HHS Public Access

Author manuscript

Int J Neural Syst. Author manuscript; available in PMC 2019 October 01.

Published in final edited form as:

Int J Neural Syst. 2018 October ; 28(8): 1850008. doi:10.1142/S0129065718500089.

Identification of Retinal Ganglion Cell Firing Patterns Using Clustering Analysis Supplied with Failure Diagnosis

Alireza Ghahari[*,‡], **Sumit R. Kumar**[§], and **Tudor C. Badea**[†,‡]

Retinal Circuit Development and Genetics Unit, National Eye Institute, 6 Center Drive, Bethesda, MD 20892, USA

Abstract

An important goal in visual neuroscience is to understand how neuronal population coding in vertebrate retina mediates the broad range of visual functions. Microelectrode arrays interface on isolated retina registers a collective measure of the spiking dynamics of retinal ganglion cells (RGCs) by probing them simultaneously and in large numbers. The recorded data stream is then processed to identify spike trains of individual RGCs by efficient and scalable spike detection and sorting routines. Most spike sorting software packages, available either commercially or as freeware, combine automated steps with judgment calls by the investigator to verify the quality of sorted spikes. This work focused on sorting spikes of RGCs into clusters using an integrated analytical platform for the data recorded during visual stimulation of wild-type mice retinas with whole field stimuli. After spike train detection, we projected each spike onto two feature spaces: a parametric space and a principal components space. We then applied clustering algorithms to sort spikes into separate clusters. To eliminate the need for human intervention, the initial clustering results were submitted to diagnostic tests that evaluated the results to detect the sources of failure in cluster assignment. This failure diagnosis formed a decision logic for diagnosable electrodes to enhance the clustering quality iteratively through rerunning the clustering algorithms. The new clustering results showed that the spike sorting accuracy was improved. Subsequently, the number of active RGCs during each whole field stimulation was found, and the light responsiveness of each RGC was identified. Our approach led to error-resilient spike sorting in both feature extraction methods; however, using parametric features led to less erroneous spike sorting compared to principal components, particularly for low signal-to-noise ratios. As our approach is reliable for retinal signal processing in response to simple visual stimuli, it could be applied to the evaluation of disrupted physiological signaling in retinal neurodegenerative diseases.

Keywords

Microelectrode array; spike sorting; parametric features; clustering routines; failure diagnosis

[‡]Corresponding authors.

[*]alireza.ghahari@uconn.edu

[†]badeatc@mail.nih.gov

[§]Current address: Department of Internal Medicine, Yale University School of Medicine, New Haven, CT 06520, USA.

1. Introduction

Retinal ganglion cells (RGCs) carry visual signals from the eye to the brain, and can be subdivided into 20–30 cell types based on neuroanatomical, physiological, molecular and functional criteria.[1] Information contained in the visual stimulus is detected by photoreceptors and decomposed into multiple information streams by the retinal circuitry.[2–5] Aspects of the visual scene, such as luminance, local contrast, color and movement, are then passed on to specific RGC types that relay the information to distinct brain nuclei. Given the topographic organization of the visual scene, some aspects of the visual stimulus can be reported over the entire retinal surface to construct a coherent representation of the surrounding world. Thus, it is suitable to record visual stimulus responses from the entire RGC population tiling the retinal surface to understand how our nervous system encodes the visual scene.

The three most promising techniques capable of simultaneous recording of potential activity from large neuronal ensembles are microelectrode arrays (MEAs), intracellular calcium (Ca^{2+}) indicators and membrane voltage indicators. MEA recordings yield high spatial and temporal resolution for measuring extracellular action potentials.[5–7] However, preparations resulting in a good transmission of the action potential from the cell to the electrode are not trivial. A major improvement in the field has come from the use of very dense arrays, or the introduction of CMOS-based devices that promise a higher resolution, nearing complete reporting of the RGCs in the ensemble.[8–10] Ca^{2+} indicators are chemical or genetic probes whose fluorescence intensity or spectrum is modulated upon Ca^{2+} concentration changes and are amenable to full-field imaging of the retinal preparation.[3,11,12] However, these changes are more prominent when the large depolarization waves associated with action potentials open voltage-dependent Ca^{2+} channels. Although these techniques are more successful at revealing activity of all neurons in the imaged preparation, they have poor temporal resolution. This is due to both the physicochemical limits of the Ca^{2+}-dependent fluorescence change and the slow dynamics of Ca^{2+} clearance from the neuron after the signaling or action potential has subsided. Membrane voltage indicators are the most recent addition to the toolset, and have better temporal resolution than the Ca^{2+} indicators, however yield a far lower signal-to-noise (SNR) ratio.[13]

Spike sorting algorithms serve to separate extracellular spike recordings from individual cells into distinct neuronal units. After detecting spikes from the measurements and representing them by their salient features, a clustering method groups spikes into different clusters. Then, a template waveform can be defined to orient the rest of spike sorting to semi-supervised or unsupervised threads. Several spike inference approaches have been applied to solve this problem, including template matching,[14–17] matched filter design[18,19] and Bayesian techniques.[20–23] Given that these approaches are limited in scope by their modeling assumption, and pose heavy computation challenges for large datasets, there is a need to explore alternative solutions. As assignment of the spikes to clusters is typically subject to false-positive (FP; contamination of clusters by spikes belonging to other cells) and false-negative (FN; artifactual splitting of clusters of spikes that belong to the same cell) errors, it is necessary to introduce some metrics that quantify these errors. It is then possible to inform methods to reduce these errors by refining the quality of initial clusters. In the

literature, this step is simply known as splitting and merging, either manually[15,24,25] or systematically.[26–28] Apart from this error analysis, an ideal clustering technique should be robust against modes of firing, namely, regular spiking versus bursting,[29–31] multispike trains[7,21,32,33] and local field potentials.[34,35]

Classical cluster analysis techniques do not easily generalize to other unsupervised data partitioning tasks. Although they perform well on the specific task for which they were originally designed, their performance degrades when applied to other clustering tasks at scale, requiring huge data flow and intense data mining. If spike sorting is performed offline by clustering procedures, quality is usually improved by user's verification of the clustering results. This typically includes visual inspection of raw data traces for sortable neuronal units, and/or generated clusters, to confirm whether each cluster contains exclusively the sorted spikes of individual units. To reduce this burden, thus facilitating automation, we considered iterative, exploratory computation routines to improve the quality of initial clusters in an integrated platform; that is, the initial representation of clusters was evaluated through a set of diagnostic tests, whose outputs were used to minimize the sum-of-squared-error (SSE) from the initial clustering for a number of clusters. This combination of cluster analysis and failure diagnosis is the core of our approach to spike sorting: clustering supplied with diagnosis (ClusDiag). Our approach provides a new spike inference methodology for visual neuroscience, and herein we report the first illustration of its performance.

The dataset employed here was acquired from an MEA during visual stimulation of wild-type mice retinas with whole field square wave stimuli. This study has three objectives: (i) sorting spikes recorded by individual electrodes into different clusters by using two feature extraction techniques; (ii) finding the number of active RGCs from the number of final clusters; and (iii) classifying those cells into three classical light-sensitive types. To meet the first objective, spike sorting was achieved by applying three analytical layers within ClusDiag to the data for each electrode: representation, evaluation and optimization. For representation, in the data preparation step, we first filtered baseline fluctuations caused by local field potentials out of the data and next enhanced the signal-to-noise ratio using the eigendecomposition technique. Second, an adaptive threshold method identified spike waveforms from the filtered data. Third, we inserted two feature extraction methods by projecting each detected spike waveform onto two feature spaces: a parametric space and a principal components space. Fourth, spikes were sorted into clusters based on leader–follower and iterative minimum-squared-error (MSE) clustering algorithms. Subsequently, a template waveform for each cluster per electrode was defined. To evaluate the initial clustering results, we performed three diagnostic tests on each clusters template and its corresponding spikes, detected the FP and FN errors and tabulated them into a diagnostic dependency matrix (D-matrix). Depending on the number of clusters, failures and spikes, a logical decision indicated whether a reclustering was required. In case of reclustering for electrodes containing more than one cluster, a conjugate gradient method[36] was used to iteratively mitigate the FP and FN errors. The updated clusters were derived by a second run of the MSE algorithm. Subsequently, a second iteration of reliability tests was run to see the extent of reduction of the errors after reclustering. The new D-matrix was examined next to see how well the errors were resolved and to discard those clusters or electrodes exhibiting

least error improvement from the rest of analysis. The remaining electrodes were checked by the third iteration of reliability tests that aimed to reduce the effect of spike events causing FP and FN failures. Finally, we quantified all FP and FN errors after complete execution of ClusDiag. We then turned to the second objective of finding the actual number of active RGCs based on the cluster identified in each electrode. To that end, we detected spike train coincidences between clusters of nearby electrodes, thus revealing RGCs whose spike trains had been recorded by more than one neighboring electrode. This allowed us to report the number of detected cells alongside the exact temporal features of each spike per cell per electrode. Lastly, to classify the sensitivity of the detected RGCs in response to the visual stimuli, we calculated the post-stimulus time histogram (PSTH) for spikes per cluster and defined a ratio index that indicated the light responsiveness of a cell.

2. Methods

This study analyzed electrophysiological responses of wild-type mouse RGCs to whole field stimuli. All mouse handling procedures during this study were approved by the Animal Care and Use Committee of the National Eye Institute (protocol NEI–640). The *in vitro* retina recording was performed with a standard MEA from Multi Channel Systems (MCS GmbH, Germany). This MEA allowed simultaneous light stimulation and recording of 60 electrodes. The light stimulation and data acquisition system are shown in Fig. 1. Light stimulation consisted of fullfield black and white stimuli alternating every 2 s, for a total of 10 iterations. Typically, multiple pieces of retina were derived from each mouse, and recorded successively on the same day. For this study, we used datasets from five such retina preparations (labeled WF_1–WF_5). We illustrate our approach over a 5×5 electrodes region of interest (ROI) centered on electrode E_{54}.

For algorithmic data analysis suiting spike sorting, we considered an approach that drew on iterative, exploratory computation routines in an integrated platform. The approach evolved, altering specific parameters within itself, to become more consistent and robust over time through three analytical layers: representation, evaluation and optimization. First, the purpose of representation is to transform the time series data into a feature vector space that mines the data to learn the salient structure of the high-dimensional data space. Classification and regression techniques (such as logistic regression, decision trees and naive Bayes classifier), cluster analysis methods and dimensionality reduction techniques [such as principal component analysis (PCA) and partial least squares] have been commonly investigated. Second, an evaluation layer leverages diagnostic inference about the quality of the representation in place (e.g. deviation of the results from the expected values). This evaluation underpins rigorous improvement of the former presentation and is integral to reducing manual intervention. Third, continuous optimization techniques, such as gradient descent and quadratic programming, are versatile elements that provide iterative optimization of a criterion function describing the error incurred from the previous representation. Parameters within the representation are tuned iteratively to reinforce a more error-resilient representation of data to follow. It is noteworthy that choosing efficient, scalable algorithms and evaluations that fit together into this platform allows to continually improve the performance of the platform over its lifecycle.

The data representation comprised four stages per electrode: (i) data preparation, (ii) spike detection, (iii) feature extraction and (iv) cluster analysis. After data enhancement, the spikes were detected using an adaptive threshold technique.[37] Each detected spike was projected onto two different feature spaces: a parametric space (f_{spike}) and a principal components space. The rest of the analysis remained the same for both feature representations. The cluster analysis involved the sequential application of two clustering techniques: the leader-follower algorithm and the iterative MSE clustering.[38] Spikes from each electrode were grouped into clusters, and a template waveform was defined to represent their collective activity. To evaluate the clusters, a set of three diagnostic tests served to detect the FP and FN instances over the first run of clustering. The reclustering decision logic detected the faulty electrodes and, for optimization, called up the conjugate gradient method[35] to minimize the sum-of-squared-error from the first run of clustering. The clustering routines were then run for only the faulty electrodes, and their new recovered clusters were examined by the second round of the diagnostic tests. This iteration served to detect the remnant severely faulty clusters and electrodes and, upon detection, excluded them from the rest of analysis. Finally, the last run of diagnostic tests resolved the rest of the contaminant clusters per electrode. This step finalized the diagnostic nature of ClusDiag, and an evaluation of total error quantified all the final FP and FN errors per cluster per electrode.[39] At this point, all the clusters were represented by their templates. For cell detection, spike coincidence events were found among close-by cells whose spikes had been spread among nearby electrodes. For cell identification, the post-stimulus time histogram for the spiking pattern of each cell throughout the stimulation was calculated. Based on the relative difference of peak values during all ON and OFF periods from the histogram, each cell was classified into ON, OFF or ON–OFF type. The block diagram of the spike sorting approach is shown in Fig. 2. The clustering approach, together with the diagnostic tests, found the most reliable template waveforms representing individual neuronal units per electrode. The properties of each analytical layer of ClusDiag are outlined in the rest of this section up to Sec. 2.8.

2.1. Data preparation

Extracellular recording at sampling rate of f_s = 25 kHz from low-impedance electrodes yields a typically low signal-to-noise ratio. In addition, when the dendrite–soma axes of the neighboring cells are aligned, their signals summate, and a compound fluctuating signal (local field potential) is generated. To ameliorate these existing physiological and bioelectrical constraints, the data preparation stage processed the raw data at two steps. First, the periodogram estimate of power spectral density of the data was found. This estimate showed that low-frequency dominant range for local field potentials appeared to exist below 200 Hz, beyond which the power spectral density decreased notably. A zero-phase-shift finite impulse response (FIR) high-pass filter with cutoff frequency of 200 Hz was defined to offset the baseline fluctuation caused by local field potentials. Second, the eigendecomposition technique was applied.[40] This technique has been central in the compass of discrete-time linear filter theory. An FIR filter designed with the optimization criterion to maximize the output signal-to-noise ratio is an eigenfilter. The coefficient vector of the impulse response of such optimum FIR filter is the eigenvector corresponding to the largest eigenvalue of the sample correlation matrix of the stochastic input process.

2.2. Spike detection

Spike is the signal of interest in the extracellular recording stream from RGCs. Techniques used for spike detection include amplitude threshold,[38,40–42] local minima detection,[15,44] spectral analysis,[13,44,45] and adaptive threshold.[37,46,47] A dependable approach needs to accommodate for potential fluctuations in RGCs firing rate during stimulus presentation. Most approaches to spike detection include adaptive corrections, which consider spike amplitude variations to refine the accuracy of spike detection. Here, an adaptive threshold spike detection[37] is implemented, where the objective is to estimate the root-mean-square (rms) value of the background noise in an adaptive way. To this end, the approach uses the frequency of the threshold crossing, rather than spike amplitude, to adjust the rms level of background noise over time. A duty-cycle module is locked to the output of the first clocked comparator and is thus synchronized with each spike recurrence. It incorporates proportional and proportional-integral feedback controls of the rms value to maintain a duty cycle of 0.159 for the output of the first comparator (representing the probability of crossing the one standard deviation (s.d.) threshold for a band-limited Gaussian noise). To emulate the electronics circuit here, the threshold detection employed the discrete samples of its input (filtered voltage) on each electrode to refine the noise rms level over time. This was done through the basic sample-and-hold discrete events for the first comparator's output model, and provided a piecewise constant estimation of the adaptive threshold through each spike recurrence. Setting the integer multiplier of the second comparator to seven led to robustness against false detection due to occasional fictitious threshold crossings. Figure 3 shows the detection of spikes from three different electrodes.

