Improving data quality in neuronal population recordings

Kenneth D Harris[1,2], Rodrigo Quian Quiroga[3], Jeremy Freeman[4] & Spencer L Smith[5]

Understanding how the brain operates requires understanding how large sets of neurons function together. Modern recording technology makes it possible to simultaneously record the activity of hundreds of neurons, and technological developments will soon allow recording of thousands or tens of thousands. As with all experimental techniques, these methods are subject to confounds that complicate the interpretation of such recordings, and could lead to erroneous scientific conclusions. Here we discuss methods for assessing and improving the quality of data from these techniques and outline likely future directions in this field.

The more powerful an experimental method, the more care must be taken to ensure its correct application. The two leading methods for measuring the activity of many neurons simultaneously, multichannel electrophysiology and population calcium imaging, have benefited from an exploding range of technical innovations. The increasing complexity of these methods, however, requires increasingly sophisticated approaches to ensure the quality of the data recorded. Quality control is more essential than ever to ensure that the scientific conclusions based on data from these methods are correct and not a result of experimental artifacts.

In this review, we discuss some factors that can affect data quality in neuronal population recordings. Careful experiments, of course, are the foundation of high-quality data. Experimental design involves inevitable tradeoffs, for example, between the number of neurons that can be recorded and the error rates that are acceptable. It is also important that scientists use appropriate data processing techniques to reliably identify individual neurons, and detect when signals are likely to be corrupted. At present, as with many other techniques, the involvement of human operators has an important role in these quality-control procedures and is currently unavoidable. Although effective at catching or correcting data quality problems, input from human operators can also produce bias if not carefully applied.

There can be multiple types of experimental errors in neurophysiology, and their importance depends on the scientific question. False-positive errors, the assignment of spikes to a neuron that did not actually fire them, can lead to invalid conclusions about how information is encoded. False-negative errors, the omission of spikes that a neuron genuinely fired, will have a potentially milder consequence of underestimating firing rate and reliability, but only if the errors occur at random: if the errors are themselves correlated with other factors (such as particular patterns of network activity, bursting or movement), then invalid conclusions could again arise. In population recordings, false-positive and false-negative errors often arise together, resulting from the incorrect assignment of one cell's spikes to another, from the incorrect merging of multiple cells' spikes together or from incorrect splitting of a single neuron's spikes into multiple detected cells. These correlated errors can lead to potentially invalid conclusions about population coding and correlation patterns. Furthermore, selection bias (a systematic failure to detect certain types of cells) can give an incorrect picture of how information is encoded at the population level.

Ultimately, a proper understanding of the limitations in current experimental techniques will only be achieved when sufficient 'ground truth' data have been collected. In practice, 'ground truth' refers to the measurement of neural population activity simultaneously with a method such as on-cell electrophysiology that offers nearly perfect detection of all spikes fired by a single neuron. Such data are presently rare. Nevertheless, the existing ground truth data, together with other approaches such as simulation, have helped the field develop an understanding of the types of confounds that can occur and methods to identify or correct errors. Careful application of these approaches can help to ensure that scientific conclusions based on data from population activity measurements are robust.

Extracellular electrophysiology

Extracellular neuronal recordings are typically performed by inserting microelectrodes, insulated everywhere except one or more small recording sites[1]. The signals from the electrodes are amplified and digitized with a sampling frequency in the range of 20–30 kHz, the rate required to resolve extracellular action potentials waveforms (spikes) lasting on the order of 1–2 ms. In addition to spike waveforms, the extracellular voltage contains a higher amplitude, lower frequency 'local field potential' signal, which is typically separated from the spike signals by filtering and used to provide an indirect measure of ongoing global activity patterns.

An electrode with a single recording site can detect the activity of multiple neurons, but to separate the activity of these cells requires appropriate computational analyses. Typically, spikes are detected as

[1]UCL Institute of Neurology, University College London, London, UK. [2]UCL Department of Neuroscience, Physiology and Pharmacology, University College London, London, UK. [3]Centre for Systems Neuroscience, University of Leicester, Leicester, UK. [4]Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, Virginia, USA. [5]Department of Cell Biology and Physiology, UNC School of Medicine, Chapel Hill, North Carolina, USA. Correspondence should be addressed to K.D.H. (kenneth.harris@ucl.ac.uk) or S.L.S. (slab@email.unc.edu).

Received 20 April; accepted 20 July; published online 26 August 2016; doi:10.1038/nn.4365

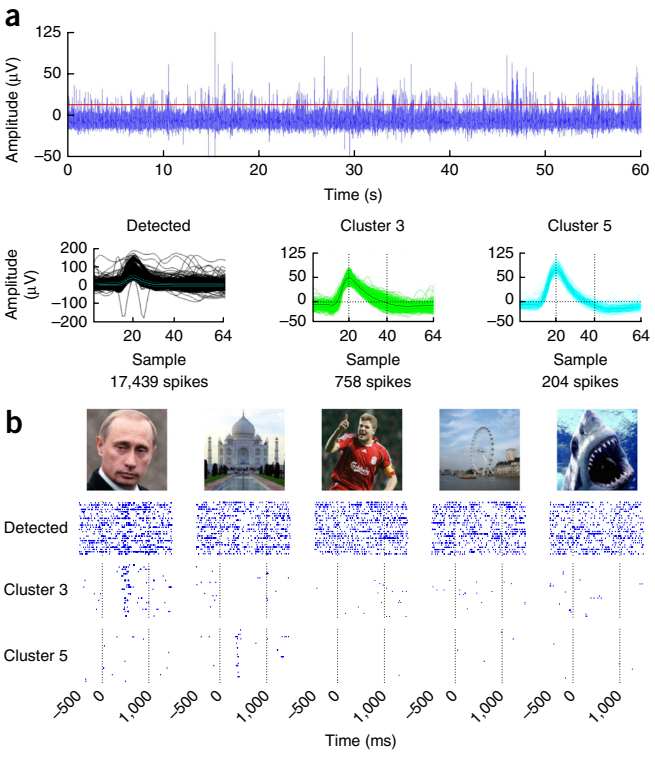

Figure 1 Spike sorting is required to draw valid conclusions in extracellular electrophysiology. (**a**) Top, extracellular recording from a single microwire electrode in the hippocampus of a patient implanted with intracranial electrodes for clinical reasons. Bottom, overlapped all detected spikes (left) and the sorted spikes corresponding to two single units (clusters 3 and 5). (**b**) Responses to five pictures presented in an experimental session. Considering all the detected spikes together, no response can be observed in the raster plots. However, a clear response to Vladimir Putin appears when considering only the spikes corresponding to cluster 3 and a response to the Taj Mahal appears when considering the spikes corresponding to cluster 5. Time zero corresponds to stimulus onset. For space reasons, only 2 of 8 identified clusters and only 5 of the 14 presented pictures are shown, but there were no responses for these other clusters and pictures. Adapted from ref. 9.

crossings of an amplitude threshold, and a waveform is extracted for each spike and temporally realigned to subsample resolution. Because the extracellular spike amplitudes and waveform shapes produced by different neurons at any one point in space can differ[2], the firing of individual neurons can be separated by classifying the waveforms into discrete groups, a process known as spike sorting[3–7]. The peak amplitude of the spike decreases with the distance from the neuron to the recording site, and for neurons located closer than ~50 μm from the electrode tip, the spikes are large enough to be detected over background activity and it is often possible to separate them according to their shapes[1,8]. For neurons further away, up to about 150–200 μm from the electrode tip, spikes can be detected, but the difference in their shapes is masked by the noise and they are grouped together as 'multiunit activity'. Neurons further away cannot be detected and they contribute to the background activity in the recording.

Why is spike sorting necessary? The importance of spike sorting is illustrated by intracranial recordings made in the human brain. Consider an example recording made with an intracranial microwire electrode, implanted for clinical reasons into the hippocampus of an epilepsy patient who viewed pictures presented in random order (**Fig. 1**)[9,10]. If one were to consider all detected spikes, without spike sorting, no obvious increase over baseline firing rate is visible for any of the stimuli. After separating the neurons on the basis of spike shapes, however, a very different conclusion emerges: the spikes in the original signal reflected the mixed activity of multiple neurons, and these neurons were extremely selective to individual pictures: one neuron responded reliably to a picture of Vladimir Putin and another responded to a picture of the Taj Mahal. The spikes from these two neurons represent only about 4% and 1% of the total number of spikes recorded in this electrode, respectively, and this extreme selectivity could not have been detected without spike sorting. In general, the selectivity of single units in the human medial temporal lobe is higher than the selectivity observed for multiunits, where it is not possible to separate the contribution of the different units[11]. Correct spike sorting is therefore essential for understanding the neural code employed by this brain circuit. Without it, one would not only miss very sparse responses, but also make the erroneous conclusion that the circuit employs a 'dense code', in which individual units conveyed information about multiple stimuli, rather than a 'sparse code', in which single cells are exquisitely tuned for particular stimuli.

