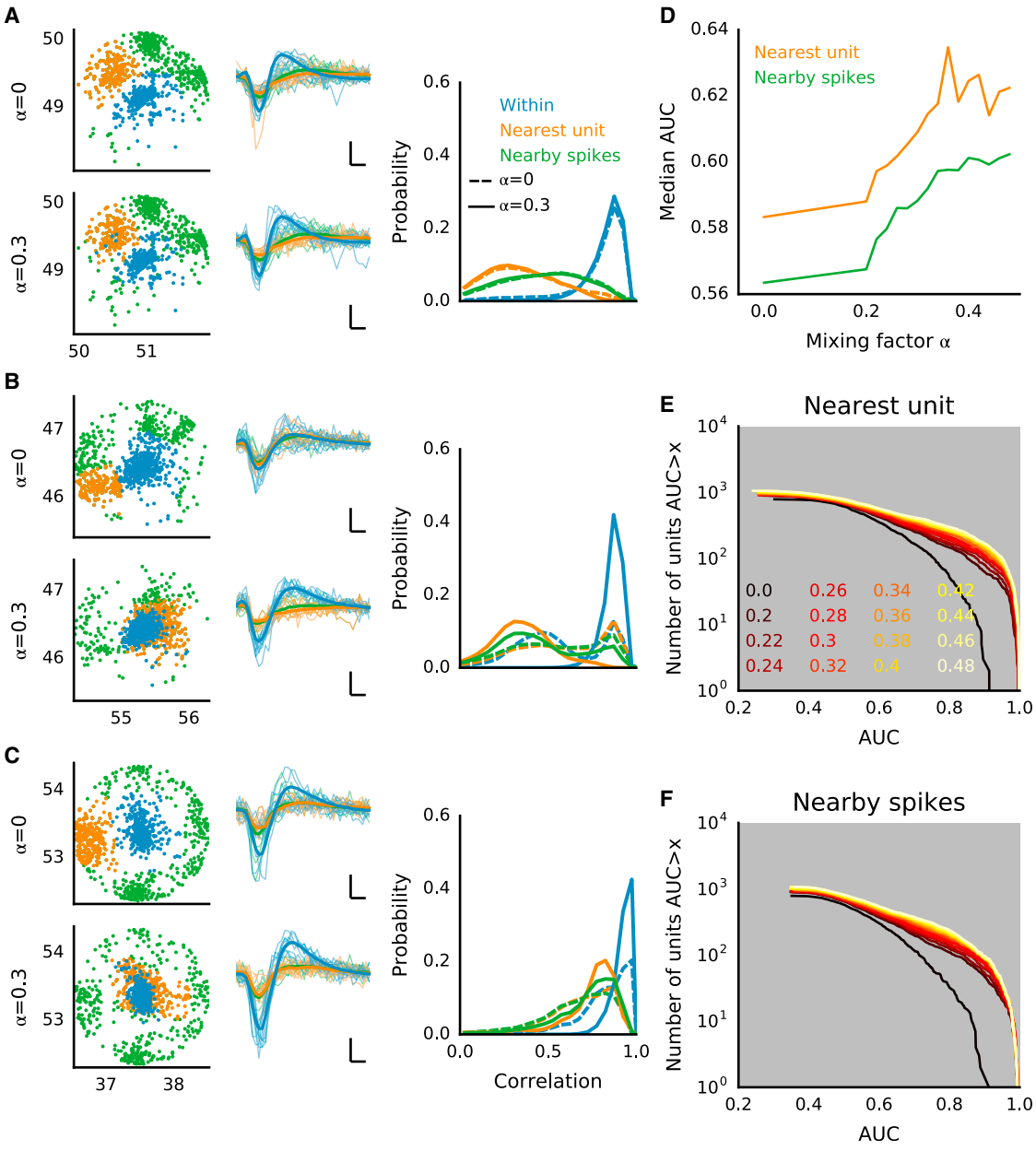


Figure 3. Waveform Correlations Demonstrate Improved Clustering for Combined Event Locations and Waveform Features

All data are from the same experiment as in Figure 1, acquired at 24 kHz.

(A) An example comparing the same unit obtained using spatial clustering alone (with mixing coefficient $\alpha = 0$) with clustering based on combined event locations and waveform features ($\alpha = 0.3$). Shown are event locations (left), example (thin lines) and average (thick lines) waveforms (center; scale bars, 0.2 ms and 100 μV), and normalized distributions of waveform correlations (right; dashed lines, $\alpha = 0$; solid lines, $\alpha = 0.3$). The selected unit is colored in blue (within), the nearest unit in orange, and the remaining events within a radius of 42μm of the target unit location in green (nearby spikes; these also include the spikes of the nearest unit). In this example, spatial clustering is sufficient to isolate the blue unit.

(B and C) Same as (A), but illustrating two units that spatially overlap with their neighbors.

(D) Median AUC for all units, quantifying the overlap between the normalized distributions of waveform correlations for each unit as a function of the mixing coefficient α. The comparison was either done with the spatially closest unit (orange) or with all neighboring spikes (green).

(E) Full distributions of AUC values obtained from comparison with the nearest unit and for different values of α.