2.3. Feature extraction

We defined a 2.5 ms long temporal window for each detected spike, which was registered to the peak amplitude, with the depolarization interval set to 1.15 ms and that of repolarization set to 1.35 ms. At the recording frequency of 25 kHz, each spike temporal cutout was represented by a vector of 63 samples. We then considered two different feature extraction methods for each detected spike by projecting it onto two feature spaces: (i) a spike-parametric space and (ii) a principal components space. Feature extraction was not trivial, however, and required several trials to identify the most effective set of features to be used in clustering. Previously, we described a fast algorithm for spike clustering based on decomposition of mixture distribution of spike peak depolarization.[44] Here, among the possible spike parameters previously reported,[31,48–50] we initially chose two features: depolarization amplitude and hyperpolarization amplitude. Clustering algorithms were run to inspect the quality of generated clusters. Although some cluster separation was observed using these two features, the quality was reasonably improved by incrementing the number of features. A few temporal parameters were also considered. The rationale to add these features was to make the feature set flexible to the disparate modes of firing, in particular to the decrease in amplitude and increase in width for the trailing spikes of a burst.[39] Also, it is assumed that adding temporal features allows the clustering algorithms to respond more robustly in face of nonstationarity.[51] We therefore included two temporal parameters: (i) the disparity in time between the depolarization and hyperpolarization amplitudes and (ii) the temporal width of the spike at the threshold crossing. Cluster analysis of these combined four features showed a fair improvement in quality. Adding more features (amplitude

variance and zero crossing number) did not yield obvious improvements to between-cluster separation. The combined four-dimensional feature vector is shown in Fig. 4. For cross-evaluation of the clustering through a reference benchmark, we used MC_Rack.[52] For PCA, the number of principal components that captured (95±1)% variation in the normalized spike waveform was calculated. The number of components varied from five to eight accordingly, and the components built the second feature vector.

2.4. Initial cluster analysis

Clustering algorithms here serve to separate the firing activity of neuronal units captured by their salient, representative features.[38,53] Generally, they are best suited to situations where the underlying data structure may not be expressed well by a parametric model. Indeed, as motivated by Ref. 38, in such situations any assumption of parametric form may be imposing an unsound structure to the data. When resorting to clustering algorithms, otherwise, the challenge is pronounced when designing clustering algorithms under conditions of nonstationarity. A solution is to select algorithms that render a reliable stability–plasticity tradeoff.[38] That is, they are stable enough to find the data structure via finite number of clusters, and plastic enough to adapt to a new pattern.

The iterative MSE clustering algorithm was applied to each of the feature sets.[38] To find an optimal partition as one that minimizes the criterion, the procedure started from an initial set of clusters. This algorithm required careful choice of initial number of clusters together with their centers. Instead of choosing random samples as the initial centers for a chosen number of clusters, we applied the basic leader–follower clustering algorithm[38] for learning of centers prior to clustering. To illustrate an example algorithm incorporated in this work, Appendix A includes the pseudocode of this clustering algorithm.

With the initial learning phase for number of clusters and their centers given by the basic leader–follower clustering algorithm, the iterative MSE clustering algorithm was applied to formalize clusters.[38] After the assignment by the algorithm, updating of the centers at the last step resulted in some clusters becoming very close to each other. Therefore, we implemented a merging mechanism for such clusters. Figure 5 shows the result of initial run and sorted spikes for three electrodes. In the three-dimensional cluster space, we observed that clusters formed mostly ellipsoidal clouds of various sizes and orientations.

2.5. Failure diagnosis

We sought to evaluate the quality of spike sorting after the first round of clustering to address possible overclustering[54,55] and intermixing of clusters.[22,55] These potential error sources could result in either overestimating (FP cases) or underestimating (FN cases) the average firing rate of individual RGC units or misrepresent their correlation to visual stimuli. A possible treatment of this issue is to apply splitting[15] and merging[55] procedures based on inference from the average firing rate. To check the clustering quality, we employed a set of tests for diagnostic knowledge representation. We then developed an approach to reiterate the clustering routines for additional improvement of sorting quality.

2.5.1. Diagnostic tests—Fault isolation tests report diagnostics in order to improve the clustering results. The test strategy often involves a testability tool in the form of a tabular or graphical fault tree for failure mode and effects analysis.[56–58] A well-known testability tool consists of a data structure that relates fault isolation modes to the test results. This data structure is mostly referred to as a D-matrix, which is a pivotal element of the test strategy computation.[60] The D-matrix tabulates signatures of binary digits indicating whether the failure modes are detected by the designated tests. Failure modes herein refer to FP and FN cases in clustering. We here describe the structure of each test and explore the diagnosability of the failures. We assessed the quality of the spike sorting through three diagnostic tests per electrode. This set of tests was applied to the clusters generated after each clustering run (diagnostic tests Run 1 and Run 2, see Fig. 2) and a third time after the removal of faulty electrodes or clusters (Run 3, Fig. 2), each serving its distinct diagnostic purpose.

2.5.1.1. Unimodal amplitude scale factor: This test was applied to each cluster from each electrode to detect the FP cases over the first run of clustering. We first identified the template waveform for each cluster as the pointwise median of all the spikes per cluster. In this test, we defined the amplitude scale factor as the ratio between the maximum depolarization amplitude of individual spikes (feature A_1) and the corresponding amplitude of the template waveform for the same cluster. We then calculated the histogram of the distribution of amplitude scale factors for each cluster. The distribution of this histogram is shown to be unimodal around unity for reliable templates.[22] A normal distribution centered at unity, with standard deviation equal to that of the amplitude scale factor, was compared against the normalized histogram. We then inferred the mean absolute error (MAE) statistic to calculate the difference between the normalized histogram and the normal distribution for each cluster on each electrode, and proceeded to test its statistical significance.

The unimodality assumption can be tested between the two competing hypotheses: $H_0 : \theta = 0$ null hypothesis against the one-sided alternative hypothesis $H_1 : \theta > 0$. For the observed MAE values z from θ across the electrodes, a half-normal distribution as an approximation fit to the histogram of likelihood function under the null hypothesis was drawn. The test was run at the 30% level of significance for the null hypothesis. This level was a proper likelihood for failure detection due to the multimodality of amplitude scale factor, among all possible levels between 10% and 30%, in increment of 5. The significance statistic to rule out H_0 in favor of H_1 is then $z > \lambda$, where λ is the 70% quantile of the half-normal distribution. For any electrode, then the corresponding diagnostic test can be formed:

$$T_c = \begin{cases} 1 & \text{if } z > \lambda, \\ 0 & \text{otherwise,} \end{cases} \qquad (1)$$

where c denotes the cluster index, and the test can be run for all clusters. Here, three clusters were generated at max by the clustering routines, so their corresponding T_cs were listed in Table 1 (T_1–T_3). This test can detect contaminations by FP cases that violate unimodality. Such FP cases were seen in situations where spikes from different cells were included into the same cluster, i.e. some spikes were falsely assigned to a given cluster. Figure 6 shows the

distributions of amplitude scale factor for the three electrodes by their first clusters, and histograms of the MAE statistic for clusters 1 and 2 across all the electrodes.

2.5.1.2. Mutual refractory period incident: When not all the spikes of a cell unit are sorted together, FN cases may occur. In that case, one can proceed to assess the correlation among the spikes of seemingly disparate clusters for intercluster separation. This correlation is determined by counting near-coincident spike events to derive their likelihood functions.[59] Therefore, the spike-time cross-correlation function was found for different pairs of clusters per electrode to check if spikes in each cluster had been emitted from a unique cell.[39] For two independent clusters, this function should have no falling gap within [−1 ms, +1 ms]. A gap may signify that the spikes are mutually respecting the refractory period across clusters, and thus some or all spikes from the two clusters might be derived from one unique cell. This then raises an FN case, which affects the clustering specificity. The refractory period window was set to 2 ms, and the inference of mutual refractory period (MRP) between any pair of clusters per electrode was performed by the following test:

$$T_e = \begin{cases} 1 & \text{if MRP exists,} \\ 0 & \text{otherwise,} \end{cases} \qquad (2)$$

where e is the electrode index. T_e signifies the MRP incident on each electrode and is the fourth tabulated test (T_4) in Table 1. It effectively detected the FN cases that in turn violated autonomy of the clusters. Figure 6 shows the histogram distributions of the cross-correlation functions between clusters 1 and 2 on two electrodes.

2.5.1.3. Interspike interval violation: Tests for the refractory period condition reveal contamination of a given cluster with spikes from other clusters or incorrectly included noise fluctuations (FP instances).[22] We therefore found the distribution of interspike interval (ISI) separately for each cluster per electrode. A refractory period violation was defined as an ISI less than 2 ms. We identified all instances of violation in a cluster, and recorded the probability of violations per cluster per electrode. This probability represents a fraction of the total FP error computed for overall performance assessment. It also indicates the separation quality of clusters and is expected to be less than a certain threshold for reliably identified neuronal units.[22,55] A cluster was deemed to have failed this test if its probability of having ISI violations (denoted as P_{ISI}) was moret han 3%. For each cluster c, then the following test was assumed:

$$T_c = \begin{cases} 1 & \text{if } P_{\text{ISI}} > 0.03, \\ 0 & \text{otherwise.} \end{cases} \qquad (3)$$

This T_c accounts for additional tests (T_5–T_7) in Table 1 that illustrates the D-matrix including all the tests for three sample electrodes across five whole field datasets (WF_1–WF_5) evaluated in this work. This represents the outcome of diagnostic tests performed on the clusters derived from the first run of the clustering routines, applied to the spike parameters derived by both methods of feature extraction (f_{spike} and PCA). Figure 6 shows

two example ISI histograms with different violation rates from cluster 1 on all three electrodes.

2.6. Reclustering through optimization

The outcomes of diagnostic tests, combined with the clustering results, can be used to reassess the clustering routines for diagnosable electrodes, to further minimize the error due to the first run of clustering. The SSE in clustering depends on how the features are clustered and the number of clusters. Therefore, the decision logic for cluster reassessment should consider the number of clusters. A decision logic was devised to harness the extant knowledge of clustering and diagnostic tests (Appendix B). This logic relativized itself on the number of clusters, number of failures and the average firing rate in cluster 2 per electrode. The decision was to either retain the electrode, or to rerun the leader–follower routine (with a reduced threshold) followed by the iterative MSE routine (one cluster), or to employ the conjugate gradient method[36] (two or three clusters) before rerunning the iterative MSE routine for assignment.

The SSE across clusters can be improved by methods that minimize error in individual clusters over successive update of cluster center. An efficient method should provide a reasonable convergence rate and an optimal performance in reaching the global minimum of SSE. We chose the conjugate gradient method since it achieved a balance between convergence rate and computation time overhead. This quadratic optimization technique led to iterative descents of SSE per cluster from the successive update of cluster center along the conjugate gradient directions. The outputs were the updated cluster centers (at last iteration, reaching solution) for each diagnosed electrode. Subsequently, the new cluster centers for each electrode were passed to the iterative MSE clustering for the second run. The results of second run of clustering on the three electrodes are depicted in Fig. 7. Across all datasets, the improvement was statistically significant. For example, after reclustering, when Run 2 of diagnostic tests was applied to data on E_{35} and E_{54} in WF_3 dataset, these electrodes were not among the faulty electrodes nor did they have even a faulty cluster to be discarded. This exemplifies the role of conjugate gradient method in improving the specificity and sensitivity of newly formed clusters.

2.7. Final clusters

Table 2 represents the D-matrix for diagnostic tests after Run 2. The improved clustering quality is demonstrated in a few cases, e.g. E_{35} in WF_3 dataset with f_{spike} turns out error-resilient because of reclustering. The implementation of merging of clusters, due to the reclustering, is also evident (e.g. see number of clusters for E_{52} in WF_3 dataset with PCA between Tables 1 and 2).

Having refined the clustering through two runs of diagnostic tests, we proceeded to discard a few clusters or electrodes that were still severely faulty. The second run of diagnostic tests was based on the number of FP cases in terms of spike amplitude and ISI violation rate. The decision for each electrode was either to discard it, or to discard only few of its clusters, or otherwise to retain it for the last run of diagnostic tests (Run 3). Appendix B conveys the

flowchart of the logic to detect faulty clusters or electrodes. Some quantities of the approach from the initial clustering to the second run of diagnostic tests are listed in Table 3.

The previous two runs of diagnostic tests dealt with the failure detection across all the electrodes. The last run, Run 3, served to make additional corrections into spikes within and across clusters, inferred from observed instances. First, within the amplitude scale factor test as described, if the test outcome was significant, then we excluded from the cluster the spikes for which the amplitude scale factors were less than a threshold. This refinement of the cluster was necessary only for a few electrodes. Second, for mutual refractory period if the gap existed for any two clusters, then we retrieved the clusters whose spikes led to this gap. We ultimately found the cluster that resulted in most gap violations per electrode and merged that cluster with the cluster that had the second most gap violations. This step made the required merges only for a few electrodes. Finally, we recorded the ISI violation rate per cluster. This was a fraction of the total FP error in overall performance assessment. The spurious spikes of violating ISI were removed from the corresponding cluster. The remaining electrodes and clusters thereof were represented with their templates for the rest of the analysis, namely, cell detection and error assessment. Table 5 includes the final number of clusters and spikes for each stimulus.

2.8. Cell detection and identification

At this point, the final clusters through the execution of ClusDiag were formed. If the identified clusters represent the cell activity recorded on different electrodes, what is the number of active cells during each stimulation? The distinction between number of clusters and number of cells appears because some clusters may represent the same cell whose spikes were detected on very few nearby electrodes. To see if this was the case for any cluster c_1 on an electrode, we searched for time coincidence between any spike of c_1 and spikes from all the other clusters of nearby electrodes (nearest lateral and diagonal). For example, one iteration of search compared the time differences (lags) between a spike of c_1 with all the spikes of another cluster c_2 on a nearby electrode. Having set the coincidence interval to be 1 ms,[7,60] then a coincident spike from c_2 on the other electrode was detected if the corresponding time difference was less than 1 ms. This step continued for all the spikes of c_1 with respect to those of c_2. If most of the spikes of c_1 each had a paired coincident spike from c_2, then the corresponding electrodes were identified to have recorded the activity of one cell (represented with two clusters). We did not observe such firing coincidence in our data. Therefore, the final number of generated clusters and the total number of detected cells were equal, as shown in Table 4.