Electrode design. The physical design of the electrodes used for extracellular electrophysiology makes a great difference to the type of signals recorded. Intracellular recordings represent a gold standard in neurophysiology, offering perfect spike detection and the ability to measure and control membrane potentials, estimate synaptic conductances, control the cell's chemical environment, and stain for later anatomical reconstruction. Nevertheless, the difficulty of intracellular recording severely limits its use for large-scale studies. Electrodes placed directly outside the cell (juxtacellular or on-cell recordings) provide a somewhat easier way to obtain perfect isolation, but are still impractical to use at scale. Extracellular recordings, which rely on detection of electric fields tens of microns away from the recorded cell, are more straightforward, but also involve greater errors, which increase with the number of neurons recorded.

The quality of extracellular unit isolation depends on the amplitudes of extracellular spikes relative to background noise. This noise arises from two sources. The first is thermal noise, which increases with the resistance of the recording electrode (for example, see ref. 12) and can be ameliorated by coating with materials such as PEDOT[13]. The second noise source is the firing of the large number of neurons that are too far from the recording site to produce sortable spikes (>~50 μm), but superimpose to produce 'hash' in the same frequency range[1,8].

Different electrode designs offer complementary advantages. Small electrodes (with diameter < 5–10 μm and impedance >1 MΩ) record from only a few nearby neurons[14], but can show excellent unit isolation, as they can be positioned very close to the neurons of interest; given that spike amplitudes decay rapidly with distance[2,8], this results in large amplitudes relative to background activity. Larger electrodes (diameter several tens of microns) have impedances typically below 1 MΩ and record the activity of neurons in a larger area. For single sites, a smaller ratio of the amplitudes of nearby neurons to those of the distant cells contributing to the background activity leads to a lower signal-to-noise ratio (SNR). The optimal design of an electrode therefore depends on the techniques that are used to process the data: with limited spike sorting algorithms, it might be preferable to have small electrodes with large SNR, whereas with more advanced algorithms and multisite probes, larger, low-impedance electrodes can increase the yield of identified neurons[15,16].

An important and understudied question concerns the degree to which electrode insertion damages the neural tissue that is being recorded. Calculations based on the decay of amplitudes with distance suggest that a single tetrode should be able to detect the activity of up to 100 single neurons in hippocampus, but in reality one typically finds an order of magnitude fewer[1]. Although it is possible that the missing 'dark matter' neurons are healthy, but not firing, or firing but not identified by current spike sorting algorithms[17], insertion of the electrode may also have damaged or killed a substantial fraction of them. Systematic investigation of how electrode geometries, materials and insertion strategies affect tissue damage would greatly help to optimize electrode design.

Multichannel electrophysiology

Understanding complex brain processes requires the analysis of large and simultaneously recorded neuronal populations[1,18–22]. Current multielectrode designs allow recording from hundreds of electrodes, and thus hundreds of neurons simultaneously[4,23–27], and these improvements have been matched by increased capabilities of data acquisition systems.

There are two major approaches to the design of multielectrode arrays. The first approach—exemplified by microwire arrays that have been used for animal studies[28], or the microfabricated 'Utah arrays' that have been implanted in human cortex to enable brain-computer interfaces[29]—consists of a large number of single-contact electrodes with an intersite distance of at least 100 µm. As any individual neuron can be detected by at most one of the recording sites, data processing for such electrodes (and the corresponding quality concerns) is the same as for single-site electrodes. Thus, the recorded population size increases linearly with site count, although without independently movable contacts, many sites may not record neurons. Furthermore, isolation quality does not improve with increasing site count.

The second approach to multielectrode array design is to use dense arrays with an intersite distance of less than 50 µm. This approach employs either twisted microwire bundles such as tetrodes (four microwires twisted together[30,31]) or micro-machined silicon probes[1,32–35]. High-density probes allow the recording of spikes of a single neuron from multiple sites. This improves spike-sorting performance, as the spikes of two neurons can frequently look identical on one channel, but differ on others[31,36]. With this method, unit isolation quality is expected to scale with site density, as confirmed by spatially subsampling of data from dense electrodes[37]. Nevertheless, the achievable number of sites is limited by constraints of manufacturing technology, for example, shaft diameter (probes that are too thick may damage brain tissue), or the increased noise found with small, high-impedance recording sites. Furthermore, the benefits of high-density probes require using spike-sorting algorithms that can combine information from different channels.

Spike-sorting methods. Spike sorting is more complex for dense arrays than for single contact electrodes. For tetrodes, the traditional method of spike sorting, still commonly applied in many labs, is purely manual cluster cutting. In this procedure, a set of features are computed for each spike, such as the peak amplitude on each channel, or waveform features evaluated by principal component analysis. Using a graphical interface, an operator manually draws boundaries around the resulting clusters, which correspond to putative single neurons. The operator is guided in this process by a number of additional tools, such as the computation of auto- and cross-correlograms, which can help to identify poorly isolated units by the presence of refractory period violations.

Today, the most commonly employed spike-sorting method is 'semi-automatic'. With this approach, spike detection and feature extraction proceeds as before, but an automatic cluster analysis algorithm is run on the spike data and a human operator uses a graphical interface similar to those used for manual sorting to check its output and correct mistakes the algorithm has made where necessary (some examples of such possible mistakes are described below). Compared to purely manual spike sorting, the semi-automatic approach has two major advantages. The first is time: it takes substantially less human time to check the output of an automatic algorithm than to perform a fully manual sort. The second advantage is that this approach achieves substantially lower errors than purely manual sorting, as demonstrated using ground-truth tetrode data[36]. These lower error rates occur because the optimal boundary between clusters is a high-dimensional surface, which can be found by automatic algorithms, but not drawn by human hand using a two-dimensional computer interface[36].

The presence of a human operator in the data-processing pipeline raises the potential for subjectivity and bias to occur. However, although a fully automatic spike sorting system would clearly be desirable, it has not yet proved possible to implement algorithms that work robustly in real-world extracellular recordings. Similar constraints are faced in several other fields of data-intensive biology, such as electron-microscopic connectomics, which also rely on manual operator curation[38]. Thus, although substantial developments have occurred to reduce the amount of manual operator time required for the sorting process[37,39], fully automatic systems are rarely applied in current practice. Fortunately, analyses comparing the decisions made by multiple expert operators have shown that their corrections of automatic cluster performance tend to be similar[17,37].

Validation of spike sorting. One of the largest problems for the development of spike sorting methods is the paucity of ground truth data, in which extracellular arrays are combined with other methods to provide unambiguous recording of firing times. Although invertebrate and *in vitro* preparations have provided some data on which to test algorithms[40–42], the noise conditions and non-stationarity found in mammalian systems *in vivo* are substantially more challenging. The difficulty of obtaining ground truth data *in vivo* has made such data very rare[8,36,43] (https://crcns.org/data-sets/hc/hc-1). The data that are available, however, suggest that error rates for semi-automatic clustering with tetrodes can be on the order of 5–10%, but the error rates of purely manual cluster cutting may be substantially higher[36].

In the absence of suitable ground truth data, a number of methods have been used to generate 'surrogate' ground truth to validate spike-sorting algorithms. One approach consists of performing detailed biophysical simulations of extracellular activity[44–46]. The computational expense of this method, combined with uncertainty in exactly how to model the challenges of real data, has encouraged other authors to try a different method. In this hybrid approach, spikes isolated from one recording are digitally added at known times to a second recording or to a simulation of background activity[16,47,48]. This approach has been used to estimate the errors expected from semi-automatic analysis of high-count array data, again yielding an estimate of errors on the order of 5–10%[37]. An alternative approach relies on measuring the reliability of the spike-sorting algorithm under perturbations of the data set[49].

Most current methods of spike sorting fail at times of high neuronal synchrony. In the hippocampus, for example, transient events of highly synchronous neural activity known as sharp waves are of great interest because of their proposed role in memory. Ground truth data suggest that error rates during sharp waves can be fivefold higher than average, reaching levels as high as 50%[36]. Such errors could come

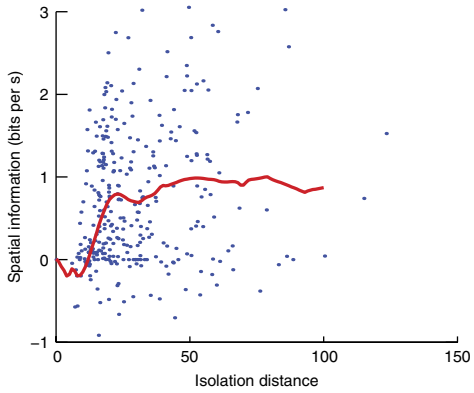

Figure 2 Quantitative measures of unit isolation in extracellular electrophysiology. Each point represents a single neuron recorded in CA1 of a rat exploring an environment, showing isolation distance (a measure of unit isolation quality[54]) versus estimated spatial information content (which can be negative, as it is computed by cross-validation[142]). The red curve represents a running median. The curve reaches an asymptote of ~1 bit per s for values of isolation distance greater than ~20, indicating that this is the true average for well-isolated cells. Data were reanalyzed from ref. 142, with some points above the top of the *y* axis value truncated for visualization purposes.

from two sources. First, synchronous activity might lead to temporally overlapping spikes, which cannot be resolved by traditional spike-sorting algorithms, but might be resolvable by newer algorithms based on template matching[50–53]. Alternatively, sharp waves might lead to the firing of otherwise silent cells with waveforms too similar to be distinguished accurately; consistent with this possibility, sharp waves are accompanied primarily by an increase in false-positive, but not false-negative, errors[36].