(F) Same as (E), but taking all nearby spikes into account.

stimulation of ChR2-expressing RGCs. In addition, 77 units were significantly less active during light stimulation than during ChR2 stimulation, presumably reflecting neurons unresponsive to photoreceptor activation but nevertheless expressing ChR2. The responsiveness of each unit to ChR2 activation was assessed by determining the correlation of an individual unit's

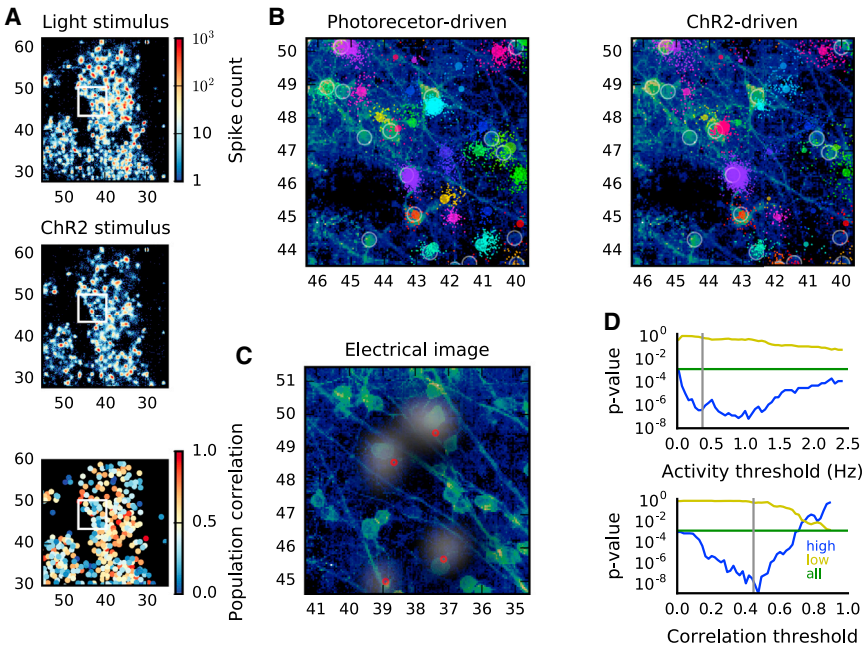


Figure 4. Comparison of Optogenetically Evoked Spikes with Anatomical Imaging

(A) Activity maps obtained during photoreceptor stimulation (top) and ChR2-expressing RGC stimulation under blockade of the glutamatergic pathway from photoreceptors to RGCs (center). The bottom graph shows the correlation of the activity of each unit with the overall ChR2-driven population activity, which quantifies the responsiveness to optogenetic stimulation.

(B) Alignment of neural activity with a confocal image. Individual spikes are shown as dots, colored according to unit membership (note that only a subset of all recorded spikes is shown for clarity). Annotated somata are highlighted by circles and the unit's centroids as colored circles with areas proportional to the spike rate.

(C) A different imaged area with superimposed electrical images of four selected units. Cluster centroids are indicated by red circles.

(D) The distribution of spatial distances between each unit and its closest soma is significantly different from randomness. The one-tailed Kolmogorov-Smirnov test shows incompatibility with the distribution obtained by assuming that somata and units are unrelated ($p = 0.001$, green line). When the units are separated into two sets according to activity level (top) and population correlation (bottom), the effect is strongest for highly active/highly correlated units (blue), whereas weakly active/correlated units are randomly distributed (yellow). The gray line indicates the threshold value for which the two sets have the same number of units. The data in these graphs summarize an imaged area of 0.78 mm^2.

activity with the overall population activity (Figure 4A, bottom). Almost all highly active units during ChR2 stimulation also showed higher correlation, with some exhibiting uncorrelated activity, which we attribute to intrinsic spontaneous activity that could not be blocked. Of all detected units, about 40% had a correlation larger than 40%, close to the expected fraction of Thy1-expressing RGCs.

Next we co-localized the activity with confocal micrographs of labeled neurons (Figure 4B). We analyzed an area of 0.78 mm^2, where 195 somata were manually annotated, and 211 units were detected. An example of the alignment of activity and anatomical image is shown in Figure 4B for activity obtained during photoreceptor stimulation (left) and ChR2 activation (right). All units with significant activity during ChR2 stimulation were closely co-localized with a labeled soma. Similarly, there is a tight co-localization between the neurons and electrical images generated from the raw traces (Figure 4C).

To verify whether labeled somata and localized units where significantly close to each other, we computed the distance to the closest soma for every unit. If units and somata were randomly distributed, the probability of a distance r would be $2\pi n r e^{-\pi n r^2}$, where n is the density of somata (Chandrasekhar, 1943). We compared the distribution of 198 distances to this null model using a one-tailed Kolmogorov-Smirnov test (Figure 4D), confirming that the distances are significantly smaller than predicted by the random model. To account for the effect of spontaneous activity, we applied the test after separating the units into two groups according to their activity level or population correlation, varying the threshold that separates the two sets. The locations of the less active and less correlated units are compatible with a random distribution, whereas the more active and better correlated units are significantly closer to their anatomical counterparts.

Model-Based Validation and Quality Control

As pointed out above, detection was performed with a low threshold to minimize false negatives. Hence some units are expected to contain ambiguities the clustering algorithm cannot fully resolve. For instance, the localization error is typically larger for spikes with small amplitudes (Muthmann et al., 2015); hence, it may not be possible to spatially cluster these events reliably.

To assess the cluster assignments' quality and automatically reject poorly separated units, we followed an approach proposed by Hill et al. (2011). Under the assumption that spike locations and waveform features can be described by a multivariate normal distribution, a comparison of the clusters assignments with those predicted by a Gaussian mixture model provides an estimate of the classification performance. Each unit was investigated in turn, including all of its immediate neighbors, by fitting a six-dimensional Gaussian mixture model with the number of components equal to the number of units (Experimental Procedures). We included four PCA dimensions to ensure that the model best exploits all available waveform features while ensuring reliable convergence. To evaluate the relevance of spatial locations and waveform features for clustering, the model was also fit to each of these features separately.