Controlled light stimulation of the retina triggers highly correlated bursting of RGCs, signifying spike frequency adaptation in relation to luminosity and pattern of the light stimulus. That is, certain cells fire more frequently in response to increase of luminosity, whereas others are silenced or maintain an unchanged firing rate. From correlation of the visual stimulus properties to the spike train of individual identified RGCs, one can characterize and classify RGC types into distinct classes. To examine the frequency response of the detected cells in relation to the stimuli, we calculated the cumulative PSTH for spikes per cluster. This histogram is statistically depicted to identify the short latency responses to

light spot within three various classes.[61] They are ON, OFF and ON–OFF classes. This classification was evident in our data likewise. The whole field stimuli evoked RGCs' light responses, proving the reception of a visual input.

To find the cumulative PSTH, we first retained and aggregated the spike times of a cell during all ON periods (2 s long), as well as separately for those of all OFF periods (2 s long). Then, the post-stimulus time histogram of firing was found (bin width: 50 ms). Next, to proceed with a basic classification of responses, we defined a ratio index that indicated the repertoire of light responsiveness of a cell. The index definition for each cell was $I_I = (A_{ON} - A_{OFF})/(A_{ON} + A_{OFF})$, where A_{ON} is the maximum amplitude of cumulative PSTH during the ON stimulation periods and the maximum amplitude is A_{OFF} for the OFF periods. This index accounts for relative difference during ON and OFF periods, normalized by sum of the two amplitudes. As per discrete observations from cumulative PSTH over the results to different whole stimuli, we set the limits to identify cells' responses. They were $0.2 < I_I \leq 1$ for ON cells (if $A_{ON} = 1.5 \times A_{OFF}$), $-1 \leq I_I < -0.2$ for OFF cells (if $A_{OFF} = 1.5 \times A_{ON}$) and $-0.2 \leq I_I \leq 0.2$ for ON–OFF cells. There are two example cumulative PSTH responses illustrated in Fig. 8. Cell identification summaries for all the datasets are provided in Table 4. It illustrates that both feature extraction techniques yield similar numbers of identified cell types.

2.9. Error assessment

We recorded the FP and FN errors due to each cluster as well as the overlap between pairs of clusters after each run of clustering routines. We now assess the total error in spike sorting with respect to the raw data SNR ratio for each electrode. First, the noise segments were taken as the portions of raw recordings in which there was no spike of any type. From the detected and sorted spikes per electrode, we thus estimated the noise signature on each electrode by removing each spike cutout from the raw data. Next, we examined the stationarity and distributional spread of the noise over the electrodes. A sketch of space–time dependency of the observed noise correlation function over a smaller ROI is shown in Fig. 9. Particularly, the cross-correlation function of noise showed that it was isotropic along x- and y-directions, with a significant decay at minimum interelectrode distance of 100 μm. This distance may well explain the degree of noise and, hence, recorded voltage correlatedness. Hence, the fact that electrodes have little cross-talk with each other is consolidating our finding that the number of final recorded RGCs coincides with the number of clusters reliably detected over all electrodes. In addition, we tested the Gaussian assumption for the noise by the negentropy criterion.[62] It determines how Gaussian a distribution is by taking a differential measurement of entropy of any distribution with respect to a Gaussian distribution with the same first two moments. This metric is statistically robust, and for a Gaussian, it is equal to zero. Figure 9 includes the negentropy values for all the electrodes. The mean of all negentropy values across the electrodes was 0.88 bit.

Because here we estimated the noise signature based on the sorted spikes, the SNR will be an estimated SNR.[14] It can be defined as the difference between the sum of averaged variances of all spikes in a cluster and the sample variance of noise, divided by the sample variance of noise. The total error for single-unit recording is defined in Ref. 14. We extended

this definition to the MEA system. The total error of spike sorting per cluster c per electrode e was defined as the ratio $TE_{e,c} = (FP_{e,c} + FN_{e,c})/(FN_{e,c} + TP_{e,c})$. The FP error, FN error and true-positive probability are represented by $FP_{e,c}$, $FN_{e,c}$ and $TP_{e,c}$, respectively. The FP error for each cluster per electrode was defined to be the maximum of single-cluster error and the multiple-cluster errors.[39] The single-cluster error was the ISI violation rate, whereas the multiple-cluster errors factored in the probability that overlap between a cluster and the others had made an FP. Between each pair of clusters this probability was calculated by fitting a multivariate Gaussian distribution into the spikes in each cluster.[39] In case of singularity of the covariance matrix, a Choleskey-like covariance decomposition was applied. The FN error for each cluster per electrode was calculated by combining the terms of the single-cluster error and multiple-cluster errors. The single-cluster error ascertained the error from spike detection. The multiple-cluster errors deal with the overlap between pairs of clusters by expressing the errors in terms of the probability that a sample spike from one cluster is misassigned to another cluster. In addition, the true positive probability was calculated by finding the probability that spikes in a cluster indeed belong to that cluster. We calculated this probability together with fractional FP and FN errors for each cluster per electrode to find the corresponding total error. We found that the total error was dominated by FP error per cluster across the electrodes. The FN error turned out to be quite low, so ClusDiag performed optimally in the sense that it minimized the probability of the FN error.

Figure 8 shows the total error for cluster 1 as a function of the estimated SNR across all electrodes for different stimuli. It can be seen that the error remains low for the range of estimated SNRs for these experiments. Moreover, comparison between the feature extraction methods indicates that f_{spike} is more error-resilient than PCA by a good margin for SNRs fewer than a half, although performance becomes close for SNRs between 0.5 and 1. Beyond unity SNR, the former outperforms the latter. Some elaborate quantities of the error assessment, including example specificity and sensitivity values from cluster 1, are listed in Table 5. We also paid attention to computation times for different stages of ClusDiag to run their course during its execution. In particular, for illustrative values here, we divided the overall runtime to two cumulative runtimes. First, we recorded the runtime of the two rounds of clustering algorithms, together with that of the optimization routine and first round of diagnostic tests, per electrode (ASR1 in Table 5). Second, the runtime for the execution of the second and third rounds of diagnostic tests per electrode was saved (ASR2 in Table 5). These two runtimes nearly balance each other out across all datasets and feature sets (e.g. for PCA the average ASR1 over all datasets is 1.84 s, while that of ASR2 is 1.71 s). The last run of diagnostic tests had extra corrective steps applied to each diagnosed cluster, each of which imposing its own load on the ASR2 division. The workstation here was an ×64-based PC with Intel Xeon eight-core CPU E5-2630 v3 at 2.40 GHz.

3. Discussion

Neuro-inspired machine learning algorithms are amenable to address the spike sorting problem in different biological substrates. A survey of previously implemented approaches indicates that the existing challenges fall into four categories from an algorithmic perspective:

(i) Robustness against modes and frequency of firing, when confronted with: sparsely firing neurons,[63] regular spiking versus bursting,[29,30,35] multispike trains[21,32,33] and compounding local field potentials.[34,36]

(ii) Automated spike sorting to reduce as much time due to user's intervention as possible for evaluating or enhancing cluster assignment,[22,64] and processing time tradeoff versus sorting quality.[55,63]

(iii) Error analysis such as relating FP and FN along with diagnosis,[39,65] low signal-to-noise ratio,[25,66] intercluster correlation[39,67] and interelectrode correlation.[15,68,69]

(iv) Potential blueprint for a toolbox that manages delivery across varying signal conditioning from acquisition electronics,[39,70] number of micro-electrodes[69,71–73] and geometrical differences of array.[51,72]

These various challenges were considered in our approach at different steps to balance sorting accuracy with computational cost. We ultimately converged on an approach built upon iterative, exploratory computation routines in an integrated platform. The implementation employed methods across three analytical layers: representation (data preparation and unsupervised clustering), evaluation (D-matrix and error rate) and optimization (conjugate gradient method).

Various modes of firing in RGCs and spike morphology variability require holistic and tolerant spike detection and feature extraction routines. An adaptive threshold spike detector and a parametric spike feature set provide robustness against nonstationary dynamics and result in improved detection and representation of spikes despite varying firing modes (e.g. regular, burst or tonic firing). For PCA-based feature sets, it is often observed that projection onto the few first principal components may not result in optimal intercluster separability.[20,22] PCA-based techniques also cannot assign weights to the various aspects of the shapes of the spikes to be sorted. Here, we noticed that, as per the initial number of clusters in Table 3, PCA is more likely to overcluster the data. Furthermore, as Fig. 8 illustrates, even after implementing various refinement steps, the FP error from the overlap between the clusters appears to dominate the error in PCA-based approach. In addition, deviation of the distribution of a feature set from normality restricts the generalization of most unsupervised clustering algorithms.[20,22] Therefore, user's verifications become part of the cluster-cutting evaluation process to gradually derive more compelling clusters. However, these verifications could in fact reinforce human bias, and the cluster-cutting process captures the underlying natural clusters solely through excessive trials,[22,74] which is the reason behind to evaluate and diagnose the clusters against computation routines.

Clustering-based spike sorting approaches infer statistics about the spatiotemporal correlation of spike events along with additive noise, yet they may be subject to error in validation of the results of clustering and may not perform well at scale for long recordings. ClusDiag achieved high sensitivity and specificity to sort the spikes of different extracellular recordings. That was mainly due to automation and inferential robustness by improving clustering results through a trial of diagnostic tests. To test the feasibility for a long recording, we applied ClusDiag on 5-min long recordings of retinas carrying wild-type and

phosphonull versions of the photopigment melanopsin, under conditions of complete synaptic blockade. Despite the small percentage of intrinsically photosensitive RGCs, which respond to light in the absence of classical photoreceptor synaptic input, our approach recovered them reliably and revealed the difference in firing rates between RGCs carrying normal or mutated photopigments.[75]

The field of failure analysis provides unique frameworks to design, test and implement scalable algorithms for failure diagnosis using systems engineering methods, such as simulation, mathematical analysis and failure mode and effects analysis. The tests applied here were effective in providing diagnostics for the reclustering routines of spike sorting. They attributed different properties of clusters to the spiking dynamics of their underling cells. Typically cells of distinct spiking dynamics show unique spike shapes and frequencies whose patterns could be revealed by applying inferential statistics. For example, Fig. 6 reveals a systematic difference in MAEs between the populations of cluster 1 and cluster 2 across electrodes. This distinction could be due to the nature of the algorithm in the clustering procedure. The leader–follower algorithm found the number of clusters in the order of temporal appearance of spikes, which possibly depended on the spiking frequency of each cell. Therefore, a fast spiking cell was more likely to have had its spikes assigned to cluster 1 on each electrode, hence cluster 1 included more spikes versus cluster 2. This led to a more accurate template waveform across cluster 1, and thus the respective MAE statistics of spikes were smaller. Structural differences as per cluster type on the other two tests signatures were also observed across all datasets. In addition, the tests collectively served to diagnose failure cases, which gave rise to the optimization algorithm to improve the quality of initial clusters. An obvious challenge for offline clustering-based spike sorting methods is that they tend to overestimate the number of clusters.[22,55] To address this challenge here, the set of diagnostic tests reliably identified the failure cases per electrode over the evaluation cycles of this platform, as seen in Table 5. However, additional tests may leverage certain diagnostic structures for a more dynamic test design under large noise contaminations and error propagation properties. Current set of tests turned out a reasonable computation time complexity for spike sorting per electrode.

Several toolboxes have achieved a compromise between sorting quality, speeding up processing time and scaling up to larger amounts of data.[20,31,51,64,72–78] The generalization performance of toolboxes across various electronics and biological systems, however, is at the frontier of neuroscience, inviting additional research and development. We attempted to implement a spike sorting algorithm employed in cortical and hippocampus recordings,[51] although with limitations, likely due to changes in geometry of the electrode arrays. While continuing our cross-checking efforts, it is possible to generalize the modular and diagnostic core of ClusDiag into other spike sorting domains; that is, due to its modular core from the representation of data to the final evaluation of the platform, other inferential algorithms could be applied, tested and gradually accreted. We envision that additional improvement over performance could be achieved through testing more complex representation models, but one caveat may be that they consume more computation cycles per electrode. Augmenting our current findings with those of others, dealing with visual stimulation paradigms in real-time condition, could spawn new applications for spike inference in visual neuroscience.

4. Conclusion

Timely, elaborate progress across the growing scale of data-driven tasks in visual neuroscience requires iterative, exploratory cycles of development, technical validation, experiments and performance optimization. We reported an integrated analytical platform for spike sorting, with engineered modules from initial representation to intermediary optimization to final evaluation. The centerpiece of this work was the unsupervised clustering supplied with failure diagnosis (dubbed the ClusDiag) for data collected during visual stimulation of wild-type mice retinas with whole field stimuli. In retrospect, an adaptive threshold spike detector and a parametric spike feature set led to robustness against various dynamics of firing in RGCs. The whole field stimuli resulted in massive, highly-correlated RGC bursting, combined with robust field potential responses from the retina. Final evaluation of this approach marked a dependable specificity and sensitivity in performance for all the whole field stimuli. Hence, we envision that this approach will perform reliably when extended to more complex stimulation conditions, where lower levels of synchronized activity are expected. To sort spikes of the retinal ganglion cells with widespread firing dynamics, ClusDiag performed robustly across the stimuli, thanks to its automated, fault-tolerant and computation-efficient platform. Distributed machine learning platforms will be an active theme for sourcing, processing and integrating large-scale data from high-density micro-electrode arrays. This would let us explore the prime potential of transforming ClusDiag into a distributed platform with specialized computing frameworks.

Appendix A. Cluster Analysis: Leader–Follower Algorithm

For unsupervised learning prior to iterative MSE clustering algorithm, the leader–follower algorithm (see Algorithm A.1) was applied.[38] Feature vectors of spikes were incorporated into the analysis in the temporal succession of their spike detection. In the second run of clustering the locations of the centers of clusters were updated within the conjugate gradient method, making the combined training more robust to the order of presentation of feature vectors. The algorithm measured the similarity of each feature vector to the centers of clusters by 1-norm distance. If the distance from the winner cluster (closest) was below a threshold, the spike was incorporated in the cluster, and the algorithm updated the center of this winner cluster by a learning rate, to adjust for the newly incorporated spike. In contrast, when the distance was higher than the threshold, the number of clusters incremented, and the current spike seeded a new cluster. The algorithm iterated through all the spike feature vectors and finally returned the centers for the formed clusters.