Common confounds in extracellular electrophysiology

Isolation quality. In even the best-quality extracellular recordings, most spikes will come from neurons far from the probe, with amplitudes too low for effective isolation. It is therefore important to identify which clusters correspond to well-isolated single cells and which represent mixtures of several neurons. The importance of such metrics is underscored by the fact that different operators, although they generally agree on which corrections to make in semi-automatic clustering, can have very different opinions on what constitutes a well-isolated unit[37].

To address this issue, several quantitative metrics of cluster quality have been proposed. Given that all neurons exhibit an absolute refractory period, any cluster that shows a large fraction of inter-spike intervals (ISIs) less than 1–2 ms cannot be a well-isolated unit. However, the converse does not apply: an apparently clear refractory period does not imply good-quality isolation. Indeed, a cluster that contained the intermixed spikes of two different neurons that fired at separate times (for example, hippocampal neurons with non-overlapping place fields) would have a completely clear refractory period. Furthermore, manual examination of autocorrelograms in the presence of bursting can lead to an erroneous impression of a clean refractory period even for very poorly isolated cells[36].

A second class of quality metrics measures how well the spikes of one cluster are separated from those of neighboring neurons[12,54–56]. It is important to note, however, that there is no single threshold value that objectively defines 'good' isolation quality: the criterion must depend on the scientific application. For example, an analysis of the structure of complex spike bursts, which involves a progressive decrease of amplitude as the burst continues (for example, see refs. 57,58), required a highly stringent criterion[56]. Accurate isolation is also critical for studying pairwise correlations of spike trains[59–61]; for example, if the spikes of a single neuron are artificially divided into two clusters, these clusters will show a spurious negative correlation because their spikes will be always separated in time. Estimation of neuronal tuning and selectivity can be also highly sensitive to clustering errors, as demonstrated by the earlier example of human intracranial recordings. By contrast, for brain-machine-interface applications, the exact identity of the neuron generating each spike might not be crucial, and it may be advantageous to use the largest possible number of recording sites in an unsupervised way, even if not sorted at all[62,63]. Spike sorting may also be less critical where there is a topographic organization of responses—that is, when nearby cells tend to fire to similar stimuli—compared with cases when nearby neurons fire to unrelated stimuli, as has been described in the rodent[64] and the human hippocampus[9] (**Fig. 1**).

When using an isolation quality metric, how should a scientist decide what threshold value of isolation quality to require for a particular scientific question? A simple method is to consider how the quantity being measured depends on isolation quality. For example, the cross-validated spatial information encoded by putative pyramidal cells of rat CA1 drops substantially for values of isolation distance[54] less than 20, but reaches an asymptote above this value (**Fig. 2**). This suggests that a threshold of 20 is suitable for analysis of spatial information coding in these cells.

Selection bias. Simultaneous intra- and extracellular recordings suggest that there should be approximately 140 single neurons in the radius recordable by a single tetrode in the hippocampus[1,8]. However, actual tetrode recordings rarely detect more than a dozen neurons at a time. The reason for this disagreement has been attributed to the presence of silent neurons[65], electrical insulation[66], damage produced by electrode insertion[67] or a potential inability of spike sorting algorithms to deal with large numbers of neurons[17].

Both manual and automatic spike-sorting methods are likely to be biased against low-rate neurons: if a cell fires only a few spikes, these will not be sufficient to define a cluster and the cell will be missed. Given the preponderance of low-rate cells in brain circuits, a failure to account for this selection bias could lead to incorrect estimation of the firing distribution[68]. In neocortex, where superficial-layer pyramidal cells fire with lower rates than deep-layer pyramidal cells and fast-spiking interneurons of all layers[69–72], bias toward high-rate neurons has led most population electrophysiology to focus on deep cortical layers.

Historically, single-neuron recordings have been performed by advancing the electrodes until neural activity is detected[73]. This method can lead to a different form of selection bias: not only will recordings be made primarily from high-rate neurons (perhaps again leading to a bias toward fast-spiking interneurons), but also from cells responding to the specific stimuli or conditions present at the time of the recording. Without care, this could lead to a "confirmation bias": an investigator would find an over-abundance of neurons that respond precisely to the stimulus or condition being investigated.

Sampling bias in extracellular electrophysiology can be ameliorated by performing non-stop chronic recordings using fixed electrodes over very long time periods: recording continuously for days or weeks can lead to sufficient spike numbers to define clusters even for cells of very low firing rate[39].

Operator bias. Because spike sorting has a manual curation step, the possibility of subconscious operator biases must be very carefully excluded.

As with many other procedures, when comparing recordings of subjects in different conditions, it is essential that the operator performs manual curation blinded to the condition of each recording. This is particularly important when analyzing quantities such as stability of firing patterns, which can be easily altered by the manual curation step[74].

Another important concern involves the use of the neuron's firing correlates (for example, sensory receptive fields or place fields) in the spike-sorting process. Although observing a similar firing pattern in two clusters does make it more likely that they represent the same cell, the use of receptive field information during clustering may bias results. Indeed, if neurons genuinely show receptive field plasticity, this will be underestimated if response stability is taken as a criterion for good isolation.

Electrode and waveform drift. Errors in spike sorting are of two types. First, the spikes of different cells can be erroneously merged together. Second, the spikes of a single cell can be erroneously separated into two or more clusters (overclustering), which often occurs when the spike waveforms of a particular neuron vary during the course of the experiment.

Waveform variability can occur for multiple reasons. The most common reason is 'electrode drift', the physical movement of the electrode relative to the brain. Given the highly localized electric fields neurons produce[2,8,16], even a few microns' movement is enough to cause substantial variability in spike amplitudes. This variability is largest for high-amplitude spikes, consistent with the sharply peaked structure of extracellular electric fields. This problem is particularly severe in acute recordings, in which the electrode is fixed not to the skull, but to an external manipulator, meaning that small movements of the head, or relaxation of the brain after compression caused by probe insertion, will cause a movement of the cells relative to the electrodes. Physical electrode drifts are less serious in chronic recordings, where stable recordings have been observed for periods of days or weeks[39,75].

Not all waveform variability is caused by physical movement. Extracellular spike amplitudes decrease during the course of complex-spike bursts and after prolonged firing, even without bursting[56,58]. Moreover, extracellular waveform shapes can depend on cellular factors such as dendritic action potential backpropagation or electrogenesis, which can vary with firing history, inhibition and neuromodulation[76,77]. The difficulty of quantitatively modeling these phenomena is one reason fully automatic spike sorting has so far proved challenging; nevertheless, the fact that these effects can be caught during manual curation suggests that automatic systems may also be eventually possible.

Outlook

Although today's silicon probes have a few hundred channels at most, probes with thousands of channels are currently under development. These probes will raise new challenges for data processing and quality control, the most important concerning manual spike sorting. Purely manual sorting is clearly impossible for this size of data, and even curation of semi-automatic sorting will become a serious burden. This burden can be markedly reduced by the development of algorithms that minimize operator time by directing attention to only those decisions that cannot be made automatically. Fully automatic spike sorting not only becomes more desirable with very high count probes, but it may also become more achievable. Electrode drift and the consequent spike shape changes present one of the biggest barriers to fully automatic sorting; it is possible that large dense probes might sample the extracellular electric fields with enough spatial resolution to allow drift to be tracked and corrected in software.

A second challenge for spike sorting involves the long-term tracking of clusters to study plasticity, for example, during learning experiments. With a few exceptions, scientists have been hesitant to use extracellular electrophysiology to study long-term plasticity of firing properties, as clusters may appear, disappear, merge or separate[78–80], and it is critical not to confuse changes in tuning of neuronal populations with changes in the recording conditions or small electrode movements. Three techniques may ameliorate these problems. The first is the gradual refinement of chronic recording methods, which can now ensure high stability of many cells over multiple days or even weeks of recordings[39,75,81]. The second is the development of quantitative methods for unbiased assessment of cluster similarity that may help to identify neurons across multiple days[78,82–84]. The third is the use of 24-h recording[39], which reduces the problem to a much easier one of tracking slow continuous changes, rather than tracking across sudden jumps between recording sessions.

Finally, and most importantly, there is a need for research into fundamental questions underpinning extracellular array recording. What electrode geometries, surface contact diameters, impedances and materials provide the best data quality while avoiding tissue damage? How do these properties interact with the choice of sorting algorithms? The most critical experiments to solve this problem involve collection of ground truth data to quantify the performance of different electrodes and sorting methods without relying on simulations.