The model comparison produces a confusion matrix with the estimated number of false positives and negatives for each unit, which is then summarized into a single measure (F-score). Two typical outcomes of this procedure are illustrated in Figures 5A and 5B for relatively crowded areas on the array. Figure 5A shows a unit with a distinct waveform (blue) and four neighbors

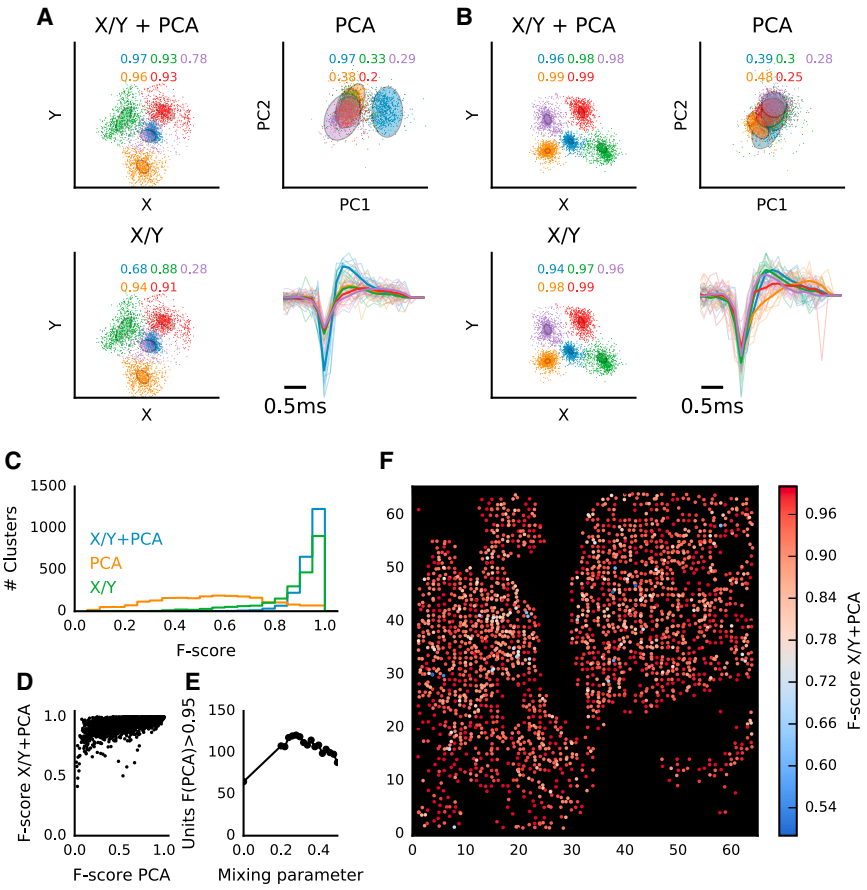


Figure 5. Quantitative Assessment of Sorting Quality with Gaussian Mixture Models

(A) A Gaussian mixture model (GMM) fit to a group of neighboring units. All units within a radius of 42 μm around the unit colored blue were included in the model. The model was then fit to combined spike locations and waveforms (X/Y+PCA), waveforms alone (PCA), or locations alone (X/Y). Spikes are colored to indicate the original cluster assignments. The numbers in each panel are the F-scores for each unit, indicating the average number of false positives and negatives between the two assignments. Examples of spike waveforms and the unit average (thick line) are shown using the same color scheme. In this example, the unit colored in blue is well separated both spatially and by waveform features.

(B) Same as (A), but illustrating a group of units with very similar waveforms, which can only be separated using spike locations.

(C) Histogram of F-scores of all units in one recording, computed as in (A) and (B).

(D) Relationship between F-scores evaluated from waveforms alone and the combined features.

(E) Number of units with an F-score > 0.95, evaluated from waveforms alone for different values of the shape mixing parameter α. The best overlap is obtained for α = 0.28, the value used in the other examples in this paper.

(F) Spatial distribution of F-scores for all units.

within one electrode radius. The blue unit was already well isolated based on waveform features alone (PCA, F-score = 0.97) but not when only spike locations were considered (X/Y, F-score = 0.68). Combining locations and waveforms did not yield further improvement, although it helped to isolate its neighbors based on their spike locations. Figure 5B shows five spatially well separated units with smaller and very similar waveforms. Waveform-based clustering alone gave poor results, but adding spike locations improved it considerably.

Figures 5C–5F summarizes the analysis performed on a 7.6-million spikes dataset. Each of 2,234 units with a spike rate of at least 0.3 Hz took, in turn, the role of the blue unit in Figure 5A, and all units within a radius of 42μm were combined into mixture models. When location and waveform features were used for quality control, 55% of the units (1,230) had an F-score > 0.95 and 15% (334 units) an F-score > 0.99 (Figure 5C; X/Y+PCA). These fractions decreased only slightly when locations were used on their own but substantially for waveforms (PCA) alone. Comparing F-scores for waveforms or combined features shows that adding locations improves fits in most cases, but poor scores for waveforms also result in lower combined scores (Figure 5D). An inspection of the waveform scores for different α values shows an optimum for α = 0.28 (Figure 5E). A spatial overview of these results showed that units with low F-scores are primarily found in crowded areas (Figure 5F).

Functional Assessment of Single-Unit Activity

We recorded RGC responses to full field flashes, allowing us to evaluate whether individual sorted units exhibit the typical On, Off, or On-Off light responses. Figures 6A and 6B show spike locations, spike waveforms, raster plots, and peri-stimulus time histograms (PSTHs) of all units in a small retinal patch, demonstrating excellent separation into fast and slow On, Off, and On-Off responses. Importantly, immediately adjacent neurons generally exhibit different responses, as expected from the mosaic functional organization of RGCs.

The fact that the majority of these units, even those with very small waveforms, exhibit reliable light responses demonstrates that the signal variance is mainly due to physiological causes rather than electrical noise (Muthmann et al., 2015). Units with well defined waveforms are typically also well separated in their PCA projections, whereas small waveforms are mainly clustered based on spatial locations (compare units 1–3 with units 5–7 in Figure 6C). The cluster F-scores (shown above the waveforms in Figure 6B) are lower for units with small waveforms; hence, further analysis for well isolated cells can rely on this measure.