The learning rate was tweaked for the best cluster resolution to be $\eta = (0.4)^s$ for the sth spike. Finding a suitable threshold θ demanded more trials for the best possible heuristic value. To find a threshold for each electrode, individual two s.d.s along each feature were calculated and added with those of the rest. The pseudocode description follows, where the feature vector f_s is either four-dimensional extracted from the detected spikes and normalized on each dimension or multidimensional comprised of the principal components for each detected spike. The variables N_c, N_s and c_i are in order representing the number of clusters, the number of spikes and each formed cluster. The first cluster is formed from the first feature vector in line 2. Cluster update is represented in lines 6 and 7, and so forth is

cluster learning in lines 8–11. The cluster set $\{c_i\}_{i=1}^{N_c}$ was obtained through the run for each electrode. Their centers, $\{\mu_i\}_{i=1}^{N_c}$, were stored in an array indexing through the electrodes. Each array element included N_c centers of clusters for an electrode.

Algorithm A.1.

The leader–follower clustering

```
1   Declare μ and initialize η, θ, N_c
2   μ_1 ← f_1
3   for e = 1,… ,E do
4     for s = 1,… ,N_s do
5       c ← arg min_c' ||f_s − μ_c'||
6       if ||f_s − μ_c|| < θ then
7         μ_c ← μ_c + ηf_s
8       else
9         N_c ← N_c + 1
10        μ_Nc ← f_s
11        μ_Nc ← μ_Nc/||μ_Nc||
12      end if
13    end for
14  end for
15  return {c_i, μ_i}_{i=1}^{N_c}
```

Appendix B. Decision Logics for Iterative Evaluation

Shown here is the flowchart (see Fig. B.1) of factors of decision for reclustering per electrode based on outcomes of the first run of clustering and diagnostic tests. This logic formed according to the number of clusters per electrode, number of failures and average firing rate. In case of two clusters, decisions were made based on the number of failures and the average firing rate in cluster 2. This was based on our observation that, for cluster 2, a failure when testing on amplitude scale factor was typically the result of insufficient number of spikes. For this reason, only large clusters were routed to reclustering. Small clusters, showing failures in amplitude scale factor, were addressed via the last run of the diagnostic tests. The thresholds for large clusters were set based on the number of failures. The average firing rate, $\overline{r_2}$, was found diagnostic in case of one failure, and a lower threshold, $\overline{r_3}$, was so in case of two or three failures. With three clusters, any failure prompted a reassessment of the clustering, since the quality of sorting was very likely affected. In case of one cluster, if there was any failure, the average firing rate, $\overline{r_1}$, was used as a threshold. For an average firing rate more than $\overline{r_1}$, clustering routines were run again beginning with the leader–follower algorithm (Run 2). The decision for each electrode was either to retain it as was, or

to run the conjugate gradient method (when there were two or three clusters), or to rerun the leader–follower algorithm (one cluster).

Next shown here is the flowchart (see Fig. B.2) of metrics to detect the faulty clusters and electrodes after the second run of clustering and diagnostic tests. The progression based itself on observations of FP failures in terms of spikes amplitudes and ISI violation rate. Severely faulty clusters were detected and discarded when fewer tests had failed. At higher numbers of failures (three or more), however, an entire electrode was discarded. For fewer than three failures, the reclustering results indicated that the FP failures dominated the failure set, so they were dealt with first. A threshold value of 6% ISI violation rate (a two-fold increase compared to the initial limit) detected clusters with severe contamination. If this condition was not met (severity was low), the next check was on violation of unimodality for spikes[1] amplitudes per cluster. To do so, the threshold of one-sided alternative hypothesis test on MAE values was increased to 95th percentile point. Thus, if the MAE value fell above this threshold, the null hypothesis was rejected, and the cluster was discarded. Finally, for all the retained electrodes, we ran a set of diagnostic tests (Run 3).

References

1. O'Brien BJ, Isayama T, Richardson R and Berson DM, Intrinsic physiological properties of cat retinal ganglion cells, J. Physiol 538 (2002) 787–802. [PubMed: 11826165]
2. Badea TC and Nathans J, Morphologies of mouse retinal ganglion cells expressing transcription factors Brn3a, Brn3b, and Brn3c: Analysis of wild type and mutant cells using genetically-directed sparse labeling, Vision Res 51 (2011) 269–279. [PubMed: 20826176]
3. Baden T et al., The functional diversity of retinal ganglion cells in the mouse, Nature 529 (2016) 345–350. [PubMed: 26735013]
4. Volgyi B, Chheda S and Bloomfield SA, Tracer coupling patterns of the ganglion cell subtypes in the mouse retina, J. Comp. Neurol 512 (2009) 664–687. [PubMed: 19051243]
5. Zeck GM and Masland RH, Spike train signatures of retinal ganglion cell types, Eur. J. Neurosci 26 (2007) 367–380. [PubMed: 17650112]
6. Farrow K and Masland RH, Physiological clustering of visual channels in the mouse retina, J. Neurophysiol 105 (2011) 1516–1530. [PubMed: 21273316]
7. Meister M, Pine J and Baylor DA, Multi-neuronal signals from the retina: Acquisition and analysis, J. Neurosci. Methods 51 (1994) 95–106. [PubMed: 8189755]
8. Eversmann B. A 128 × 128 CMOS bio-sensor array for extracellular recording of neural activity; IEEE International Solid-State Circuits Conf.: Digest of Technical Papers (ISSCC); 2003;
9. Fiscella M et al., Recording from defined populations of retinal ganglion cells using a high-density CMOS-integrated microelectrode array with real-time switchable electrode selection, J. Neurosci. Methods 211 (2012) 103–113. [PubMed: 22939921]
10. Lambacher A et al., Identifying firing mammalian neurons in networks with high-resolution multi-transistor array (MTA), Appl. Phys. A 102 (2011) 1–11.
11. Feller MB, Wellis DP, Stellwagen D, Werblin FS and Shatz CJ, Requirement for cholinergic synaptic transmission in the propagation of spontaneous retinal waves, Science 272 (1996) 1182–1187. [PubMed: 8638165]
12. Yuste R, Peinado A and Katz LC, Neuronal domains in developing neocortex, Science 257 (1992) 665–669. [PubMed: 1496379]
13. Yang HH and St-Pierre F, Genetically encoded voltage indicators: Opportunities and challenges, J. Neurosci 36 (2016) 9977–9989. [PubMed: 27683896]

14. Kim S and McNames J, Automatic spike detection based on adaptive template matching for extracellular neural recordings, J. Neurosci. Methods 165 (2007) 165–174. [PubMed: 17669507]
15. Marre O, Amodei D, Deshmukh N, Sadeghi K, Soo F, Holy TE and Berry MJ, Mapping a complete neural population in the retina, J. Neurosci 32 (2012) 14859–14873. [PubMed: 23100409]
16. Nguyen TK et al., Closed-loop optical neural stimulation based on a 32-channel low-noise recording system with online spike sorting, J. Neural Eng 11 (2014) 046005. [PubMed: 24891498]
17. Zhang PM, Wu JY, Zhou Y, Liang PJ and Yuan JQ, Spike sorting based on automatic template reconstruction with a partial solution to the overlapping problem, J. Neurosci. Methods 135 (2004) 55–65. [PubMed: 15020089]
18. Franke F, Real-time analysis of extracellular multi-electrode recordings, thesis, Technische Universiäat Berlin, Germany (2011).
19. Vollgraf R, Munk M and Obermayer K, Optimal filtering for spike sorting of multi-site electrode recordings, Network 16 (2005) 85–113. [PubMed: 16350435]
20. Franke F, Quian Quiroga R, Hierlemann A and Obermayer K, Bayes optimal template matching for spike sorting — Combining Fisher discriminant analysis with optimal filtering, J. Comput. Neurosci 38 (2015) 439–459. [PubMed: 25652689]
21. Pillow JW, Ahmadian Y and Paninski L, Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains, Neural Comput 23 (2011) 1–45. [PubMed: 20964538]
22. Prentice JS, Homann J, Simmons KD, Tkacik G, Balasubramanian V and Nelson PC, Fast, scalable, Bayesian spike identification for multi-electrode arrays, PLoS ONE 6 (2011) e19884. [PubMed: 21799725]
23. Thiel A et al., Contribution of individual retinal ganglion cell responses to velocity and acceleration encoding, J. Neurophysiol 98 (2007) 2285–2296. [PubMed: 17596411]
24. Leibig C, Wachtler T and Zeck G, Unsupervised neural spike sorting for high-density microelectrode arrays with convolutive independent component analysis, J. Neurosci. Methods 271 (2016) 1–13. [PubMed: 27317497]
25. Pillow JW, Shlens J, Chichilnisky EJ and Simoncelli EP, A model-based spike sorting algorithm for removing correlation artifacts in multi-neuron recordings, PLoS ONE 8 (2013) e62123. [PubMed: 23671583]
26. Swindale NV and Spacek MA, Spike sorting for polytrodes: A divide and conquer approach, Front. Syst. Neurosci 8 (2014) 6. [PubMed: 24574979]
27. Takahashi S, Anzai Y and Sakurai Y, Automatic sorting for multi-neuronal activity recorded with tetrodes in the presence of overlapping spikes, J. Neurophysiol 89 (2003) 2245–2258. [PubMed: 12612049]
28. Zouridakis G and Tam DC, Identification of reliable spike templates in multi-unit extracellular recordings using fuzzy clustering, Comput. Methods Programs Biomed 61 (2000) 91–98. [PubMed: 10661394]
29. Cessac B and Palacios AG, Spike train statistics from empirical facts to theory: The case of the retina, in Modeling in Computational Biology and Biomedicine: A Multidisciplinary Endeavor (Springer, Berlin, 2013), pp. 261–302.
30. Chen AH, Zhou Y, Gong HQ and Liang PJ, Firing rates and dynamic correlated activities of ganglion cells both contribute to retinal information processing, Brain Res 1017 (2004) 13–20. [PubMed: 15261094]
31. Quiroga RQ, Nadasdy Z and Ben-Shaul Y, Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering, Neural Comput 16 (2004) 1661–1687. [PubMed: 15228749]
32. Abeles M and Goldstein MH, Multispike train analysis, Proc. IEEE 65 (1977) 762–773.
33. Brown EN, Kass RE and Mitra PP, Multiple neural spike train data analysis: State-of-the-art and future challenges, Nat. Neurosci 7 (2004) 456–461. [PubMed: 15114358]
34. Kremkow J et al., Neuronal nonlinearity explains greater visual spatial resolution for darks than lights, Proc. Natl. Acad. Sci. USA 111 (2014) 3170–3175. [PubMed: 24516130]
35. Quian Quiroga R and Panzeri S, Extracting information from neuronal populations: Information theory and decoding approaches, Nat. Rev. Neurosci 10 (2009) 173–185. [PubMed: 19229240]

36. Bertsekas DP, Nonlinear Programming (Athena Scientific, Belmont, 1999).
37. Watkins PT, Santhanam G, Shenoy KV and Harrison RR, Validation of adaptive threshold spike detector for neural recording, in Proc. 26th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (2004), pp. 4079–4082.
38. Duda RO, Hart PE and Stork DG, Pattern Classification (Wiley, New York, 2001).
39. Hill DN, Mehta SB and Kleinfeld D, Quality metrics to accompany spike sorting of extracellular signals, J. Neurosci 31 (2011) 8699–8705. [PubMed: 21677152]
40. Makhoul J, On the eigenvectors of symmetric Toeplitz matrices, IEEE Trans. Acoust. Speech Signal Process 29 (1981) 868–872.
41. Chandra R and Optican LM, Detection, classification, and superposition resolution of action potentials in multiunit single-channel recordings by an on-line real-time neural network, IEEE Trans. Biomed. Eng 44 (1997) 403–412. [PubMed: 9125825]
42. Jackel D, Frey U, Fiscella M, Franke F and Hierlemann A, Applicability of independent component analysis on high-density microelectrode array recordings, J. Neurophysiol 108 (2012) 334–348. [PubMed: 22490552]
43. Neumann T, Ziegler C and Blau A, Multielectrode array recordings reveal physiological diversity of intrinsically photosensitive retinal ganglion cells in the chick embryo, Brain Res 1207 (2008) 120–127. [PubMed: 18377877]
44. Ghahari A and Badea TC, Robust spike sorting of retinal ganglion cells tuned to spot stimuli, in Proc. 38th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (2016), pp. 1745–1749.
45. Nenadic Z and Burdick JW, Spike detection using the continuous wavelet transform, IEEE Trans. Biomed. Eng 52 (2005) 74–87. [PubMed: 15651566]
46. Chan HL, Lin MA, Wu T, Lee ST, Tsai YT and Chao PK, Detection of neuronal spikes using an adaptive threshold based on the maxmin spread sorting method, J. Neurosci. Methods 172 (2008) 112–121. [PubMed: 18508127]
47. Liu X, Yang X and Zheng N, Automatic extracellular spike detection with piecewise optimal morphological filter, Neurocomputing 79 (2012) 132–139.
48. Lewicki MS, A review of methods for spike sorting: The detection and classification of neural action potentials, Network 9 (1998) R53–R78. [PubMed: 10221571]
49. Paraskevopoulou SE et al., Feature extraction using first and second derivative extrema (FSDE) for real-time and hardware-efficient spike sorting, J. Neurosci. Methods 215 (2013) 29–37. [PubMed: 23403106]
50. Zamani M and Demosthenous A, Feature extraction using extrema sampling of discrete derivatives for spike sorting in implantable upper-limb neural prostheses, IEEE Trans. Neural Syst. Rehabil. Eng 22 (2014) 716–726. [PubMed: 24760942]
51. Rossant C et al., Spike sorting for large, dense electrode arrays, Nat. Neurosci 19 (2016) 634–641. [PubMed: 26974951]
52. Multi Channel Systems, MC_Rack (2015), http://www.multichannelsystems.com/software/mc-rack.
53. Everitt B, Cluster Analysis (Wiley, Chichester, 2011).
54. Pedreira Gallego C, Recordings in the human temporal lobe: Data analysis and spike sorting, Ph.D. thesis, University of Leicester, Leicester, UK (2010).
55. Rutishauser U, Schuman EM and Mamelak AN, Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, *in vivo*, J. Neurosci. Methods 154 (2006) 204–224. [PubMed: 16488479]
56. Sampath M, Sengupta R, Lafortune S, Sinnamohideen K and Teneketzis DC, Failure diagnosis using discrete-event models, IEEE Trans. Control Syst. Technol 4 (1996) 105–124.
57. Zad SH, Kwong RH and Wonham WM, Fault diagnosis in discrete-event systems: Framework and model reduction, IEEE Trans. Autom. Control 48 (2003) 1199–1212.
58. Rish I et al., Adaptive diagnosis in distributed systems, IEEE Trans. Neural Netw 16 (2005) 1088–1109. [PubMed: 16252819]