Population recording via calcium imaging

Calcium imaging is a complementary technique for measuring the activity of neuronal populations. Depolarization during action potentials opens voltage-gated Ca^{2+} channels and results in a transient increase in intracellular [Ca^{2+}], which can be detected optically using fluorescent reporters: calcium-sensitive dyes or proteins. Calcium imaging can be used to infer patterns of spiking activity across hundreds to thousands of identified cells *in vivo*[85,86]. Like any measurement, however, it demands careful application and analysis.

Calcium signals are correlated with neuronal spiking, but are an indirect reflection of it, and biophysical variations make the precise relationship between calcium signals and spiking variable[87]. Calcium reporters can themselves limit the precision of spike inference; although synthetic calcium dyes can be used in a linear regime, they still exhibit nonlinear features including saturation, and genetically encoded calcium indicator (GECI) proteins are highly nonlinear as a result of cooperative Ca^{2+} binding[88].

The temporal resolution of the calcium signal is limited. Indicator kinetics are relatively slow (for example, rise times of ~10 ms for single action potentials measured with synthetic dyes, and >50 ms for many GECIs). Furthermore, imaging methods that scan over space trade off recordable population size against sampling time. Recordings from large neuronal populations often require time steps on the order of tens to hundreds of milliseconds[86].

As with electrophysiology, the ultimate check on recording quality is ground truth, typically obtained by simultaneous on-cell patch recordings. Such recordings are feasible with some, but not all, calcium-imaging instrumentation. However, even in the absence of ground truth, good experimental design and rigorous analysis can improve data quality. Below, we discuss these considerations. Our discussion primarily relates to two-photon laser scanning microscopy, but several points are relevant to one-photon imaging, including wide-field imaging through GRIN lenses[89] and light-sheet imaging in transparent specimens[90].

Experimental design and measurement noise. As with any technique, instrumentation, preparation and recording parameters must

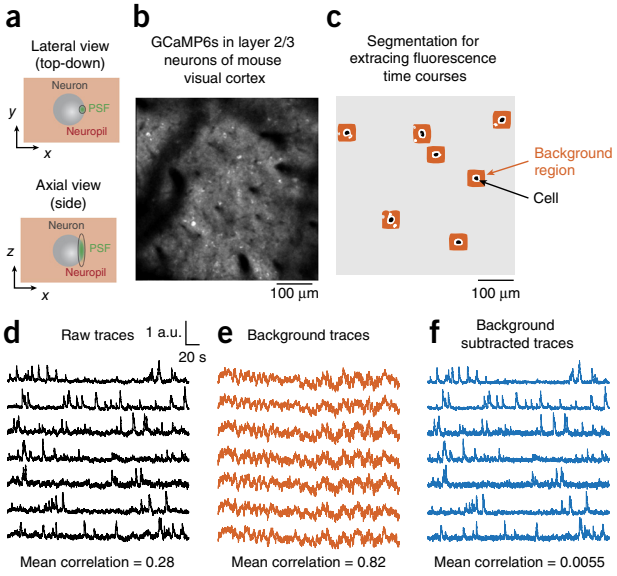

Figure 3 Subtracting neuropil contamination from raw fluorescence time courses. (**a**) In two-photon imaging, the point spread function (PSF) is elongated in the axial dimension even in high numerical aperture systems. Pixels in the borders of cell bodies still contain signals from the surrounding neuropil. (**b**) The GECI GCaMP6s was expressed in mouse visual cortex neurons, resulting in brightly labeled cell bodies and neuropil. (**c**) Binary masks for cell body ROIs (black) were identified semi-automatically and neuropil regions were algorithmically constructed (avoiding pixels belonging to other potential cell bodies or black regions). (**d**) Raw traces for the fluorescence time courses of the selected cells. (**e**) Fluorescence time courses for the background regions for each selected cell. Note the high temporal correlation. (**f**) Fluorescence time courses after background subtraction. Note the reduced mean correlation[141]. All traces have been scaled to the same maximum height to better exhibit details in the time courses. Image courtesy of J.N. Stirman, Y. Yu, S.L.S.

be tailored to the demands of the specific scientific application. Some experiments require the detection of single action potentials and/or the resolution of precise spike counts in each neuron with minimal uncertainty. For others, detecting qualitative increases and decreases in spike rate suffices. Experimental design is critical because most optimizations involve tradeoffs. For example, high zoom (many pixels per neuron) and high frame rate can provide faithful estimates of spiking based on calcium signals[91], but also limit the number of neurons that can be sampled in each imaging frame. Low frame rate acquisition can be sufficient to map population responses when the stimulus changes slower than the acquisition rate: for example, Ohki and colleagues[92] mapped the orientation-tuned responses of hundreds of neurons at 0.61 frames per s, with a stimulus that changed at 0.0625 Hz. By contrast, Dombeck and colleagues[93] acquired 15.6 frames per s over smaller fields of view to capture the activity of hippocampal place fields with subsecond resolution.

Imaging quality depends primarily on recording a sufficient number of photons per pixel. Photon emission exhibits Poisson-like variability (shot noise), with SNR scaling as the square root of the photon count[94]. Optimizing quality therefore requires tuning parameters such as laser power and scan configuration, which includes parameters such as pixel dwell time, field of view and pixels per neuron, which are subject to the limitation that high laser power can cause tissue damage[95]. Optimized scan patterns[96–98] can target sampling to individual neurons, but lack the spatial coverage required for *post hoc* motion correction. Photons per pixel can be estimated from parameters including the gain and offset of the photomultiplier tube (http://labrigger.com/blog/2010/07/30/measuring-the-gain-of-your-imaging-system/). Tradeoffs between photons per pixel, numbers of pixels and SNR can be quantified using signal detection theory and compared and optimized for specific experiments[99–101]. For example, GCaMP6s imaging can provide single action potential detection with nearly 100% detection of all spikes when imaged at a high frame rate over a small field of view (30 µm × 30 µm at 60 frames per s)[91]. However, larger field-of-view, population-level imaging (265 µm × 265 µm at 59.1 frames per s) yields spike rate estimates that correlate with the true spike rate at an average level of ~0.5 (Pearson's *R*, using a 50-ms spike rate binning window)[87]. Quality is also affected by indicator properties and labeling intensity[88,94]. Typically, however, these quantities are not measured for *in vivo* preparations, and parameters yielding high SNR data are identified by trial and error (for example, dye concentration, viral vector titer and number of days after transfection to image).

Signal contamination. In densely labeled tissue, structures adjacent to cell bodies can contribute contaminating signals[102]. In areas such as neocortex, somata are distributed sparsely enough to make contamination from adjacent cell bodies rare[103], permitting moderate-resolution imaging systems to accurately measure cellular-resolution dynamics[92,103,104]. However, in structures with densely packed cell bodies, such as the dentate gyrus of the hippocampus or cerebellar granular layer, cell-to-cell contamination cannot be ignored.

Calcium imaging of large populations typically leads to contamination from signals arising in the neuropil: the axons and dendrites of nearby cells, whose activity produces a substantial contamination of the signal recorded at any cell soma[91,105,106]. Even with two-photon microscopy, axial resolution is often limited to several microns or more, although in-plane (lateral) resolution can be sub-micron (**Fig. 3a**). Furthermore, even high resolution (that is, high numerical aperture) imaging systems are precluded from realizing their full resolution in practice as a result of optical aberrations caused by brain tissue and the loss of marginal rays when imaging deep[107,108]. Imaging system aberrations can also limit resolution, particularly when imaging outside of the very center of the field of view[109,110]. GRIN lenses, used to access deep structures, also suffer from substantial aberrations[111,112]. Adaptive optics can compensate for aberrations of the focused excitation light[113], but the emitted fluorescence photons are still subjected to scattering and aberrations in the tissue and optical systems. Neuropil contamination can be somewhat mitigated by expressing GECIs sparsely, for example, using transgenic mice with *Thy1* promoters, or conditional viral strategies[91,114–116]. However, neuropil contamination represents a serious concern for calcium imaging. The neuropil signal reflects the summed activity of a large number of neurons, the majority of which will typically be located close to the imaged region, and it can therefore be tuned for similar stimuli or behavioral variables[117]. If not accounted for, neuropil contamination can lead to spurious conclusions about neuronal encoding. It is therefore essential that this contamination be understood and, to the extent possible, removed during analysis. Approaches will be discussed below.

Motion artifacts. When imaging in living subjects, heartbeat, breathing and motor behavior can all contribute substantial movement artifacts. Although heartbeat-associated movements in the brain are typically on the order of 1 µm or less[118], breathing and motor behavior can cause larger movements of over 10 µm[119,120]. Movement amplitudes

vary by brain area and can be reduced by appropriate surgical preparation[121]. During movements, cells can change their pixel location in the imaging plane (XY), or in and out of the plane of focus (Z). If temporal resolution is high enough to make frame-to-frame XY movement small, movement can be corrected through image registration, such as global cross-correlation or line-by-line alignment[117,122,123]. Large and faster XY movements may require more complex model-based algorithms[119,124]. Motion in Z is more difficult to correct, as the neighboring planes are typically not recorded. Generally, point spread functions (PSFs) are extended in Z, making motion artifacts resulting from small Z movements less problematic[125,126]. In cases with large Z movement, multiplane imaging[127] and online motion correction[120,128] can help to minimize artifacts.