Comparison with Conventional Spike Sorting

Conventional spike sorting relies on differences in spike waveforms. To evaluate how our approach scores in comparison with such methods, we compared our method with the outcome of manually curated spike sorting done on each MEA channel separately. Conventional spike sorting was performed using

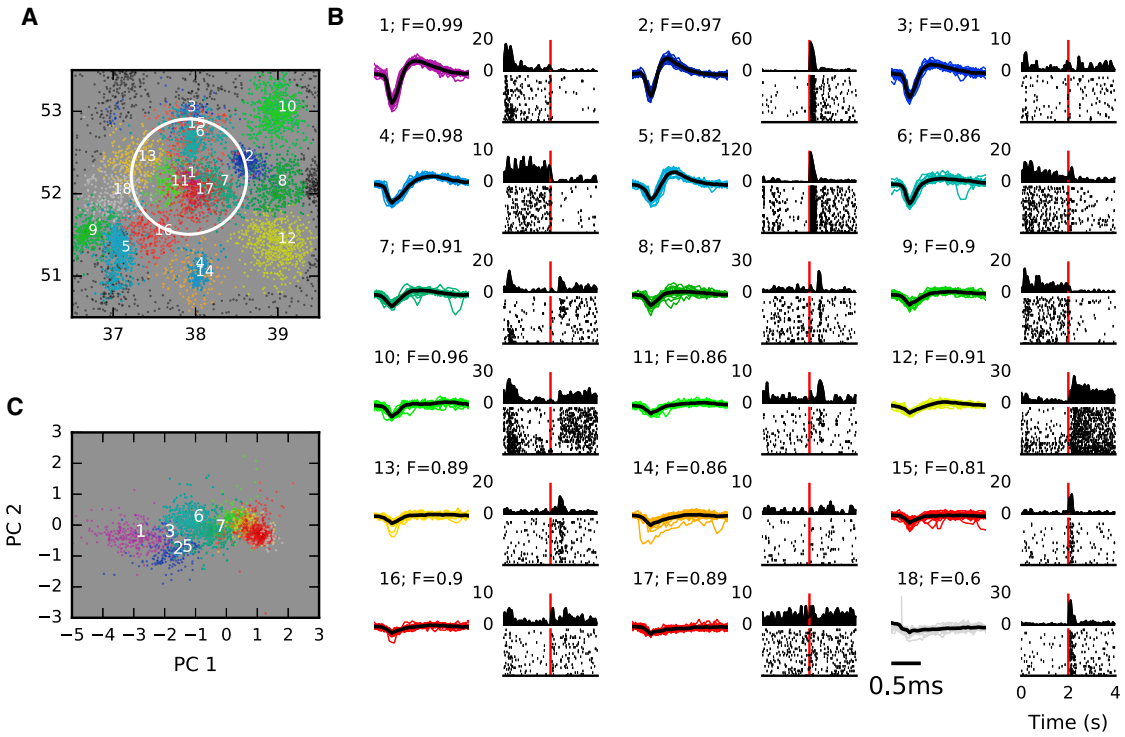


Figure 6. Functional Characterization of Spike-Sorted RGCs

(A) Spatial locations of individual spikes within a small area on the MEA. Only a subset of spikes are shown for clarity. This area contained 18 units, and unit membership is indicated by color. Spikes of units centered outside of the visible area are shown as black dots. Coordinates are in units of electrode distance (42μm).

(B) Overview of the units highlighted in (A) using the same color scheme. Each panel shows example waveforms, the average spike waveform (black line), and the raster and PSTH for full field stimulation (2 s bright, 2 s dark; red lines indicate stimulus offset time). The unit number and cluster F-score are given above the spike waveforms.

(C) Spikes in the circled area in (A), with identical color coding, shown in the space of waveform principal components (PCA space).

Shown are the same data as in Figure 1 with an acquisition rate of 24 kHz.

T-distribution expectation-maximization (E-M) clustering (Shoham et al., 2003) followed by manual inspection and correction (Plexon Offline Sorter).

The data used for this comparison were recorded at 24 kHz with 1,024 electrodes (Figure 1) and included 538 clusters with at least 200 spikes each. For each cluster, we located the most similar sorted unit using spike count cross-correlation following binning (each unit is typically found on multiple electrodes) and obtained the number of spikes in the sorted unit that were not part of the cluster (false negatives) and the number of spikes in the cluster not present in the sorted unit (false positives). As for the mixture model above, we then computed precision, recall, and the F-score for each cluster (Experimental Procedures).

Figure 7A illustrate two common cases we encountered. The first example shows an almost identical assignment for both methods that we found in 96 clusters (18%), with an F-score larger than 0.95 (Figure 7B). Such pairs had very few false positives and negatives (e.g., the pair in Figure 7A, top, had nine false negatives and no false positives of 1,818 spikes).

For many of the remaining clusters, the F-score was dominated by a sizable fraction of false negatives, spikes in the sorted unit that were not included in the corresponding cluster (units with low recall in Figure 7C). An inspection of the spatial locations of these events showed that false negatives were often located far away from the cluster centroid and visually appeared to be part of another unit (Figure 7A, center and bottom, orange events). Figure 7A, center and bottom, illustrates the consequences of erroneous assignment by conventional spike sorting, changing On cells into On-Off cells by merging spikes from other nearby Off cells.

We found that the inclusion of distant spikes happened frequently, with an average distance of false negatives from the cluster centroid typically around 30 μm (Figure 7D). This suggested that they were wrongly included in a sorted unit based on waveform similarities. To see whether these failures are associated with specific waveform features, we compared the F-scores with the average projections of the waveforms along their first principal component (Figure 7E). The PC1 projection provides an indicator of signal quality for each unit (Figure 2B), and, indeed, lower F-scores were observed almost exclusively for low-scoring units. Hence, we conclude that conventional spike sorting only allows reliably isolation of units with strong, very prominent waveform features, whereas smaller, less distinct