59. Palm G, Aertsen AMHJ and Gerstein GL, On the significance of correlations among neuronal spike trains, Biol. Cybern 59 (1988) 1–11. [PubMed: 3401513]
60. Schnitzer MJ and Meister M, Multineuronal firing patterns in the signal from eye to brain, Neuron 37 (2003) 499–511. [PubMed: 12575956]
61. Carcieri SM, Jacobs AL and Nirenberg S, Classification of retinal ganglion cells: A statistical approach, J. Neurophysiol 90 (2003) 1704–1713. [PubMed: 12966177]
62. Hyvrinen A and Oja E, Independent component analysis: Algorithms and applications, Neural Netw 13 (2000) 411–430. [PubMed: 10946390]
63. Pitkow X and Meister M, Decorrelation and efficient coding by retinal ganglion cells, Nat. Neurosci 15 (2012) 628–635. [PubMed: 22406548]
64. Aghagolzadeh M, Mohebi A and Oweiss KG, Sorting and tracking neuronal spikes via simple thresholding, IEEE Trans. Neural Syst. Rehabil. Eng 22 (2014) 858–869. [PubMed: 24240005]
65. Vasserman G, Shamir M, Ben Simon A and Segev R, Coding “what” and “when” in the Archer fish retina, PLoS Comput. Biol 6 (2010) e1000977. [PubMed: 21079682]
66. Kyung Hwan K and Sung June K, A wavelet-based method for action potential detection from extracellular neural signal recording with low signal-to-noise ratio, IEEE Trans. Biomed. Eng 50 (2003) 999–1011. [PubMed: 12892327]
67. Meister M, Lagnado L and Baylor DA, Concerted signaling by retinal ganglion cells, Science 270 (1995) 1207–1210. [PubMed: 7502047]
68. Van Dijck G, Seidl K, Paul O, Ruther P, Van Hulle MM and Maex R, Enhancing the yield of high-density electrode arrays through automated electrode selection, Int. J. Neural Syst 22 (2012) 1–19. [PubMed: 22262521]
69. Yger P et al., Fast and accurate spike sorting *in vitro* and *in vivo* for up to thousands of electrodes, bioRxiv 067843 (2016), doi:10.1101/067843.
70. Buzsaki G, Stark E, Berenyi A, Khodagholy D, Kipke DR, Yoon E and Wise KD, Tools for probing local circuits: High-density silicon probes combined with optogenetics, Neuron 86 (2015) 92–105. [PubMed: 25856489]
71. Hilgen G et al., Unsupervised spike sorting for large-scale, high-density multielectrode arrays, Cell Rep 18(10) (2017) 2521–2532. [PubMed: 28273464]
72. Pachitariu M, Steinmetz N, Kadir S, Carandini M and Harris KD, Kilosort: Realtime spike-sorting for extracellular electrophysiology with hundreds of channels, bioRxiv 061481 (2016), doi: 10.1101/061481.
73. Hilgen G et al., Pan-retinal characterisation of light responses from ganglion cells in the developing mouse retina, Sci. Rep 7 (2017) 42330. [PubMed: 28186129]
74. Harris KD, Henze DA, Csicsvari J, Hirase H and Buzsaki G, Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements, J. Neurophysiol 84 (2000) 401–414. [PubMed: 10899214]
75. Somasundaram P et al., C-terminal phosphorylation regulates the kinetics of a subset of melanospin-mediated behaviors in mice, Proc. Natl. Acad. Sci. USA 114 (2017) 2741–2746. [PubMed: 28223508]
76. Ortiz-Rosario A, Adeli H and Buford JA, MUSIC:Expected maximization Gaussian mixture methodology for clustering and detection of task-related neuronal firing rates, Behav. Brain Res 317 (2017) 226–236. [PubMed: 27650101]
77. Bongard M, Micol D and Fernandez E, NEV2lkit: A new open source tool for handling neuronal event files from multi-electrode recordings, Int. J. Neural Syst 24 (2014) 1450009. [PubMed: 24694167]
78. Knieling S, Sridharan KS, Belardinelli P, Naros G, Weiss D, Mormann F and Gharabaghi A, An unsupervised online spike-sorting framework, Int. J. Neural Syst 26 (2016) 1550042. [PubMed: 26711713]

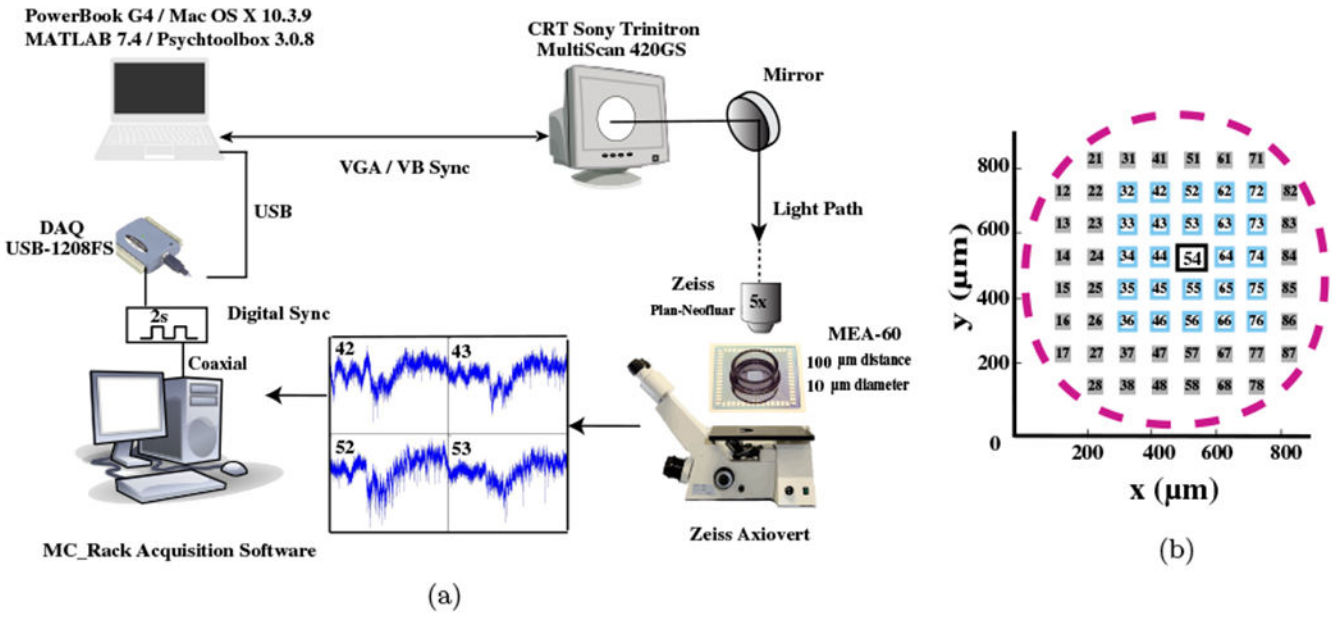

Fig. 1.

Light stimulation and data acquisition hardware and software. (a) MEA recordings. The stimulus was generated on a PowerBook G4 running Mac OS X 10.3.9, MATLAB 7.4 and Psychtoolbox 3.0.8, presented on a CRT Sony Trinitron MultiScan 420GS screen. It was demagnified using a light path consisting of two mirrors and a total reflection prism mounted on a customized Zeiss Axiovert 40 C inverted microscope, and projected onto the retina plane through a Zeiss EC Plan-Neofluar 5× objective. The stimulus consisted of 10 iterations of 2 s bright, 2 s dark whole field illumination, while synchronized with the MC_Rack recording software using a DAQ USB-1208FS (Measurement Computing Corp., Norton, MA). Retinas were placed RGC side down onto the titanium oxide electrode array (electrodes were 10 μm diameter and placed at 100 μm distance). (b) Whole field stimulus projected on a 60-channel MEA in the retinal plane (the retina covered a surface area of 0.49 mm^2 on the MEA). An arbitrary 5 × 5 region of interest centered on E_{54} is selected to exemplify the MEA response and data analysis in this work. Waveforms in (a) show 1 s raw data recorded at 25 kHz from E_{42}, E_{43}, E_{52} and E_{53}, demonstrating spikes overlaid on strong field potentials.

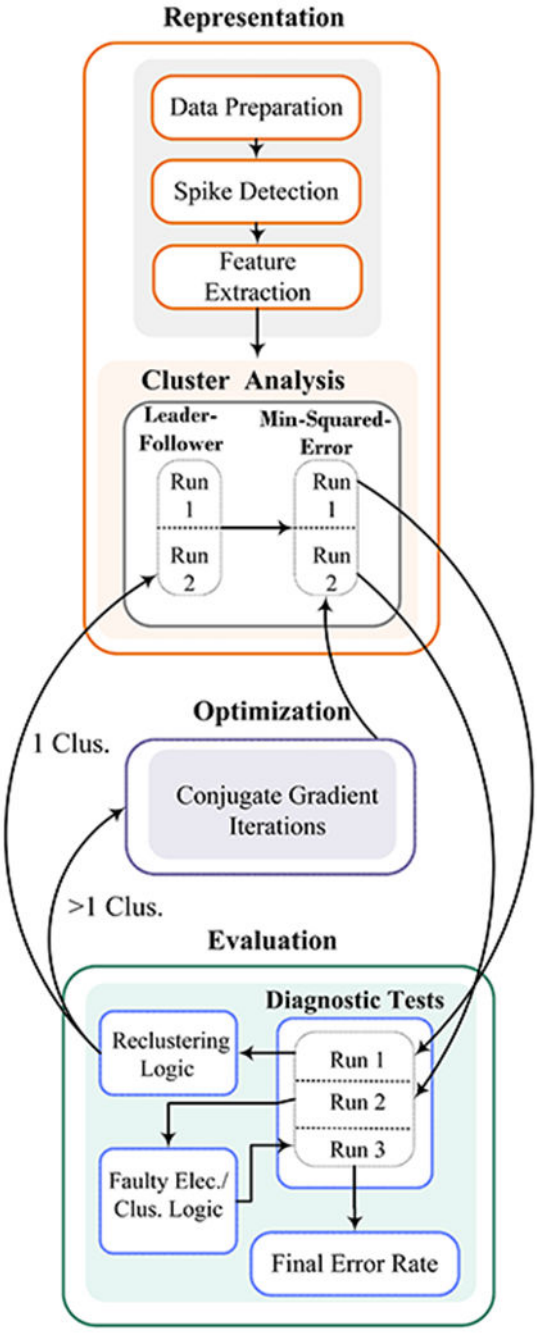

Fig. 2.

Block diagram of the three analytical layers within ClusDiag: representation, optimization and evaluation. For the latter, diagnostic tests reported FP and FN instances over the first run of clustering. The reclustering decision logic called up the conjugate gradient method to minimize the error from the first run of clustering. The second iteration of clustering routines was then applied to the faulty electrodes, and the recovered new clusters were passed on to the second run of diagnostic tests. Finally, the third run of diagnostic tests

served to enhance the clustering resolution for the remaining clusters per electrode. Error assessment quantified all the final FP and FN errors per cluster.

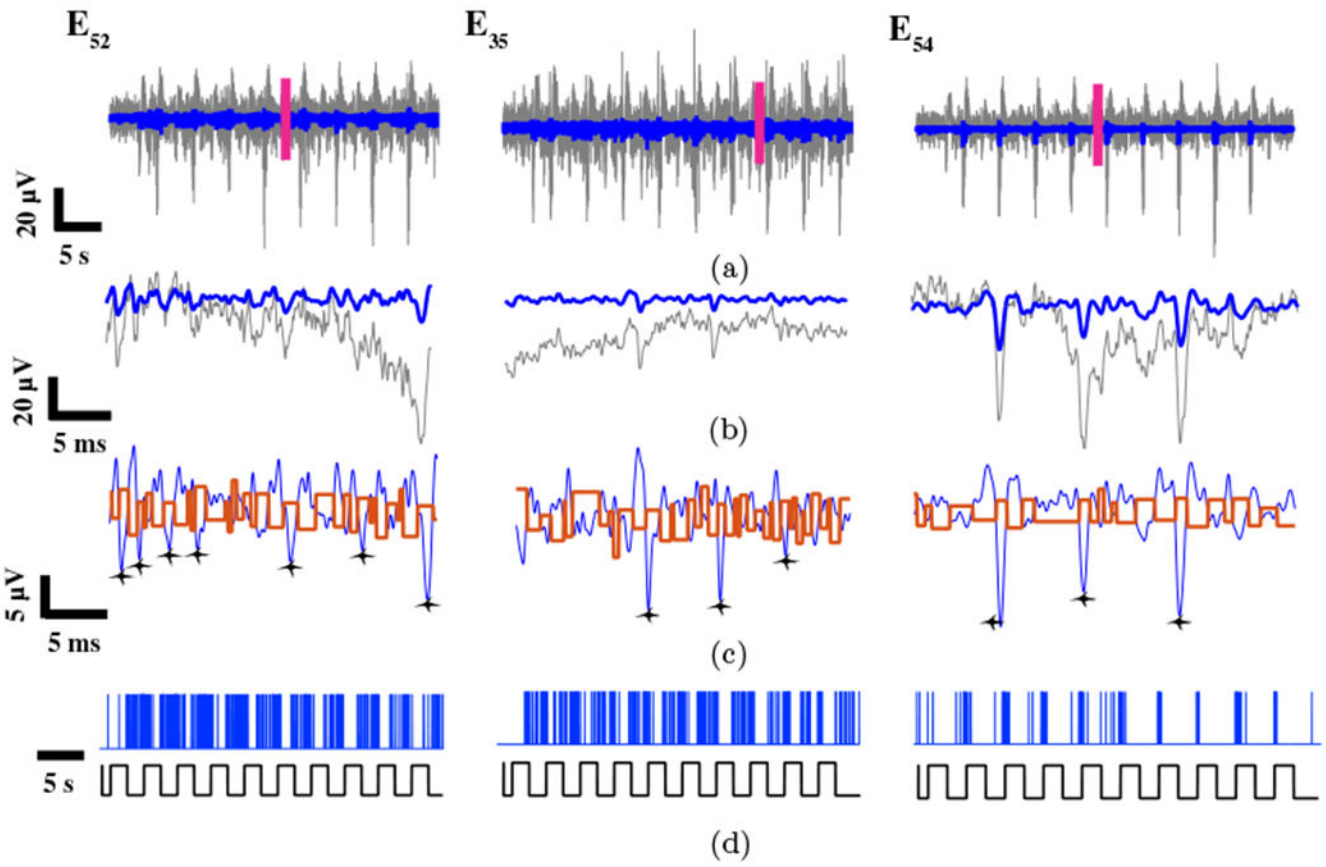

Fig. 3.