Data processing. Before calcium imaging data can be used for scientific analysis, it requires preprocessing, typically involving image alignment and motion correction; segmentation to find regions of interest (ROIs) corresponding to imaged cells; time course extraction; neuropil compensation; and, optionally, spike train estimation. In some analysis algorithms, multiple steps are performed in concert.

Segmentation and time course extraction. The simplest way to define ROIs corresponding to individual neurons is with binary masks, which can be drawn manually or with varying degrees of automation[85]. With binary masks, neuropil contamination must be estimated and subtracted out in a separate step, and can be estimated using either an average of the neuropil signal surrounding the cell of interest[129] (**Fig. 3b–f**) or by subtracting the first principal component of the contamination[130,131].

More ambitious approaches describe each pixel's calcium signal as a superposition of signals from one or more cell bodies or processes, neuropil and potentially other sources (for example, instrumental noise). This is typically framed as a matrix factorization problem, where the spatial components (neuronal ROIs) and temporal components (neuronal time courses) are simultaneously learned. As with spike sorting, the results of such automatic algorithms must be verified on a case-by-case basis. For example, methods based on principal component analysis and independent component analysis[131,132] can identify negative temporal signals, which cannot represent neuronal time courses; these methods can also produce ROIs corresponding to multiple neurons or dendritic regions, as the algorithm has no prior information about the spatial extent of signal sources. Constrained non-negative matrix factorization, optimized for the particular characteristics of calcium imaging data[133,134] provides promising results even in the presence of signal crosstalk[106], whereas a new combined clustering and factorization method may offer improvements in both accuracy and performance[135]. Given that matrix factorization methods separate temporally distinct patterns of activity, they may fail on neighboring cells that exhibit highly synchronous activity; a supervised learning approach might avoid this shortcoming[96]. Furthermore, any activity-dependent algorithm will be biased toward active cells; labeling cell nuclei with static (not calcium dependent) fluorescent proteins can help to identify cell bodies independent of activity[117]. Similarly, 'dictionary' methods have been developed to detect cells based on average and resting fluorescence activity, thereby finding a large number of inactive cells[136].

Manually verifying the results of automatic algorithms is time consuming, but some simple automatic techniques can catch many artifacts. The zero-lag cross-correlation of ROI pairs can identify contamination of one cell's signal by another. These correlations should usually be close to zero, rarely above 0.5 and values >0.8 typically indicate contamination. Minor contamination can result in smaller increases in correlation values, and so time courses from nearby cells should be checked particularly closely. Correlation resulting from contamination should be more stable in time than correlation resulting from neuronal firing, so rolling correlation analysis (with width substantially above that of a fluorescence transient) can help to identify unusually stable correlations resulting from contamination. Excluding individual highly contaminated pixels will result in less contaminated ROIs. As genuine neuronal correlations can be affected by brain state, this analysis is best performed on data obtained under conditions of relative desynchronization (for example, stimulus-evoked activity rather than spontaneous activity under anesthesia). Finally, correlating the time series of individual pixels to stimulus and behavioral variables can help to diagnose artifacts.

Evaluation of segmentation approaches would benefit from a systematic comparison of algorithms against ground truth. Ground truth for segmentation can be hard to obtain, or even define, but options include hand annotation or coexpressing anatomical markers (for example, fluorescent proteins that are confined to the nucleus) that report the presence or absence of a neuron soma at a location. The Neurofinder challenge provides several ground truth data sets and a web application for comparing algorithm results (http://neurofinder.codeneuro.org/). Ground truth data are also available from Collaborative Research in Computational Neuroscience (https://crcns.org/data-sets/methods/cai-1). Surprisingly, nearly all segmentation algorithms thus far have taken an unsupervised approach: trying to infer neurons directly from data, rather than by training a supervised model on existing annotations, as is common in object recognition, behavioral classification and anatomical segmentation (Apthorpe, N.J. *et al.* Automatic Neuron Detection in Calcium Imaging Data Using Convolutional Networks, arXiv, http://arxiv.org/abs/1606.07372 (2016)). Especially with the availability of annotated data, supervised methods could be a fruitful avenue of exploration.

Spike inference. The calcium signal is only an indirect reflection of spiking. Many analysis approaches aim to derive from the calcium fluorescence time course of each neuron an estimate of firing rates or exact spike times.

The calcium signal can be approximated as a temporally filtered version of the spike train. Although the decay kinetics of this filter are typically slow (half-decay times are hundreds to thousands of milliseconds), its rise kinetics can be fast (times to peak are tens of milliseconds). The simplest approach is therefore deconvolution with an estimated unitary response, such as an instantaneous rise and exponential decay[137,138]. In practice, deconvolution performance is limited by several factors. The relationship between calcium signals and spiking is complex, nonlinear and varies across neurons, especially during spike bursts. With GCaMP6f, for example, spike pairs can exhibit fluorescence transients whose size depends on the inter-spike interval (personal communication, D. DiGregorio and S. Wang). As with segmentation, supervised learning methods using systematic ground truth data are a promising alternative to unsupervised deconvolution[87]. With present technology, estimates of spike times from calcium imaging should always be treated as approximations, although this uncertainty can be propagated through stages of analysis (**Fig. 4**)[133].

Estimation of spike times is not necessary for many scientific questions. When neurons fire sparsely, for example, neuronal responses can be characterized by how the calcium response itself depends on stimulus or behavioral-related factors. The results of such analyses will not be numerically identical to analyses computed from actual

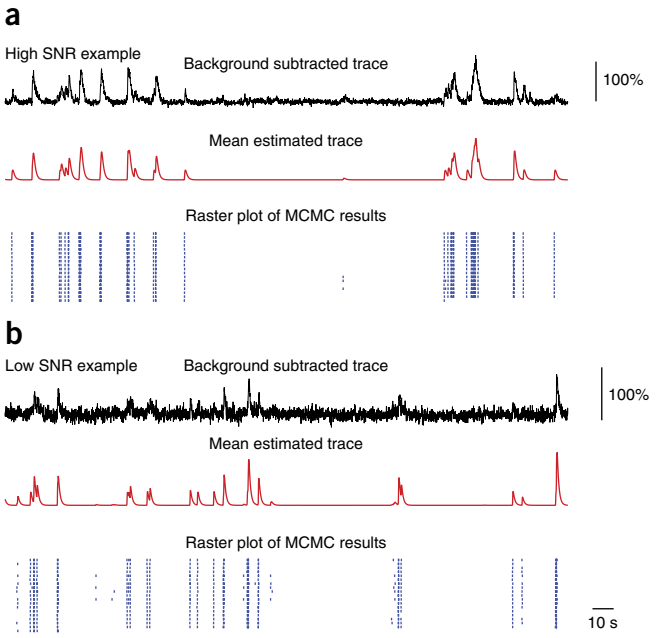

Figure 4 Probabilistic estimation of spike times from calcium signals. A Markov chain Monte Carlo (MCMC) model of calcium signal generation is used to generate a probability distribution of spike trains (blue raster) given a single observed calcium trace (black). (**a**) High signal-to-noise data lead to a highly certain estimate of spiking. (**b**) Low signal-to-noise data lead to an uncertain estimate, illustrated by the raster showing different spike trains that could have generated the calcium signal. Subsequent analysis can be performed using the entire distribution of spike train estimates. Image courtesy of J.N. Stirman, Y. Yu, S.L.S.

counts (for example when computing correlations among neurons), but if interpreted correctly, this can avoid biases introduced by explicit spike estimation. This concept is explicitly formulated in hierarchical models that describe the calcium response as a function of spike times, which are in turn a function of the stimulus and/or behavior. The parameters of such an 'encoding' function can be estimated directly without explicitly estimating spike times, instead treating them as latent factors (for example, see refs. 139,140). Finally, spike detection can be simply viewed as a nonlinear denoising step to remove spurious low-amplitude signals, rather than an explicit estimator of spike times; in this view, procedures based on template matching, which are more flexible and less computationally expensive, may be appealing[117].

Ground truth from electrophysiology. Experiments verifying two-photon calcium imaging with ground truth from simultaneous on-cell electrophysiology have yielded encouraging results. It is important to match the imaging parameters between ground truth data sets and actual experiments: highly zoomed-in imaging offers a best-case scenario about how faithfully an indicator might report spiking, but may vary markedly from those obtained when imaging is zoomed-out to increase the number of neurons imaged (which also decreases the SNR of recorded cells). Single action potentials can be resolved >80% of the time in optimized systems, and multi-spike bursts are even more reliably detected[91,141]. However, estimates of the number of spikes in multi-spike bursts is typically imprecise and, in practice, larger fields of view decrease SNR and lead to overall correlation coefficients between 0.1 and 0.5 (ref. 87), although this has ranged up to 0.8 in some studies[118,141].