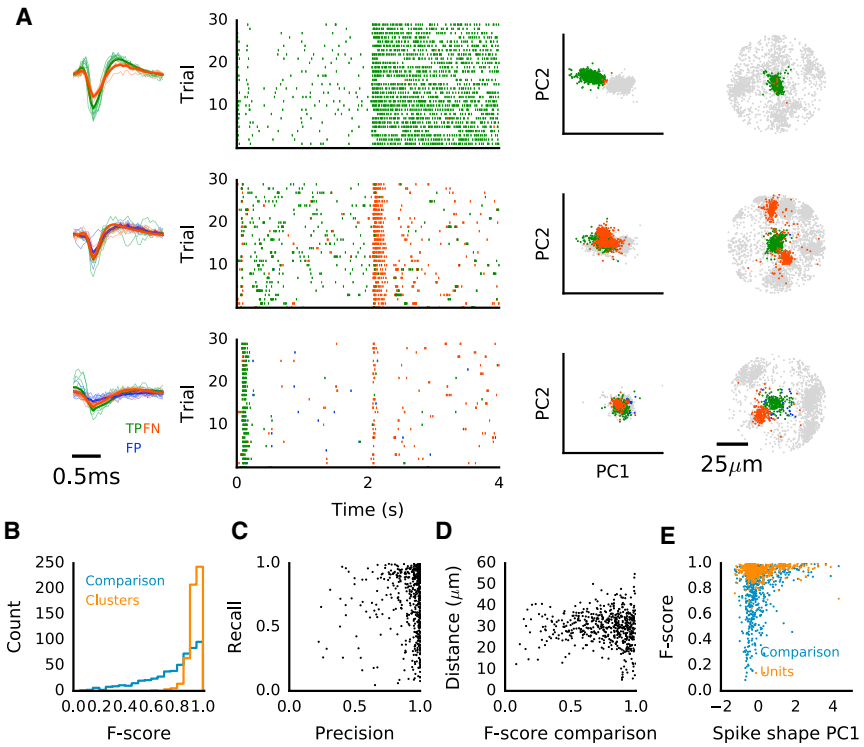


Figure 7. Failure of Conventional Spike Sorting in Isolating Single Units

(A) Examples of three units clustered with our method compared with corresponding units obtained from conventional, spike-shape based sorting. Raster plots show responses to full field flashes (left; 2 s bright, 2 s dark), principal component projections of all spikes found in the area within a radius of 78 μm around the cluster center (center), and all spikes plotted at their locations (right). Spikes colored green were found in both units, those in orange only in the sorted unit, and those in blue only in the clustered unit.

(B) Histograms of F-scores for the comparison (blue) and for mixture model fits for the sorted units (orange).

(C) Precision and recall for the comparison, illustrating that low F-scores are primarily due to spikes missing in the clustered unit (orange events in A).

(D) Average distance of spikes not included in the clustered unit, measured from the cluster centroid.

(E) Comparison of F-scores with the average projection of the waveforms along the first principal component, shown for the comparison of sorting method (blue) and for the mixture model fits of clustered units (orange).

Shown are the same data as in Figure 1.

waveforms cannot be separated reliably on the exclusive basis of their shape.

DISCUSSION

Spike sorting is a critical step in the analysis of extracellular electrophysiological recordings. An erroneous assignment of spikes can have severe consequences for the interpretation of neural activity, which has motivated the development of joint models of spike waveforms and neural activity to avoid spurious or biased correlation estimates (Ventura and Gerkin, 2012). In high-density recordings, increasingly used both for in vitro and in vivo studies, assigning spikes to single units becomes exponentially complex as a function of the number of events; hence, it requires approximate solutions. Moreover, the sheer size of the data prevents detailed manual inspection and quality control.

Here we solve this task by creating an efficient, low-dimensional data representation, based on spatial spike locations and the most prominent waveform features, that can be clustered efficiently. We found that clustering in four dimensions, with two dimensions representing waveform features, was sufficient to achieve high performance, which we attribute to the fact that the signals reliably measured with a dense MEA mainly originate from strong currents at the AIS of each neuron, with limited variability between neurons. This enables estimating their spatial origin but limits variability to support shape-based spike sorting. Comparison of optogenetically evoked spikes with anatomical images indicates that detected spikes typically cluster near the AIS and that localization alone is sufficiently precise to reliably isolate some neurons even without using additional waveform features.

Out method could be used with arrays and probes where an event location estimate can be reliably obtained. The dimensionality of the clustering step can then be adjusted to exploit higher waveform variability. The complexity of the clustering algorithm scales quadratic with the number of spikes, and the highly optimized version used here has a better performance when prominent spatial clustering is present. We developed a parallelized implementation that allows sorting of millions of spikes in minutes (ten million spikes take about 8 min on a 12-core 2.6-GHz Xeon workstation). Together with a method for quality control, this makes it possible to perform parameter sweeps to identify the optimal parameters of the clustering algorithm. Clustering is followed by an automated assessment of clustering quality, allowing the automated rejection of poorly isolated units and manual inspection of borderline cases. We also provide a visualization tool where further annotation can be performed.

The complete workflow consists of event detection, spatial localization, clustering, quality control, and, finally, optional manual inspection. The former two currently constitute the main bottleneck. Detection takes about four times real time and scales linearly with recording duration. The complexity of the spike localization scales linearly with the number of detected events and runs roughly in real time for recordings with normal spike rates. For both methods, parallelized implementations are under development.

Our work with high-density recordings has revealed significant limitations of purely shape-based spike sorting for MEA recordings. It is virtually impossible to evaluate how many units are

represented in a single electrode signal. If the electrode is positioned close to a neuronal cluster, one or two units with strong signals usually have sufficiently distinct waveforms to be separable. However, comparison with spike locations showed that weaker signals arising from more remote cells are generally not distinguishable based on shape alone. We frequently found cases where spikes of neurons with entirely different physiological signatures were mixed by shape-based sorting, a problem that cannot be avoided even by careful manual inspection. Our method, on the other hand, handles such situations much better because spatial location estimates are sufficiently precise to disambiguate borderline cases. Thus, a main factor affecting sorting performance is the noise and bias in spatial localization, both depending on signal quality (Muthmann et al., 2015).

A different strategy, outlined by Marre et al. (2012), is to estimate spatio-temporal templates that are then used to identify spikes from each neuron (Dragas et al., 2015). This shifts the computational burden from spatial interpolation and source localization in our method to the deconvolution of spikes from raw data. We found that adding shape criteria at the detection stage could lead to false negatives, suggesting that templates can only be reliably estimated for neurons with sufficiently high firing rates. A third method, recently developed by Rossant et al. (2016) for high-density in vivo probes, reduces complexity by masking irrelevant parts of the data based on geometric constraints before fitting a mixture model and clustering the data. This avoids an early discarding of potentially useful information, which our method does by using signal interpolation and Marre et al. (2012) did by creating templates. On the other hand, although potentially more precise, this method is computationally more demanding and, hence, more suitable for data from hundreds of channels.