(Color online) Detection of spike trains recorded from three electrodes of interest. E_{52}, E_{35} and E_{54} captured the activity of cells with regular spiking and burst firing throughout the recording. (a) Full 42.2 s waveform of the MEA-recorded raw data in gray and the filtered data in blue. (b) Magnified time window of the data in the first row [25 ms, vertical mark in (a)]. (c) Filtered data from (b), overlaid with the adaptive threshold of detection (red). The asterisk marks denote peak amplitudes of detected spikes. (d) Detected spikes train for the whole recording duration in (a). Square wave indicates stimulus periodicity with ON (high) and OFF (low) 2 s periods.

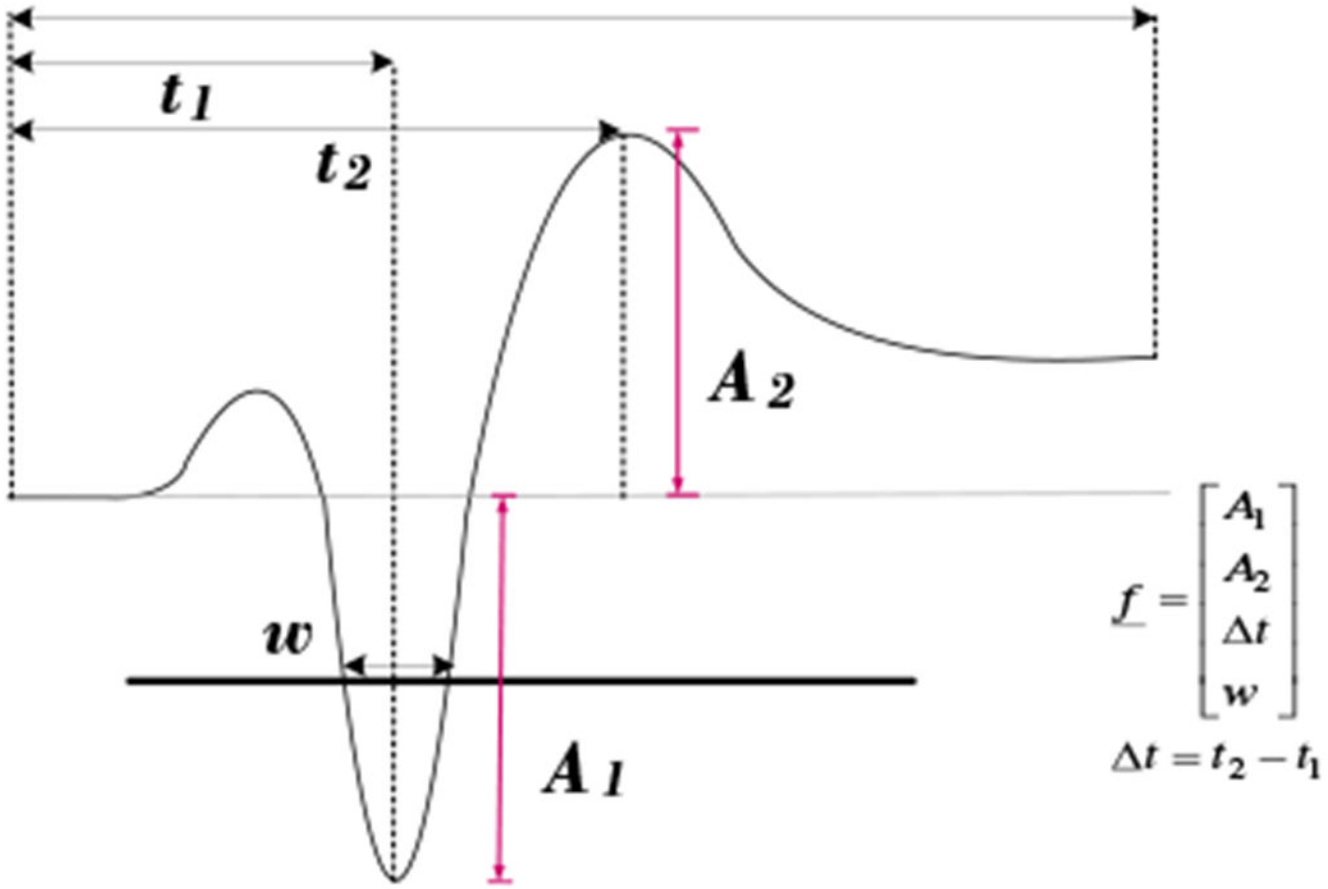

Fig. 4.

Definition of the feature vector from a detected spike and threshold of detection. The feature vector f consists of maximum depolarization amplitude A_1, maximum hyperpolarization amplitude A_2, time difference (Δt) between the time points of A_1 and A_2 and the temporal width of the spike at the level of threshold w.

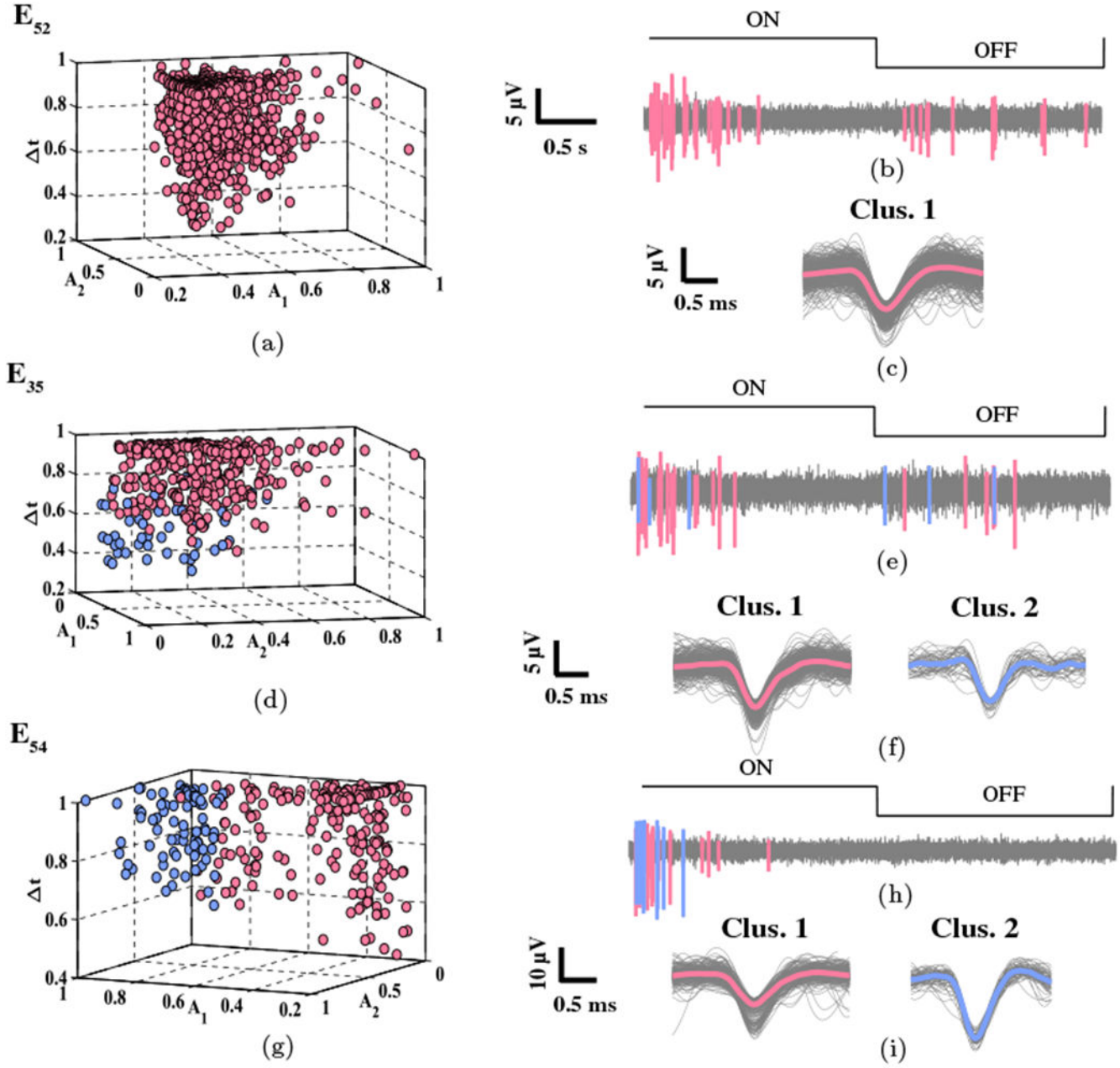

Fig. 5.

(Color online) Examples of sorted spikes from the first run of clustering. (a), (d), (g) Rotated 3D scatterplots for three of the four spike features: depolarization amplitude A_1, hyperpolarization amplitude A_2 and the difference between their time points, Δt. (b), (e), (h) Time domain representations for an ON and OFF stimulation period starting at 33.24 s and ending at 37.26 s. (c), (f), (i) Overlaid spike cutouts (2.5-ms interval) for clusters 1 and 2, together with the derived templates. Clusters, spikes and templates are pseudocolored in red and blue. (a)–(c) The 702 spikes from electrode E_{52} fall in one cluster. (d)–(f) Spikes from electrode E_{35} were sorted into two clusters, with 309 spikes in cluster 1 and 46 spikes in

cluster 2. (g)–(i) Spikes from electrode E_{54} fall in two clusters. Cluster 1 included 210 spikes and cluster 2 comprised 82 spikes.

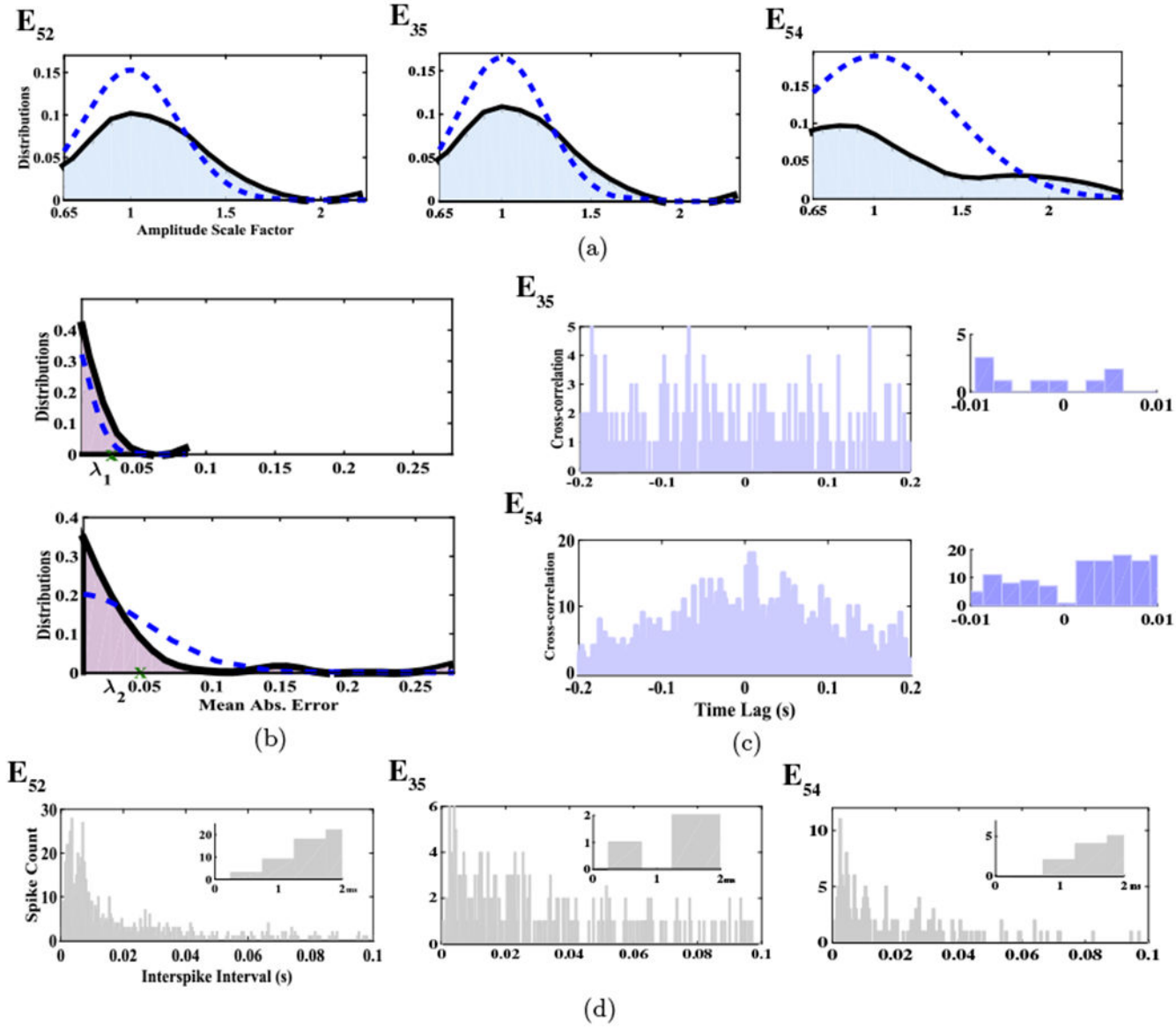

Fig. 6.

Diagnostic tests provide metrics to assess the quality of the first clustering run. The metrics check on FP and FN cases that arose from the clustering. (a) Normalized histogram of the amplitude scale factor in cluster 1 (bin width: 0.1) is represented with the smoothed estimate of distribution function (solid line). A scaled normal distribution fitted to the data (dashed line) is also shown. (b) Normalized histogram of the MAE statistic is represented for cluster 1 (top) and cluster 2 (bottom) across all electrodes (bin width: 0.01). A smoothed estimate of distribution function (solid line) is shown together with a half-normal distribution fit (dashed line). The test threshold was $\lambda_1 = 0.027$ for cluster 1 and $\lambda_2 = 0.046$ for cluster 2, as shown. Comparing the MAE value from cluster 1 on each electrode to λ_1 showed that cluster 1 failed in FP measure on E_{35} and E_{54}. (c) The spike-time cross-correlation function was computed for each pair of clusters per electrode. Shown here are the histogram distributions of spike-time firing between clusters 1 and 2 on E_{35} and E_{54} (bin width: 2 ms). The region

surrounding zero is shown magnified. A gap in vicinity of zero in the smallest bin width means dependency among clusters. Thus, both E_{35} and E_{54} failed, and the initial clusters needed to be reassessed. (d) Cluster contamination with spikes from other cells was also checked using the ISI histogram (bin width: 0.5 ms). The binned violation area is magnified. The ISI violation rates were 5.1% on E_{52}, 1.3% on E_{35} and 3.4% on E_{54}.