In some cases, it is impossible to perform calibration electrophysiology experiments. For example, the small working distance of GRIN lenses largely precludes correlative electrophysiology. Similarly, with air immersion objectives, simultaneous electrophysiology would so perturb the optical setup as to yield it largely irrelevant. In these cases, where ground truth is unobtainable, experiments should be designed to be insensitive to the expected imprecision of spike inference. To bracket the precision of the estimated spike trains, experimenters can compare their data to electrophysiological recordings under similar circumstances. For example, data from extracellular recordings in mouse visual cortex can provide a baseline for calcium imaging studies, setting both the expected spontaneous firing rates and expected distribution of maximal firing rates in mouse visual cortex in response to drifting grating visual stimuli[71], bearing in mind the selection bias in electrophysiology toward active cells.

Summary

Calcium imaging and extracellular electrophysiology can both provide high-fidelity readouts of neuronal population activity. They have complementary advantages: electrophysiology allows detection of single action potentials with submillisecond timing in deep structures and multiple brain areas, whereas calcium imaging can provide a comprehensive and less-biased view of a local population and interfaces easily with the genetic toolkit required to identify neurons by cell type or connectivity.

Both methods are subject to experimental confounds, most notably spike-sorting errors for electrophysiology and signal contamination for calcium imaging. Although these confounds cannot be avoided, they can be mitigated through careful experimental design. Furthermore, a detailed understanding of causes and consequences of these confounds makes erroneous scientific conclusions unlikely if they are used carefully. When objective measures of data quality exist, it is important that these are used and documented together with the scientific conclusions drawn. To gain a truly quantitative understanding of the error rates likely to occur in population recordings, however, it is essential that substantial further effort be put into collecting ground truth data calibrating these techniques against reliable measures of neural activity.

ACKNOWLEDGMENTS

We thank N. Steinmetz and M. Pachitariu for comments on the manuscript. K.D.H. is supported by the Wellcome Trust (95668, 95669, 100154), EPSRC (K015141, I005102), MRC and Simons foundation (SCGB 325512). S.L.S. is supported by the Human Frontier Science Program (CDA00063/2012, RGP0027/2016), National Science Foundation (1450824), Whitehall Foundation, Klingenstein Foundation, McKnight Foundation, Simons Foundation (SCGB 325407SS) and the National Institutes of Health (R01NS091335, R01EY024294). R.Q.Q. is supported by the Human Frontiers Science Program (RGP0015/2013).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Buzsáki, G. Large-scale recording of neuronal ensembles. *Nat. Neurosci.* **7**, 446–451 (2004).
2. Gold, C., Henze, D.A., Koch, C. & Buzsáki, G. On the origin of the extracellular action potential waveform: a modeling study. *J. Neurophysiol.* **95**, 3113–3128 (2006).
3. Quiroga, R.Q. Spike sorting. *Curr. Biol.* **22**, R45–R46 (2012).
4. Berényi, A. *et al.* Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals. *J. Neurophysiol.* **111**, 1132–1149 (2014).
5. Lewicki, M.S. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* **9**, R53–R78 (1998).
6. Rey, H.G., Pedreira, C. & Quian Quiroga, R. Past, present and future of spike sorting techniques. *Brain Res. Bull.* **119**, Pt B, 106–117 (2015).