EXPERIMENTAL PROCEDURES

Electrophysiology

Experimental procedures were approved and carried out in accordance with the guidance provided by the United Kingdom Home Office, Animals (Scientific Procedures) Act 1986 (Retinal Recordings), by the institutional Istituto Italiano di Tecnologia (IIT) Ethic Committee, and by the Italian Ministry of Health and Animal Care (Authorization ID 227, Prot. 4127 March 25, 2008) (neural cultures; Panas et al., 2015).

Experiments on the retina were performed on adult wild-type mice (C57BL/6, aged post-natal days [P] 27–39) or on B6.Cg-Tg(Thy1-COP4/EYFP)9Gfng/J mice (Thy1-ChR2-YFP; The Jackson Laboratory; RRID:IMSR_JAX:007615) aged P69–96. Recordings from the RGC layer were performed using the BioCam4096 platform with active pixel sensor (APS) MEA chips (type BioChip 4096S, 3Brain), providing 4,096 square microelectrodes (21μm × 21μm) on an active area of 2.67 mm × 2.67 mm, aligned in a square grid with 42 μm spacing. The platform records at a sampling rate of about 7 kHz/electrode when measuring from the full 64 × 64 MEA, but sampling increases to 24 kHz when recording from 1,024 electrodes. Raw data were visualized and recorded with the 3Brain proprietary BrainWave software. Activity was recorded at 12-bit resolution per electrode, low pass-filtered at 5 kHz with the on-chip filter, and high pass-filtered by setting the digital high pass filter of the platform at 0.1 Hz.

Mice were killed by cervical dislocation and enucleated prior to retinal isolation. The isolated retina was placed, RGC layer facing down, onto the MEA (for details, see Maccione et al., 2014). The retina was continuously perfused with artificial cerebrospinal fluid (maintained at 32°C) containing the following: 118 mM NaCl, 25 mM $NaHCO_3$, 1 mM NaH_2PO_4, 3 mM KCl, 1 mM $MgCl_2$, 2 mM $CaCl_2$, and 10 mM glucose, equilibrated with 95% O_2 and 5% CO_2. All preparations were performed under dim red light, and the room was maintained in darkness throughout the experiment.

Visual and Optogenetic Stimulation

We used a custom-made projection system to deliver visual stimuli to the retina (for details, see Portelli et al., 2016). Photoreceptor-driven responses were acquired at a maximum irradiance of 4μW/cm^2 (neutral density (ND) filter 4.5), low enough to avoid eliciting ChR2-driven responses in the ChR2 retinas. ChR2-driven responses were elicited using the broad RGB spectrum of the projector with a maximum irradiance of 0.87 mW/cm^2 (ND 2.2) following blockade of photoreceptor-driven responses by increasing $[MgCl_2]_{out}$ to 2.5 mM and by decreasing $[CaCl_2]_{out}$ to 0.5 mM (to reduce synaptic transmission) and in the presence of 20μm DNQX and 20μm L-AP4 (Tocris Bioscience) to block glutamatergic neurotransmission in the photoreceptor-bipolar cell-RGC pathway. Responses to repetitive (30×) full field stimuli (0.5 Hz) were analyzed as shown in Figures 6 and 7.

Spike Detection, Localization, and Selection

The procedures for spike detection and localization are described in detail elsewhere (Muthmann et al., 2015). Weighted interpolated signals were generated using two spatial templates to capture both spikes originating close to or between electrodes. The running baseline and noise estimate were computed as signal percentiles, and putative spikes were detected as threshold crossings. This procedure ensures detection of temporally overlapping spikes as long as they leave a distinct spatial footprint. Next, source locations were estimated for each event by considering the spatial signal spread over neighboring electrodes. The signals were baseline-subtracted and inverted, and then the median signal was subtracted to minimize bias because of noise. The signal was clipped to positive values, and the center of mass was determined. To filter out noise and poorly detected neurons in recordings at 7 kHz, we developed an automated post hoc event rejection. To this end, noise events were sampled from areas on the MEA where no activity was recorded, such as at incisions or uncovered areas (identifiable by low spike counts). Up to 1,000 of such events and up to 1,000 events with large amplitudes were used to train a support vector machine with radial basis functions. This model was then used to classify events as true spikes or noise (Figure S1).

Spike Clustering

Data points were clustered using an implementation of the mean shift algorithm (Comaniciu and Meer, 2002), available in the scikit-learn open source machine learning library (Pedregosa et al., 2011). Importantly, this algorithm does not require prior knowledge of the desired number of clusters. It depends on a single parameter, the bandwidth h, which determines the expected cluster size, which, in turn, can be estimated from a typical spatial cluster size in an activity plot (Figure 1B) and was here set to 12.6 μm (the average width of clusters). The clustering process was run on a four-dimensional space consisting of two dimensions, indicating the location of each event on the chip, x and y, and two dimensions representing the first two principal components of the event's waveform. The latter were multiplied by an additional dimensional constant α that tuned the relative importance of the waveform components compared with the spatial coordinates. To parallelize this algorithm, we exploited the fact that all points follow a local density gradient until they converge to a local maximum, the center of a cluster. Because every data point does so independently of the others, this process is run in parallel, which improved performance roughly proportionally to the number of available central processing units (CPUs). The relevant code has been merged into the scikit-learn Python library.