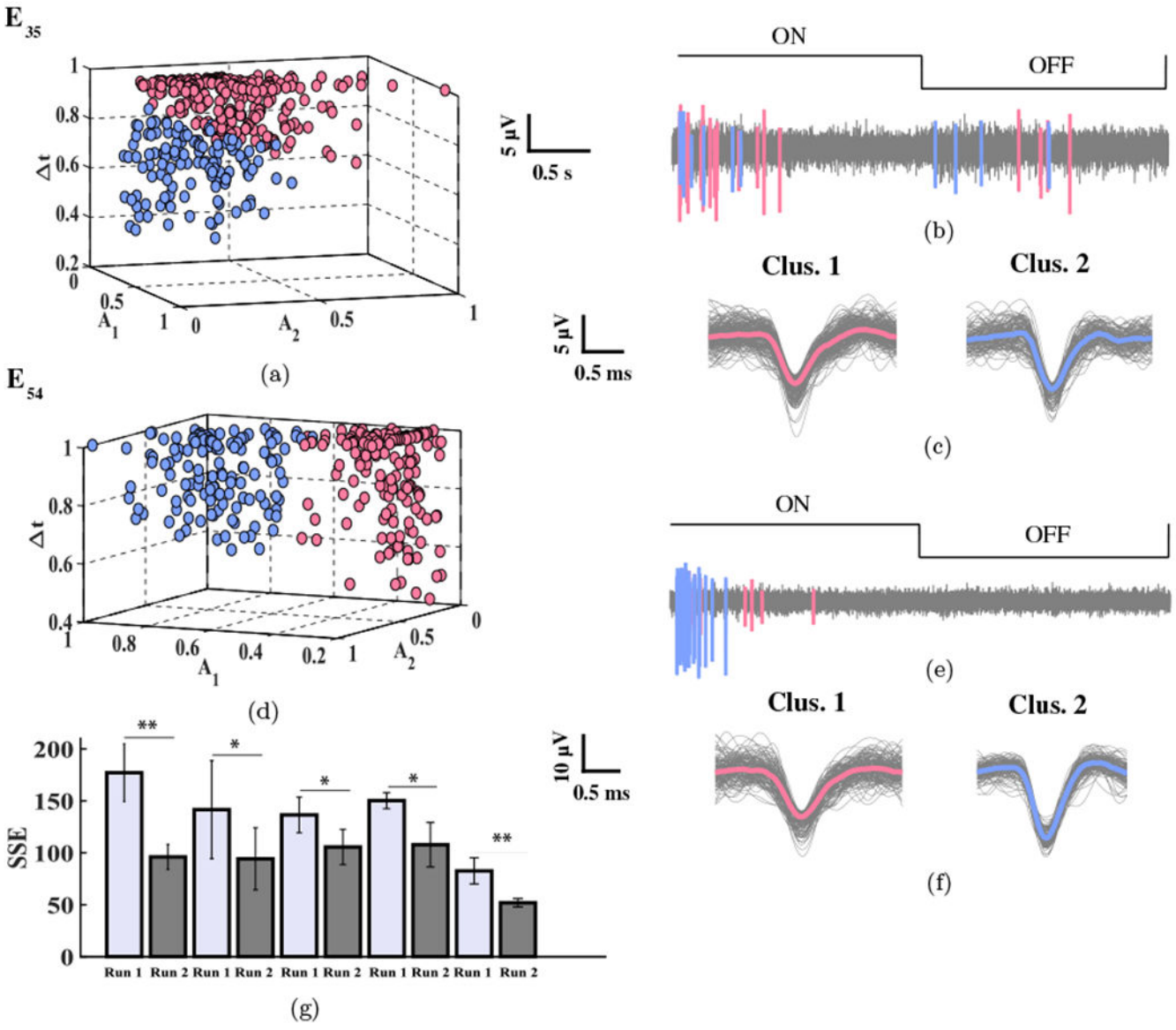

Fig. 7.

Sorted spikes from the electrodes presented in Fig. 5 after second run of clustering. For description of 3D scatter plots [(a) and (d)], time domain representations [(b) and (e)] and spike cutout overlays [(c) and (f)], see Fig. 5. E_{52} was detected by the decision logic to be reassessed by the leader-follower algorithm and reclustered. This reclustering still represented E_{52} via one cluster, so all the illustrations for E_{52} remain as shown in Fig. 5. (a)–(c) The two reassigned clusters of E_{35} now have 226 spikes (cluster 1) and 126 spikes (cluster 2). (d)–(f) Reassigned spikes for E_{54} resulted in 158 spikes in cluster 1 and 129 spikes in cluster 2. Both E_{35} and E_{54} were diagnosed as faulty electrodes through Run 1 of the diagnostic tests and were reclustered to improve each one's SSE value. (g) Comparison test between the SSE values of Run 1 and Run 2 for all diagnosed electrodes for the parametric features and each of the five experiments (WF_1–WF_5). Bars represent the mean value and error bars depict standard deviations. Improvements were all statistically

significant in either 0.5%(**) or 2%(*) level for the null hypothesis. From left to right the SSE improvement was significant for WF_1 ($p < 0.005$, $n = 50$), WF_2 ($p < 0.02$, $n = 37$), WF_3 ($p < 0.02$, $n = 40$), WF_4 ($p < 0.02$, $n = 40$) and WF_5 ($p < 0.005$, $n = 36$), where n is the total number of diagnosed electrodes. Run 2 of diagnostic tests showed that E_{35} and E_{54} were recovered after the reclustering. In comparison with clusters in Fig. 5, the FN spikes in cluster 2 due to cluster 1 are reduced for both electrodes.

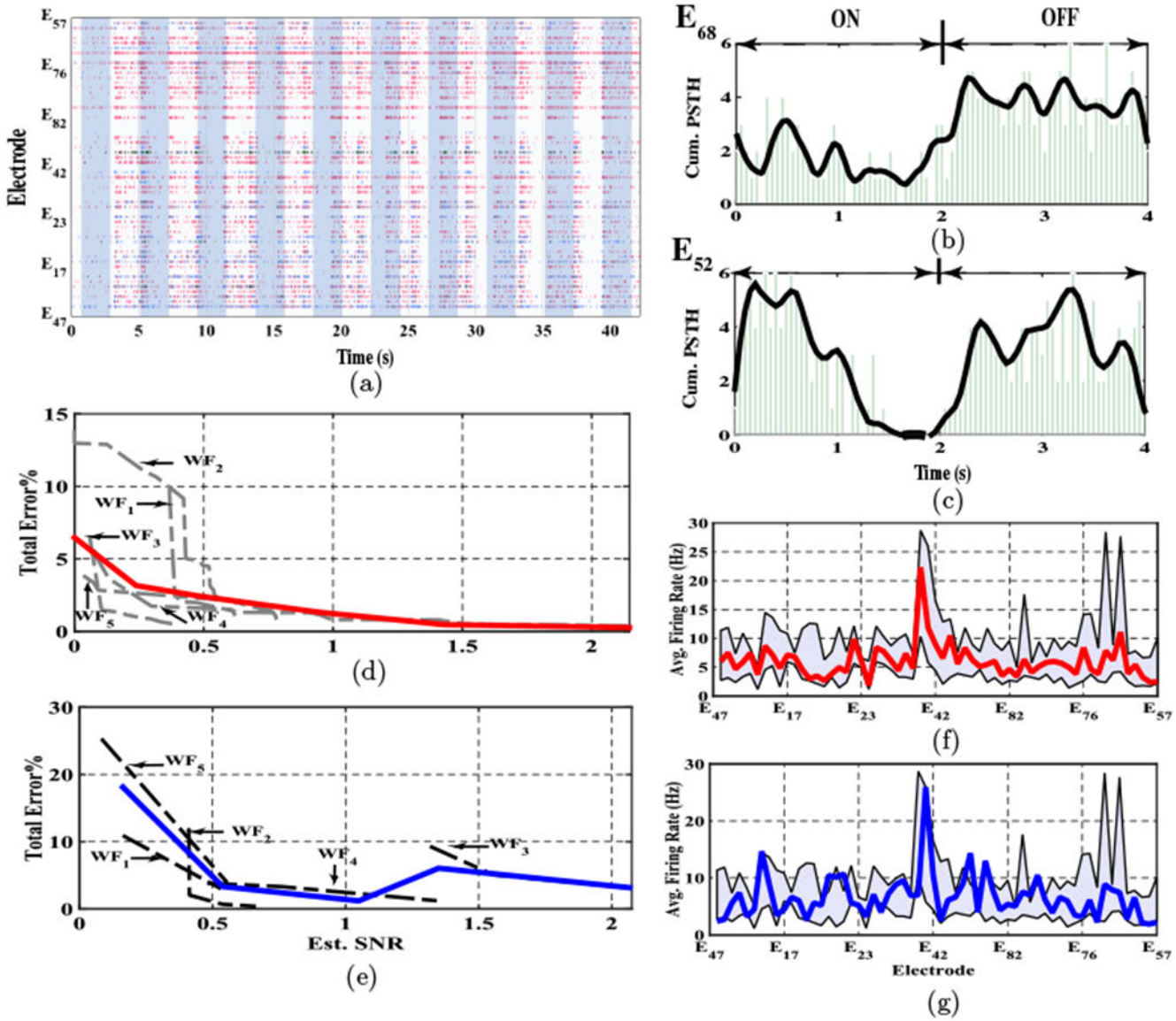

Fig. 8.

(Color online) RGC bursting classification and spike sorting assessment. (a) Spike raster plot for the 72 detected cells in WF_3 dataset using the parametric features. ON periods are shown shaded, and each point indicates firing time of a cell on an electrode (color coded to illustrate the cluster type). (b) The cumulative PSTH is examined to infer the type of the cell based on a ratio index I_I. The cumulative PSTH for an OFF cell recorded by E_{68} is shown (the I_I value was −0.2). An envelope of relative variation of cumulative PSTH is also graphed. (c) The cumulative PSTH for an ON–OFF cell recorded by E_{52} is shown (the I_I value was zero). (d) Total error against the estimated SNR in cluster 1 from the parametric features. The error trajectories for all the five whole field stimuli are shown in dotted gray. The median all over the estimated SNR values is depicted in red. (e) Total error against the estimated SNR from the PCA. The error trajectories for all the stimuli in dashed black. The median represents the error across all the estimated SNR values in blue. (f) The average

firing rates for all the detected cells, calculated for all whole field stimuli using the parametric features. The median of average firing rates of all the cells detected on an electrode is shown in red on the shaded interquartile range. (g) The average firing rates of the cells detected across all whole field stimuli, when using PCA-based features per electrode, with the median of average firing rates of all the detected cells shown in blue.

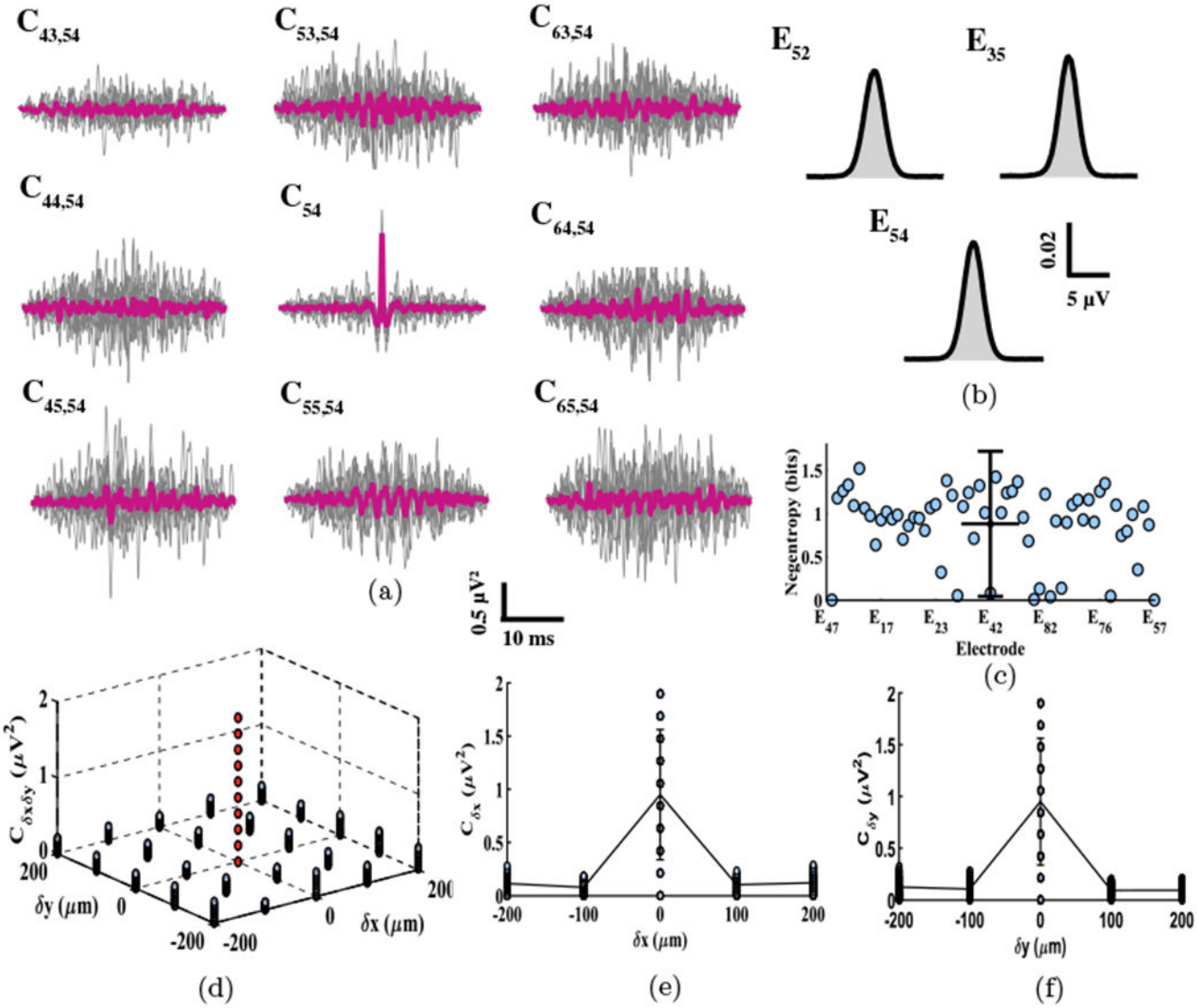

Fig. 9.

(Color online) Estimated noise characterization for the WF_3 dataset. (a) The spatiotemporal correlation function derived within a smaller 3 × 3 neighborhood of E_{54}. We chose 10 equally-spaced 20 ms long intervals across the noise waveform for each electrode, and calculated 10 of such functions. The 20 ms range was sufficient to see the decay of the functions to zero. We took the average of the 10 cross correlation functions (shown in dark red). This process was repeated for varying regions and electrodes of the MEA. The noise characterized a nonstationary stochastic process. (b) Normalized distribution of the estimated noise histogram for the three electrodes. (c) The negentropy criterion was calculated for each electrode to quantify the closeness to the Gaussian distribution. The 95% confidence interval ranged from 0.04 bits to 1.71 bits, indicating fair predictability to a Gaussian. (d) Extrapolated spatial correlation function from 10 samples of the average correlation function in (a). At each electrode location, those 10 samples are shown around

the central E_{54}. (e), (f) The values in (d) are projected onto *x*- or *y*-axis to find the respective horizontal or vertical spread of the noise cross-correlation function (avg. ± s.d. is shown).