7. Einevoll, G.T., Franke, F., Hagen, E., Pouzat, C. & Harris, K.D. Towards reliable spike-train recordings from thousands of neurons with multielectrodes. *Curr. Opin. Neurobiol.* **22**, 11–17 (2012).
8. Henze, D.A. *et al.* Intracellular features predicted by extracellular recordings in the hippocampus *in vivo*. *J. Neurophysiol.* **84**, 390–400 (2000).
9. Rey, H.G. *et al.* Single-cell recordings in the human medial temporal lobe. *J. Anat.* **227**, 394–408 (2015).
10. Quiroga, R.Q. Concept cells: the building blocks of declarative memory functions. *Nat. Rev. Neurosci.* **13**, 587–597 (2012).
11. Quiroga, R.Q., Reddy, L., Koch, C. & Fried, I. Decoding visual inputs from multiple neurons in the human temporal lobe. *J. Neurophysiol.* **98**, 1997–2007 (2007).
12. Hill, D.N., Mehta, S.B. & Kleinfeld, D. Quality metrics to accompany spike sorting of extracellular signals. *J. Neurosci.* **31**, 8699–8705 (2011).
13. Ludwig, K.A., Uram, J.D., Yang, J., Martin, D.C. & Kipke, D.R. Chronic neural recordings using silicon microelectrode arrays electrochemically deposited with a poly(3,4-ethylenedioxythiophene) (PEDOT) film. *J. Neural Eng.* **3**, 59–70 (2006).
14. Gerstein, G.L. & Clarke, W.A. Simultaneous studies of firing patterns in several neurons. *Science* **143**, 1325–1327 (1964).
15. Robinson, D.A. The electrical properties of metal microelectrodes. *Proc. IEEE* **56**, 1065–1071 (1968).
16. Camuñas-Mesa, L.A. & Quiroga, R.Q. A detailed and fast model of extracellular recordings. *Neural Comput.* **25**, 1191–1212 (2013).
17. Pedreira, C., Martinez, J., Ison, M.J. & Quian Quiroga, R. How many neurons can we see with current spike sorting algorithms? *J. Neurosci. Methods* **211**, 58–65 (2012).
18. Quian Quiroga, R. & Panzeri, S. *Principles of Neural Coding* (Taylor & Francis/CRC Press, 2013).
19. Buzsáki, G. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* **68**, 362–385 (2010).
20. Brown, E.N., Kass, R.E. & Mitra, P.P. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* **7**, 456–461 (2004).
21. Quian Quiroga, R. & Panzeri, S. Extracting information from neuronal populations: information theory and decoding approaches. *Nat. Rev. Neurosci.* **10**, 173–185 (2009).
22. Harris, K.D. Neural signatures of cell assembly organization. *Nat. Rev. Neurosci.* **6**, 399–407 (2005).
23. Spira, M.E. & Hai, A. Multi-electrode array technologies for neuroscience and cardiology. *Nat. Nanotechnol.* **8**, 83–94 (2013).
24. Lambacher, A. *et al.* Identifying firing mammalian neurons in networks with high-resolution multitransistor array (MTA). *Appl. Phys. A.* **102**, 1–11 (2011).
25. Frey, U., Egert, U., Heer, F., Hafizovic, S. & Hierlemann, A. Microelectronic system for high-resolution mapping of extracellular electric fields applied to brain slices. *Biosens. Bioelectron.* **24**, 2191–2198 (2009).
26. Litke, A. *et al.* What does the eye tell the brain? Development of a system fo the large-scale recording of retinal output activity. *IEEE Trans. Nucl. Sci.* **51**, 1434–1440 (2004).
27. Lopez, C.M. *et al.* A 966-electrode neural probe with 384 configurable channels in 0.13 µm SOI CMOS. *IEEE Int. Solid State Circuits Conf.* 392–393 (2016).
28. Nicolelis, M.A., Ghazanfar, A.A., Faggin, B.M., Votaw, S. & Oliveira, L.M. Reconstructing the engram: simultaneous, multisite, many single neuron recordings. *Neuron* **18**, 529–537 (1997).
29. Hochberg, L.R. *et al.* Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* **442**, 164–171 (2006).
30. McNaughton, B.L., O'Keefe, J. & Barnes, C.A. The stereotrode: a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *J. Neurosci. Methods* **8**, 391–397 (1983).
31. Gray, C.M., Maldonado, P.E., Wilson, M. & McNaughton, B. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J. Neurosci. Methods* **63**, 43–54 (1995).
32. Blanche, T.J., Spacek, M.A., Hetke, J.F. & Swindale, N.V. Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording. *J. Neurophysiol.* **93**, 2987–3000 (2005).
33. Drake, K.L., Wise, K.D., Farraye, J., Anderson, D.J. & BeMent, S.L. Performance of planar multisite microprobes in recording extracellular single-unit intracortical activity. *IEEE Trans. Biomed. Eng.* **35**, 719–732 (1988).
34. Bai, Q. & Wise, K.D. Single-unit neural recording with active microelectrode arrays. *IEEE Trans. Biomed. Eng.* **48**, 911–920 (2001).
35. Csicsvari, J. *et al.* Massively parallel recording of unit and local field potentials with silicon-based electrodes. *J. Neurophysiol.* **90**, 1314–1323 (2003).
36. Harris, K.D., Henze, D.A., Csicsvari, J., Hirase, H. & Buzsáki, G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophysiol.* **84**, 401–414 (2000).
37. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–641 (2016).
38. Briggman, K.L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471**, 183–188 (2011).
39. Dhawale, A.K. *et al.* Automated long-term recording and analysis of neural activity in behaving animals. *bioRxiv* http://dx.doi.org/10.1101/033266 (2015).
40. Anastassiou, C.A., Perin, R., Buzsáki, G., Markram, H. & Koch, C. Cell-type- and activity-dependent extracellular correlates of intracellular spiking. *J. Neurophysiol.* **114**, 608–623 (2015).
41. Wehr, M., Pezaris, J. & Sahani, M. Simultaneous paired intracellular and tetrode recordings for spike sorting algorithms. *Neurocomputing* **26**, 1061–1068 (1999).
42. Delescluse, M. & Pouzat, C. Efficient spike-sorting of multi-state neurons using inter-spike intervals information. *J. Neurosci. Methods* **150**, 16–29 (2006).
43. Neto, J.P. *et al.* Validating silicon polytrodes with paired juxtacellular recordings: method and dataset. *bioRxiv* http://dx.doi.org/10.1101/037937 (2016).
44. Lindén, H. *et al.* LFPy: a tool for biophysical simulation of extracellular potentials generated by detailed model neurons. *Front. Neuroinform.* **7**, 41 (2014).
45. Menne, K., Folkers, A., Malina, T., Maex, R. & Hofmann, U. Test of spike-sorting algorithms on the basis of simulated network data. *Neurocomputing* **44–46**, 1119–1126 (2002).
46. Petersen, K.E. & Einevoll, G.T. Amplitude variability and extracellular low-pass filtering of neuronal spikes. *Biophys. J.* **94**, 784–802 (2008).
47. Martinez, J., Pedreira, C., Ison, M.J. & Quian Quiroga, R. Realistic simulation of extracellular recordings. *J. Neurosci. Methods* **184**, 285–293 (2009).
48. Navajas, J. *et al.* Minimum requirements for accurate and efficient real-time on-chip spike sorting. *J. Neurosci. Methods* **230**, 51–64 (2014).
49. Barnett, A.H., Magland, J.F. & Greengard, L.F. Validation of neural spike sorting algorithms without ground-truth information. *J. Neurosci. Methods* **264**, 65–77 (2016).
50. Franke, F., Natora, M., Boucsein, C., Munk, M.H. & Obermayer, K. An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes. *J. Comput. Neurosci.* **29**, 127–148 (2010).
51. Pillow, J.W., Shlens, J., Chichilnisky, E.J. & Simoncelli, E.P. A model-based spike sorting algorithm for removing correlation artifacts in multi-neuron recordings. *PLoS One* **8**, e62123 (2013).
52. Ekanadham, C., Tranchina, D. & Simoncelli, E.P. A unified framework and method for automatic neural spike identification. *J. Neurosci. Methods* **222**, 47–55 (2014).
53. Pachitariu, M., Steinmetz, N.A., Kadir, S.N., Carandini, M. & Harris, K.D. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv* http://dx.doi.org/10.1101/061481 (2016).
54. Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K. & Redish, A.D. Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* **131**, 1–11 (2005).
55. Pouzat, C., Mazor, O. & Laurent, G. Using noise signature to optimize spike-sorting and to assess neuronal classification quality. *J. Neurosci. Methods* **122**, 43–57 (2002).
56. Harris, K.D., Hirase, H., Leinekugel, X., Henze, D.A. & Buzsáki, G. Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. *Neuron* **32**, 141–149 (2001).
57. Fee, M.S., Mitra, P.P. & Kleinfeld, D. Variability of extracellular spike waveforms of cortical neurons. *J. Neurophysiol.* **76**, 3823–3833 (1996).
58. Quirk, M.C. & Wilson, M.A. Interaction between spike waveform classification and temporal sequence detection. *J. Neurosci. Methods* **94**, 41–52 (1999).
59. Cohen, M.R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**, 811–819 (2011).
60. Pazienti, A. & Grün, S. Robustness of the significance of spike synchrony with respect to sorting errors. *J. Comput. Neurosci.* **21**, 329–342 (2006).
61. Ventura, V. & Gerkin, R.C. Accurately estimating neuronal correlation requires a new spike-sorting paradigm. *Proc. Natl. Acad. Sci. USA* **109**, 7230–7235 (2012).
62. Ganguly, K., Dimitrov, D.F., Wallis, J.D. & Carmena, J.M. Reversible large-scale modification of cortical networks during neuroprosthetic control. *Nat. Neurosci.* **14**, 662–667 (2011).
63. Fraser, G.W., Chase, S.M., Whitford, A. & Schwartz, A.B. Control of a brain-computer interface without spike sorting. *J. Neural Eng.* **6**, 055004 (2009).
64. Redish, A.D. *et al.* Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *J. Neurosci.* **21**, RC134 (2001).
65. Shoham, S., O'Connor, D.H. & Segev, R. How silent is the brain: is there a "dark matter" problem in neuroscience? *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **192**, 777–784 (2006).
66. Moffitt, M.A. & McIntyre, C.C. Model-based analysis of cortical recording with silicon microelectrodes. *Clin. Neurophysiol.* **116**, 2240–2250 (2005).
67. Claverol-Tinture, E. & Nadasdy, Z. Intersection of microwire electrodes with proximal CA1 stratum-pyramidale neurons at insertion for multiunit recordings predicted by a 3-D computer model. *IEEE Trans. Biomed. Eng.* **51**, 2211–2216 (2004).
68. Buzsáki, G. & Mizuseki, K. The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.* **15**, 264–278 (2014).
69. de Kock, C.P., Bruno, R.M., Spors, H. & Sakmann, B. Layer- and cell-type-specific suprathreshold stimulus representation in rat primary somatosensory cortex. *J. Physiol. (Lond.)* **581**, 139–154 (2007).
70. Sakata, S. & Harris, K.D. Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron* **64**, 404–418 (2009).
71. Niell, C.M. & Stryker, M.P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).
72. O'Connor, D.H., Peron, S.P., Huber, D. & Svoboda, K. Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron* **67**, 1048–1061 (2010).
73. Hubel, D.H. Tungsten microelectrode for recording from single units. *Science* **125**, 549–550 (1957).
74. Kentros, C. *et al.* Abolition of long-term stability of new hippocampal place cell maps by NMDA receptor blockade. *Science* **280**, 2121–2126 (1998).