Quality Metric

Following Hill et al. (2011), we fitted a multivariate Gaussian mixture model to a set of N clusters and then estimated their overlap using posterior probabilities to obtain the probability of incorrect assignments under the assumption of a Gaussian cluster shape. The model is fit in six dimensions, with the two spatial coordinates and the projections of the spike waveform along the first four principal components. For each cluster, we assume that only spikes in nearby clusters interfere with the sorting. Therefore, all clusters or spikes within a radius of 42 μm (electrode pitch) are included in the model. To obtain meaningful fits for sets of clusters with very disparate number of spikes, a Gaussian is fit

to each cluster individually before combining them into a mixture model. The assignment quality is evaluated as follows. Let the probability of spike s in cluster c be $P(C=c \mid S=s)$: the estimated fraction of spikes in cluster k that could belong to cluster i is given by $f^p(k,i) = (1/N_k)\sum_{s\in k} P(C=i \mid S=s)$; by generalizing to all other clusters, we obtained the number of false positives in k:

$$f_k^p = \sum_{i\neq k} f^p(k;i) = \sum_{i\neq k}\sum_{s\in k} P(C=i \mid S=s).$$

Correspondingly, the number of false negatives, the fraction of spikes in cluster c that was expected to be assigned to other (i.e., wrong) clusters, was obtained as $f_k^n = \sum_{i\neq k}\sum_{s\in i} P(C=k \mid S=s)$.

The probabilities $P(C=c, S=s)$ were given by mixture model. To obtain a single quality measure, we compute (P_k) and recall (R_k):

$$P_k = \frac{n_k - f_k^p}{n_k}$$

$$R_k = \frac{n_k - f_k^p}{n_k - f_k^p + f_k^n}.$$

The harmonic mean of these quantifies yields the F-score:

$$F_k = 2\frac{P_k R_k}{P_k + R_k}.$$

Confocal Imaging and Image Analysis

To achieve a precise alignment of RGCs with recording electrodes, the retina had to be imaged on a chip with photoreceptors facing upward. The retina was fixed with 4% paraformaldehyde (in 0.1 M PBS and 200 mM sucrose) on the MEA chip for 1 hr after recording. We have determined that tissue shrinkage, which may interfere with activity alignment, is negligible for this protocol. The retina was rinsed several times with 0.1 M PBS, embedded with Vectashield (Vector Laboratories), and sealed with a coverslip (Menzel Glaeser). Imaging was performed with a Leica SP5 confocal upright microscope supplied with a 25×/0.95 numerical aperture (NA) working distance (WD) 2.5 mm water immersion objective for optimal signal collection focusing on areas encompassing 8 × 8 electrodes (300 × 300μm field of view). In each field, images (2,048 × 2,048 pixels) were acquired in z stacks in tissue thickness of 60–100 μm (optical slicing yielding 30–50 image planes). A lateral resolution of 200 nm per pixel, just above the diffraction limit, and optical slicing of 550 nm provided an adequate trade-off between sufficient image details and acquisition time, minimizing the risk of photo damage. For image restoration, the Richardson-Lucy method (Lucy, 1974; Richardson, 1972) was used. In addition to the fluorescence signals in specific fields, large-field images, including images of the MEA, were acquired to enable co-localization of images with RGC spiking activity.

In one Thy1 YFP-ChR2 retina, RGC somata were manually annotated in selected subfields where activity was recorded, and the confocal images of the RGC layer were spatially aligned with the estimated locations of detected events. To this end, the active area of one electrode was determined, and the remaining electrode locations were computed, generating a regular grid using 42 μm electrode spacing. The images and soma locations were then transformed into array coordinates, and spike locations were overlaid with the retinal image.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and one figure and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.02.038.

AUTHOR CONTRIBUTIONS

Conceptualization, G.H., M.S., E.S., and M.H.H.; Methodology, G.H, M.S., J.O.M., S.P., I.E.K., S.U., E.S., and M.H.H.; Software, M.S., S.P., J.O.M., C.J.R., A.P.E., and M.H.H.; Formal Analysis, M.S., S.U., and M.H.H.; Investigation, G.H., M.S., S.P., I.E.K., and S.U.; Resources, A.M. and L.B.; Data Curation, G.H., A.M., L.B., and E.S.; Writing – Original Draft, M.H.H. with input from co-authors; Supervision, L.B., V.M., D.S., F.C.Z, E.S., and M.H.H.; Funding Acquisition, L.B., V.M., D.S., F.C.Z., E.S., and M.H.H.

ACKNOWLEDGMENTS

We thank Fernando Rozenblit, Vidhyasankar Krishnamoorthy, and Amos Storkey for valuable input. This work was supported by the 7th Framework Program for Research of the European Commission (grant agreement 600847: RENVISION, project of the Future and Emerging Technologies [FET] program Neuro-bio-inspired Systems FET-Proactive Initiative) and the Wellcome Trust (grant number 096975/Z/11/Z). M.S. was supported by the EuroSPIN Erasmus Mundus Program and the EPSRC Doctoral Training Centre in Neuroinformatics (EP/F500385/1 and BB/F529254/1). J.O.M. was supported by the EuroSPIN Erasmus Mundus Program and NCBS/TIFR.

Received: June 20, 2016
Revised: November 21, 2016
Accepted: February 13, 2017
Published: March 7, 2017

REFERENCES

Ballini, M., Muller, J., Livi, P., Chen, Y., Frey, U., Stettler, A., Shadmani, A., Viswam, V., Jones, I.L., Jackel, D., et al. (2014). A 1024-channel CMOS microelectrode array with 26,400 electrodes for recording and stimulation of electrogenic cells in vitro. IEEE J. Solid-State Circuits *49*, 2705–2719.

Berdondini, L., van der Wal, P.D., Guenat, O., de Rooij, N.F., Koudelka-Hep, M., Seitz, P., Kaufmann, R., Metzler, P., Blanc, N., and Rohr, S. (2005). High-density electrode array for imaging in vitro electrophysiological activity. Biosens. Bioelectron. *21*, 167–174.

Chandrasekhar, S. (1943). Stochastic problems in physics and astronomy. Rev. Mod. Phys. *15*, 1.

Comaniciu, D., and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. *24*, 603–619.

Dragas, J., Jackel, D., Hierlemann, A., and Franke, F. (2015). Complexity Optimisation and High-Throughput Low-Latency Hardware Implementation of a Multi-Electrode Spike-Sorting Algorithm. IEEE Trans. Neural Syst. Rehabil. Eng. *23*, 149–158.