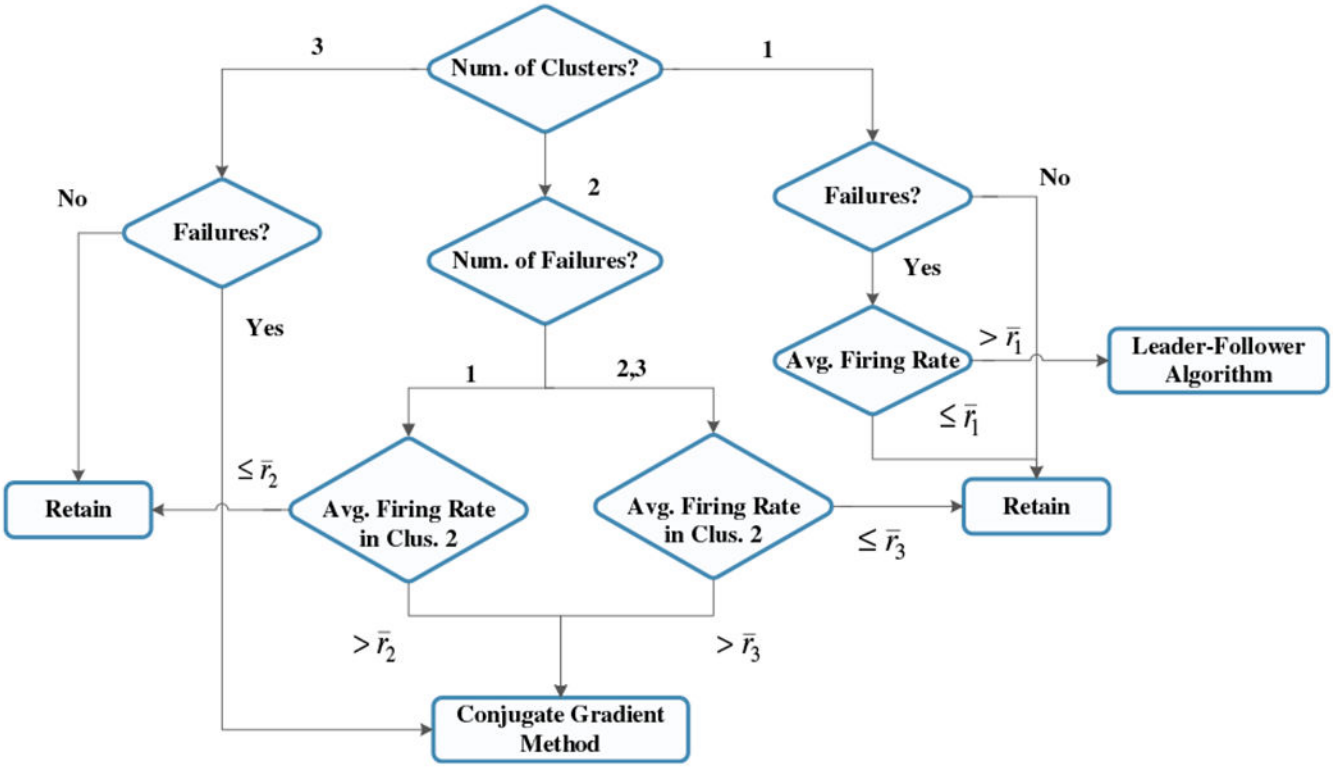

Fig. B.1.
Reclustering decision logic.

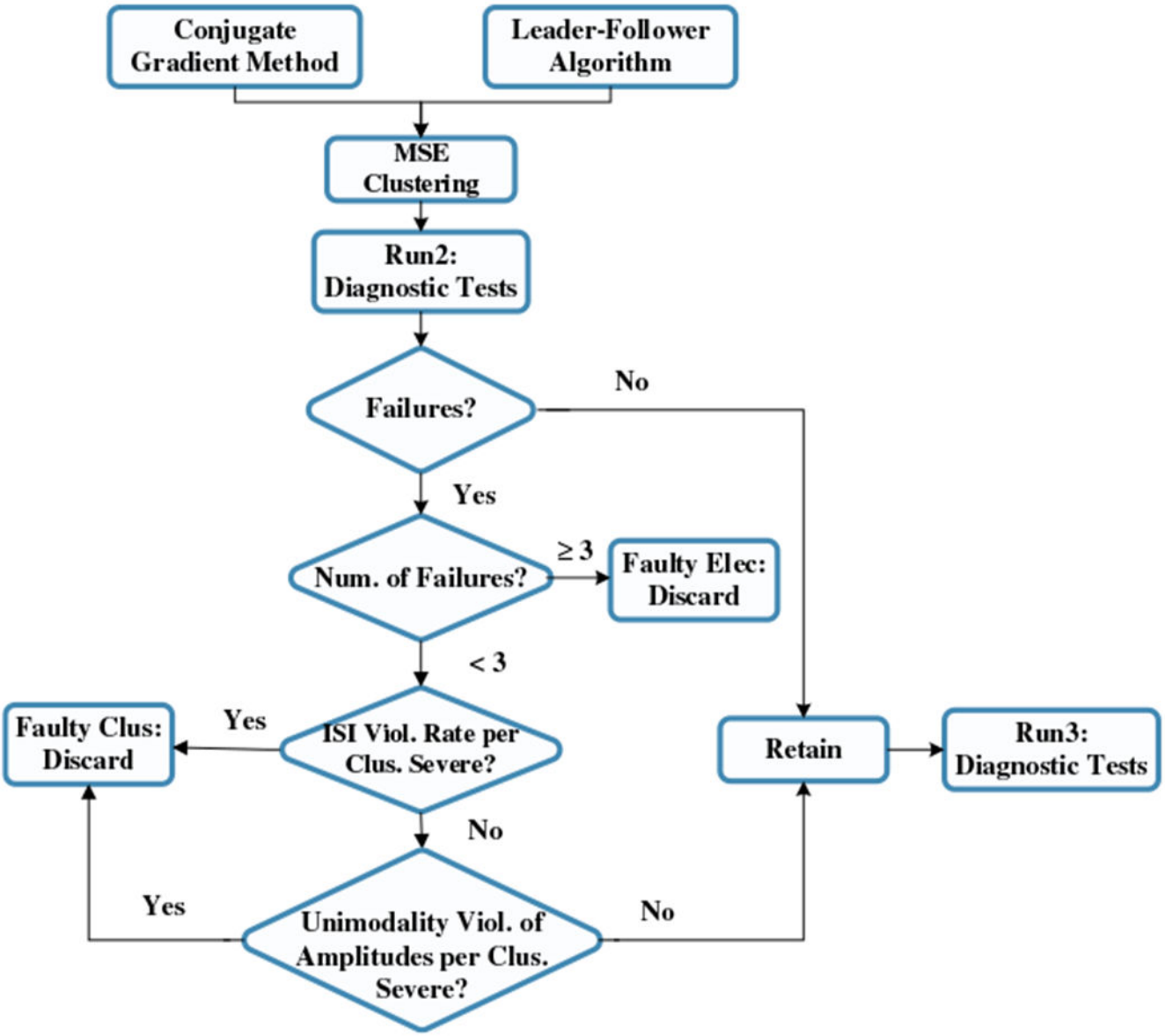

Fig. B.2.
Logic to detect faulty electrodes/clusters.

Table 1.

D-Matrix of diagnostic tests for select electrodes after Run 1. In the T_1–T_7 columns, 1s indicate failure, 0s denote no failure and ×s are not applicable. NC: number of clusters; T_1, T_2, T_3: one-sided alternative hypothesis tests on unimodality for amplitude scale factors of spikes in clusters 1, 2 and 3, respectively; T_4: mutual refractory period incident test between any pair of clusters; T_5, T_6, T_7: ISI violation rate tests for clusters 1, 2 and 3, respectively.

		NC		T_1		T_2		T_3		T_4		T_5		T_6		T_7	
Stim.	**Elec.**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**
WF_1	E_{52}	2	1	1	0	0	×	×	×	1	×	0	0	0	×	×	×
	E_{35}	2	2	1	1	0	0	×	×	0	0	1	0	0	0	×	×
	E_{54}	1	2	0	1	×	0	×	×	×	0	0	1	×	0	×	×
WF_2	E_{52}	2	1	1	1	1	×	×	×	1	×	0	0	0	×	×	×
	E_{35}	2	2	1	0	1	1	×	×	1	1	0	0	0	1	×	×
	E_{54}	3	1	1	0	1	×	0	×	1	×	0	0	0	×	0	×
WF_3	E_{52}	1	3	0	1	×	1	×	0	×	0	1	0	×	0	×	1
	E_{35}	2	1	1	0	0	×	×	×	1	×	0	0	0	×	×	×
	E_{54}	2	2	1	0	0	0	×	×	1	0	1	1	0	0	×	×
WF_4	E_{52}	1	2	0	0	×	0	×	×	×	1	0	0	×	0	×	×
	E_{35}	2	3	0	0	1	1	×	0	1	1	1	0	0	0	×	0
	E_{54}	2	2	0	0	0	0	×	×	0	1	0	0	0	0	×	×
WF_5	E_{52}	2	3	0	1	0	0	×	0	1	1	0	0	0	0	×	1
	E_{35}	1	1	1	0	×	×	×	×	×	×	0	0	×	×	×	×
	E_{54}	1	3	0	0	×	1	×	0	×	0	0	0	×	0	×	0

Table 2.

D-Matrix of diagnostic tests for select electrodes after Run 2; T_1–T_7 as given in Table 1. Here 1s indicate failure, 0s denote no failure and ×s are not applicable.

		NC		T_1		T_2		T_3		T_4		T_5		T_6		T_7	
Stim.	**Elec.**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**
WF_1	E_{52}	2	1	1	0	0	×	×	×	1	×	0	0	0	×	×	×
	E_{35}	2	2	0	1	0	0	×	×	0	0	1	0	0	0	×	×
	E_{54}	1	2	0	1	×	0	×	×	×	0	0	1	×	0	×	×
WF_2	E_{52}	2	1	1	1	1	×	×	×	0	×	0	0	0	×	×	×
	E_{35}	2	2	1	0	0	1	×	×	0	1	0	0	0	1	×	×
	E_{54}	3	1	1	0	0	×	0	×	1	×	0	0	0	×	0	×
WF_3	E_{52}	1	1	0	0	×	×	×	×	×	×	1	0	×	×	×	×
	E_{35}	2	1	0	0	0	×	×	×	0	×	0	0	0	×	×	×
	E_{54}	2	1	1	0	0	×	×	×	0	×	0	0	0	×	×	×
WF_4	E_{52}	1	1	0	0	×	×	×	×	×	×	0	0	×	×	×	×
	E_{35}	2	2	0	0	1	0	×	×	1	0	0	0	0	0	×	×
	E_{54}	2	1	0	0	0	×	×	×	×	×	0	0	×	×	×	×
WF_5	E_{52}	2	1	0	0	0	×	×	×	0	×	0	1	0	×	×	×
	E_{35}	1	1	1	0	×	×	×	×	×	×	0	0	×	×	×	×
	E_{54}	1	2	0	0	×	0	×	×	×	0	0	1	×	0	×	×

Table 3.

Select quantities of results of the cluster analysis and diagnostics. INDS: initial number of detected spikes from all electrodes; INC: initial number of clusters from all electrodes; NERC: Number of electrodes for reclustering via conjugate gradient method; NERL: number of electrodes for reclustering via leader–follower algorithm; NCR: number of clusters after reclustering; NDCR2: number of discarded clusters after Run 2 of diagnostic tests; NDER2: number of discarded electrodes after Run 2 of diagnostic tests; NCR2: number of clusters after Run 2 of diagnostic tests from all electrodes. Each stimulation lasted for 42.2 s.

	INDS		INC		NERC		NERL		NCR		NDCR2		NDER2		NCR2	
Stim.	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**
WF_1	40,071	40,071	110	131	32	31	18	2	102	91	25	30	6	0	63	61
WF_2	20,569	20,569	99	107	23	13	14	0	104	88	14	13	5	1	77	73
WF_3	33,476	33,476	102	125	27	28	13	3	97	91	12	21	2	0	72	70
WF_4	35,138	35,138	92	125	24	23	25	1	99	92	11	11	5	0	76	81
WF_5	20,763	20,763	103	111	25	23	11	6	97	95	16	26	4	0	71	69

Table 4.

Select quantities of results of the cell detection and identification. TNDC: total number of detected cells; NDON: number of detected ON cells; NDOFF: number of detected OFF cells; NDONOFF: number of detected ON–OFF cells.

	TNDC		NDON		NDOFF		NDONOFF	
Stim.	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**
WF_1	63	61	27	24	2	0	34	37
WF_2	77	73	7	11	12	8	58	54
WF_3	72	70	14	16	2	1	56	53
WF_4	76	81	16	19	4	1	56	61
WF_5	71	69	15	13	9	7	47	49

Table 5.

Select quantities of results of the cluster quality diagnostics and error assessment. FNCR3: final number of clusters after Run 3 of diagnostic tests from all electrodes; FNSSR3: final number of sorted spikes after Run 3 of diagnostic tests from all electrodes; AIVC1: average ISI violation rate in cluster 1 over all electrodes; AIVC2: average ISI violation rate in cluster 2 over all electrodes; ASeC1: average sensitivity in cluster 1 over all electrodes; ASpC1: average specificity in cluster 1 over all electrodes; ASR1: average sorting runtime per electrode over first division; ASR2: average sorting runtime per electrode over second division. Each stimulation lasted for 42.2 s.

	FNCR3		FNSSR3		AIVC1		AIVC2		ASeC1		ASpC1		ASR1(s)		ASR2(s)	
Stim.	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**	f_{spike}	**PCA**
WF_1	63	61	21,456	23,689	0.02	0.02	0.02	0.02	0.99	0.93	0.92	0.87	3.27	2.59	1.92	2.34
WF_2	77	73	18,777	19,470	0.01	0.01	0.003	0.002	0.99	0.98	0.98	0.97	1.32	1.23	1.11	1.13
WF_3	72	70	23,735	31,778	0.01	0.03	0.01	0.01	0.99	0.99	0.96	0.96	2.02	2.01	1.88	1.95
WF_4	76	81	31,042	33,979	0.01	0.02	0.01	0.02	0.99	0.99	0.95	0.87	2.21	2.09	1.97	1.99
WF_5	71	69	18,378	16,812	0.01	0.01	0.01	0.02	0.99	0.95	0.95	0.95	1.24	1.26	1.14	1.15