75. Okun, M., Carandini, M. & Harris, K.D. Long term recordings with immobile silicon probes in the mouse cortex. *PLoS One* http://dx.doi.org/10.1371/journal.pone.0151180 (2015).
76. Spruston, N., Schiller, Y., Stuart, G. & Sakmann, B. Activity-dependent action potential invasion and calcium influx into hippocampal CA1 dendrites. *Science* **268**, 297–300 (1995).
77. Buzsáki, G. & Kandel, A. Somadendritic backpropagation of action potentials in cortical pyramidal cells of the awake rat. *J. Neurophysiol.* **79**, 1587–1591 (1998).
78. Tolias, A.S. *et al.* Recording chronically from the same neurons in awake, behaving primates. *J. Neurophysiol.* **98**, 3780–3790 (2007).
79. Swindale, N.V. & Spacek, M.A. Spike sorting for polytrodes: a divide and conquer approach. *Front. Syst. Neurosci.* **8**, 6 (2014).
80. Bar-Hillel, A., Spiro, A. & Stark, E. Spike sorting: Bayesian clustering of non-stationary data. *J. Neurosci. Methods* **157**, 303–316 (2006).
81. Jackson, A. & Fetz, E.E. Compact movable microwire array for long-term chronic unit recording in cerebral cortex of primates. *J. Neurophysiol.* **98**, 3109–3118 (2007).
82. Fee, M.S., Mitra, P.P. & Kleinfeld, D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *J. Neurosci. Methods* **69**, 175–188 (1996).
83. Williams, J.C., Rennaker, R.L. & Kipke, D.R. Stability of chronic multichannel neural recordings: implications for a long-term neural interface. *Neurocomputing* **26–27**, 1069–1076 (1999).
84. Greenberg, P.A. & Wilson, F.A. Functional stability of dorsolateral prefrontal neurons. *J. Neurophysiol.* **92**, 1042–1055 (2004).
85. Ohki, K., Chung, S., Ch'ng, Y.H., Kara, P. & Reid, R.C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).
86. Stosiek, C., Garaschuk, O., Holthoff, K. & Konnerth, A. *In vivo* two-photon calcium imaging of neuronal networks. *Proc. Natl. Acad. Sci. USA* **100**, 7319–7324 (2003).
87. Theis, L. *et al.* Benchmarking spike rate inference in population calcium imaging. *Neuron* **90**, 471–482 (2016).
88. Rose, T., Goltstein, P.M., Portugues, R. & Griesbeck, O. Putting a finishing touch on GECIs. *Front. Mol. Neurosci.* **7**, 88 (2014).
89. Ghosh, K.K. *et al.* Miniaturized integration of a fluorescence microscope. *Nat. Methods* **8**, 871–878 (2011).
90. Keller, P.J., Ahrens, M.B. & Freeman, J. Light-sheet imaging for systems neuroscience. *Nat. Methods* **12**, 27–29 (2015).
91. Chen, T.W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
92. Ohki, K. *et al.* Highly ordered arrangement of single neurons in orientation pinwheels. *Nature* **442**, 925–928 (2006).
93. Dombeck, D.A., Harvey, C.D., Tian, L., Looger, L.L. & Tank, D.W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nat. Neurosci.* **13**, 1433–1440 (2010).
94. Yasuda, R. *et al.* Imaging calcium concentration dynamics in small neuronal compartments. *Sci. STKE* **2004**, pl5 (2004).
95. Podgorski, K. & Ranganathan, G.N. Brain heating induced by near infrared lasers during multi-photon microscopy. *J.Neurophysiol.* http://dx.doi.org/10.1152/jn.00275.2016 (2016).
96. Valmianski, I. *et al.* Automatic identification of fluorescently labeled brain cells for rapid functional imaging. *J. Neurophysiol.* **104**, 1803–1811 (2010).
97. Sadovsky, A.J. *et al.* Heuristically optimal path scanning for high-speed multiphoton circuit imaging. *J. Neurophysiol.* **106**, 1591–1598 (2011).
98. Cotton, R.J., Froudarakis, E., Storer, P., Saggau, P. & Tolias, A.S. Three-dimensional mapping of microcircuit correlation structure. *Front. Neural Circuits* **7**, 151 (2013).
99. Hamel, E.J., Grewe, B.F., Parker, J.G. & Schnitzer, M.J. Cellular level brain imaging in behaving mammals: an engineering approach. *Neuron* **86**, 140–159 (2015).
100. Wilt, B.A., Fitzgerald, J.E. & Schnitzer, M.J. Photon shot noise limits on optical detection of neuronal spikes and estimation of spike timing. *Biophys. J.* **104**, 51–62 (2013).
101. Young, M.D., Field, J.J., Sheetz, K.E., Bartels, R.A. & Squier, J. A pragmatic guide to multiphoton microscope design. *Adv. Opt. Photonics* **7**, 276–378 (2015).
102. Kerr, J.N., Greenberg, D. & Helmchen, F. Imaging input and output of neocortical networks *in vivo*. *Proc. Natl. Acad. Sci. USA* **102**, 14063–14068 (2005).
103. Lecoq, J. *et al.* Visualizing mammalian brain area interactions by dual-axis two-photon calcium imaging. *Nat. Neurosci.* **17**, 1825–1829 (2014).
104. Stirman, J.N., Smith, I.T., Kudenov, M.W. & Smith, S.L. Wide field-of-view, twin-region two-photon imaging across extended cortical networks. *Nat. Biotechnol.* http://dx.doi.org/10.1038/nbt.3594 (2016).
105. Helmchen, F. & Denk, W. Deep tissue two-photon microscopy. *Nat. Methods* **2**, 932–940 (2005).
106. Yang, W. *et al.* Simultaneous multi-plane imaging of neural circuits. *Neuron* **89**, 269–284 (2016).
107. Dunn, A.K., Wallace, V.P., Coleno, M., Berns, M.W. & Tromberg, B.J. Influence of optical properties on two-photon fluorescence imaging in turbid samples. *Appl. Opt.* **39**, 1194–1201 (2000).
108. Tung, C.K. *et al.* Effects of objective numerical apertures on achievable imaging depths in multiphoton microscopy. *Microsc. Res. Tech.* **65**, 308–314 (2004).
109. Zheng, G., Ou, X., Horstmeyer, R. & Yang, C. Characterization of spatially varying aberrations for wide field-of-view microscopy. *Opt. Express* **21**, 15131–15143 (2013).
110. von Diezmann, A., Lee, M.Y., Lew, M.D. & Moerner, W.E. Correcting field-dependent aberrations with nanoscale accuracy in three-dimensional single-molecule localization microscopy. *Optica* **2**, 985–993 (2015).
111. Wang, C. & Ji, N. Characterization and improvement of three-dimensional imaging performance of GRIN-lens-based two-photon fluorescence endomicroscopes with adaptive optics. *Opt. Express* **21**, 27142–27154 (2013).
112. Barretto, R.P., Messerschmidt, B. & Schnitzer, M.J. *In vivo* fluorescence imaging with high-resolution microlenses. *Nat. Methods* **6**, 511–512 (2009).
113. Ji, N., Sato, T.R. & Betzig, E. Characterization and adaptive optical correction of aberrations during in vivo imaging in the mouse cortex. *Proc. Natl. Acad. Sci. USA* **109**, 22–27 (2012).
114. Dana, H. *et al.* Thy1-GCaMP6 transgenic mice for neuronal population imaging in vivo. *PLoS One* **9**, e108697 (2014).
115. Chen, Q. *et al.* Imaging neural activity using Thy1-GCaMP transgenic mice. *Neuron* **76**, 297–308 (2012).
116. Glickfeld, L.L., Andermann, M.L., Bonin, V. & Reid, R.C. Cortico-cortical projections in mouse visual cortex are functionally target specific. *Nat. Neurosci.* **16**, 219–226 (2013).
117. Peron, S.P., Freeman, J., Iyer, V., Guo, C. & Svoboda, K. A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron* **86**, 783–799 (2015).
118. Bonin, V., Histed, M.H., Yurgenson, S. & Reid, R.C. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *J. Neurosci.* **31**, 18506–18521 (2011).
119. Dombeck, D.A., Khabbaz, A.N., Collman, F., Adelman, T.L. & Tank, D.W. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).
120. Chen, J.L., Pfäffli, O.A., Voigt, F.F., Margolis, D.J. & Helmchen, F. Online correction of licking-induced brain motion during two-photon imaging with a tunable lens. *J. Physiol. (Lond.)* **591**, 4689–4698 (2013).
121. Holtmaat, A. *et al.* Long-term, high-resolution imaging in the mouse neocortex through a chronic cranial window. *Nat. Protoc.* **4**, 1128–1144 (2009).
122. Dubbs, A., Guevara, J. & Yuste, R. moco: fast motion correction for calcium imaging. *Front. Neuroinform.* **10**, 6 (2016).
123. Thévenaz, P., Ruttimann, U.E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).
124. Greenberg, D.S. & Kerr, J.N. Automated correction of fast motion artifacts for two-photon imaging of awake animals. *J. Neurosci. Methods* **176**, 1–15 (2009).
125. Wilms, C.D. & Häusser, M. Reading out a spatiotemporal population code by imaging neighbouring parallel fibre axons *in vivo*. *Nat. Commun.* **6**, 6464 (2015).
126. Dufour, P., Piché, M., De Koninck, Y. & McCarthy, N. Two-photon excitation fluorescence microscopy with a high depth of field using an axicon. *Appl. Opt.* **45**, 9246–9252 (2006).
127. Andermann, M.L., Kerlin, A.M. & Reid, R.C. Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing. *Front. Cell. Neurosci.* **4**, 3 (2010).
128. Laffray, S. *et al.* Adaptive movement compensation for *in vivo* imaging of fast cellular dynamics within a moving tissue. *PLoS One* **6**, e19928 (2011).
129. Kerlin, A.M., Andermann, M.L., Berezovskii, V.K. & Reid, R.C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* **67**, 858–871 (2010).
130. Andermann, M.L., Kerlin, A.M., Roumis, D.K., Glickfeld, L.L. & Reid, R.C. Functional specialization of mouse higher visual cortical areas. *Neuron* **72**, 1025–1039 (2011).
131. Schultz, S.R., Kitamura, K., Post-Uiterweer, A., Krupic, J. & Häusser, M. Spatial pattern coding of sensory information by climbing fiber-evoked calcium signals in networks of neighboring cerebellar Purkinje cells. *J. Neurosci.* **29**, 8005–8015 (2009).
132. Mukamel, E.A., Nimmerjahn, A. & Schnitzer, M.J. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* **63**, 747–760 (2009).
133. Pnevmatikakis, E.A. *et al.* Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299 (2016).
134. Andilla, F.D. & Hamprecht, F.A. Sparse space-time deconvolution for calcium image analysis. *Adv. Neural Inf. Process. Syst.* **27**, 64–72 (2014).
135. Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* http://dx.doi.org/10.1101/061507 (2016).
136. Pachitariu, M. *et al.* Extracting regions of interest from biological images with convolutional sparse block coding. *Adv. Neural Inf. Process. Syst.* **24**, 1745–1753 (2013).
137. Vogelstein, J.T. *et al.* Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* **104**, 3691–3704 (2010).
138. Yaksi, E. & Friedrich, R.W. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca^{2+} imaging. *Nat. Methods* **3**, 377–383 (2006).
139. Ganmore, E., Krumin, M., Rossi, L.F., Carandini, M. & Simoncelli, E.P. Direct estimation of firing rates from calcium imaging data. *arXiv* http://arxiv.org/abs/1601.00364 (2016).
140. Harvey, C.D., Coen, P. & Tank, D.W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68 (2012).
141. Smith, S.L. & Häusser, M. Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nat. Neurosci.* **13**, 1144–1149 (2010).
142. Harris, K.D., Csicsvari, J., Hirase, H., Dragoi, G. & Buzsáki, G. Organization of cell assemblies in the hippocampus. *Nature* **424**, 552–556 (2003).