Eversmann, B., Jenkner, M., Hofmann, F., Paulus, C., Brederlow, R., Holzapfl, B., Fromherz, P., Merz, M., Brenner, M., Schreiter, M., et al. (2003). A 128 128 CMOS Biosensor Array for Extracellular Recording of Neural Activity. IEEE J. Solid-State Circuits *38*, 2306–2317.

Frey, U., Sedivy, J., Heer, F., Pedron, R., Ballini, M., Mueller, J., Bakkum, D., Hafizovic, S., Faraci, F.D., Greve, F., et al. (2010). Switch-matrix-based high-density microelectrode array in CMOS technology. IEEE J. Solid-State Circuits *45*, 467–482.

Harris, K.D., Henze, D.A., Csicsvari, J., Hirase, H., and Buzsáki, G. (2000). Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J. Neurophysiol. *84*, 401–414.

Hill, D.N., Mehta, S.B., and Kleinfeld, D. (2011). Quality metrics to accompany spike sorting of extracellular signals. J. Neurosci. *31*, 8699–8705.

Hutzler, M., Lambacher, A., Eversmann, B., Jenkner, M., Thewes, R., and Fromherz, P. (2006). High-resolution multitransistor array recording of electrical field potentials in cultured brain slices. J. Neurophysiol. *96*, 1638–1645.

Lewicki, M.S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. Network *9*, R53–R78.

Lindén, H., Tetzlaff, T., Potjans, T.C., Pettersen, K.H., Grün, S., Diesmann, M., and Einevoll, G.T. (2011). Modeling the spatial reach of the LFP. Neuron *72*, 859–872.

Lucy, L.B. (1974). An iterative technique for the rectification of observed distributions. Astron. J. *79*, 745.

Maccione, A., Hennig, M.H., Gandolfo, M., Muthmann, O., van Coppenhagen, J., Eglen, S.J., Berdondini, L., and Sernagor, E. (2014). Following the ontogeny of retinal waves: pan-retinal recordings of population dynamics in the neonatal mouse. J. Physiol. *592*, 1545–1563.

Marre, O., Amodei, D., Deshmukh, N., Sadeghi, K., Soo, F., Holy, T.E., and Berry, M.J., 2nd. (2012). Mapping a complete neural population in the retina. J. Neurosci. *32*, 14859–14873.

Mechler, F., Victor, J.D., Ohiorhenuan, I., Schmid, A.M., and Hu, Q. (2011). Three-dimensional localization of neurons in cortical tetrode recordings. J. Neurophysiol. *106*, 828–848.

Müller, J., Ballini, M., Livi, P., Chen, Y., Radivojevic, M., Shadmani, A., Viswam, V., Jones, I.L., Fiscella, M., Diggelmann, R., et al. (2015). High-resolution CMOS MEA platform to study neurons at subcellular, cellular, and network levels. Lab Chip *15*, 2767–2780.

Muthmann, J.-O., Amin, H., Sernagor, E., Maccione, A., Panas, D., Berdondini, L., Bhalla, U.S., and Hennig, M.H. (2015). Spike Detection for Large Neural Populations Using High Density Multielectrode Arrays. Front. Neuroinform. *9*, 28.

Ness, T.V., Chintaluri, C., Potworowski, J., Łęski, S., Głąbska, H., Wójcik, D.K., and Einevoll, G.T. (2015). Modelling and Analysis of Electrical Potentials Recorded in Microelectrode Arrays (MEAs). Neuroinformatics *13*, 403–426.

Obien, M.E.J., Deligkaris, K., Bullmann, T., Bakkum, D.J., and Frey, U. (2015). Revealing neuronal function through microelectrode array recordings. Front. Neurosci. *8*, 423.

Panas, D., Amin, H., Maccione, A., Muthmann, O., van Rossum, M., Berdondini, L., and Hennig, M.H. (2015). Sloppiness in spontaneously active neuronal networks. J. Neurosci. *35*, 8480–8492.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Pettersen, K.H., and Einevoll, G.T. (2008). Amplitude variability and extracellular low-pass filtering of neuronal spikes. Biophys. J. *94*, 784–802.

Portelli, G., Barrett, J.M., Hilgen, G., Masquelier, T., Maccione, A., Di Marco, S., Berdondini, L., Kornprobst, P., and Sernagor, E. (2016). Rank order coding: A retinal information decoding strategy revealed by large-scale multielectrode array retinal recordings. eNeuro *3*.

Prentice, J.S., Homann, J., Simmons, K.D., Tkačik, G., Balasubramanian, V., and Nelson, P.C. (2011). Fast, scalable, Bayesian spike identification for multi-electrode arrays. PLoS ONE *6*, e19884.

Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput. *16*, 1661–1687.

Raymond, I.D., Vila, A., Huynh, U.-C.N., and Brecha, N.C. (2008). Cyan fluorescent protein expression in ganglion and amacrine cells in a thy1-CFP transgenic mouse retina. Mol. Vis. *14*, 1559–1574.

Rey, H.G., Pedreira, C., and Quian Quiroga, R. (2015). Past, present and future of spike sorting techniques. Brain Res. Bull. *119* (*Pt B*), 106–117.

Richardson, W.H. (1972). Bayesian-based iterative method of image restoration. J. Opt. Soc. Am. *62*, 55–59.

Rossant, C., Kadir, S.N., Goodman, D.F.M., Schulman, J., Hunter, M.L.D., Saleem, A.B., Grosmark, A., Belluscio, M., Denfield, G.H., Ecker, A.S., et al. (2016). Spike sorting for large, dense electrode arrays. Nat. Neurosci. *19*, 634–641.

Shoham, S., Fellows, M.R., and Normann, R.A. (2003). Robust, automatic spike sorting using mixtures of multivariate t-distributions. J. Neurosci. Methods *127*, 111–122.

Ventura, V., and Gerkin, R.C. (2012). Accurately estimating neuronal correlation requires a new spike-sorting paradigm. Proc. Natl. Acad. Sci. USA *109*, 7230–7235.