Grids from Bands, or Bands from Grids? An Examination of the Effects of Single
Unit Contamination on Grid Cell Firing Fields

Z. Navratilova\textsuperscript{1,2}, K.B. Godfrey\textsuperscript{1}, and B.L. McNaughton\textsuperscript{2,3}

\textsuperscript{1}Neuro-electronics Research Flanders, KU Leuven, Leuven, Belgium
\textsuperscript{2}Department of Neuroscience, Canadian Centre for Behavioural Neuroscience, The University of Lethbridge, Canada
\textsuperscript{3}University of California at Irvine, Center for the Neurobiology of Learning and Memory Irvine CA, USA.

Correspondence to: Bruce L. McNaughton, The University of Lethbridge, 4401 University Dr W, Lethbridge, AB, T1K 3M4, Canada, email: bruce.mcnaughton@uleth.ca

Abstract

Neural recording technology is improving rapidly, allowing for the detection of spikes from hundreds of cells simultaneously. The limiting step in multi-electrode electrophysiology continues to be single cell isolation. However, this step is crucial to the interpretation of data from putative single neurons. We present here, in simulation, an illustration of possibly erroneous conclusions that may be reached when poorly isolated single cell data are analyzed. Grid cells are neurons recorded in rodents, and bats, which spike in equally spaced locations in a hexagonal pattern. One theory states that grid firing patterns arise from a combination of band firing patterns. However, we show here that summing the grid firing patterns of two poorly resolved neurons can result in spurious band-like patterns. Thus, evidence of neurons spiking in band patterns must undergo extreme scrutiny before it is accepted. Toward this aim, we discuss single cell isolation methods and metrics.

Key words: Spike sorting, Grid cells, Medial entorhinal cortex
Introduction

The hippocampus contains “place cells,” which fire when an animal is moving around in specific parts of an environment (O'Keefe and Dostrovsky, 1971). Several associated regions contain “head-direction cells,” which fire when the animal is facing or moving in a specific direction, regardless of location (Ranck Jr., 1984; Taube et al., 1990). More recently, “grid cells,” which fire in a rhomboidal (aka triangular or hexagonal) lattice (Fyhn et al., 2004; Hafting et al., 2005) over an environment, and “border cells,” which fire near environmental boundaries (Savelli et al., 2008; Solstad et al., 2008) have been discovered in the medial entorhinal cortex, an area that projects directly to the hippocampus. The properties of these neurons suggest that they are involved in highly elaborate computations involved in representing the animal's position in space. Grid cells in particular have been hypothesized to perform a calculation using information about the animal’s movement (proprioception, vestibular information and optic flow) to form and update a unique code for relative position (McNaughton et al., 2006). Several classes of models have been proposed for how this “path integration” is accomplished.

One class of models extends an earlier model for keeping track of head-direction with rotational movement information (Skaggs et al., 1995) to two-dimensional position (Samsonovich and McNaughton, 1997). These models propose that a specific pattern of connectivity of a group of neurons results in a quasi-continuous “attractor,” which stabilizes activity, and is thus able to keep a representation of a variable, such as position, in the absence of external input. To update this representation with rotation (in the case of the head-direction attractor), or translation (in the case of the position attractor), another group of neurons, which contain information about both head-direction (or position) and rotation (or translation) interacts with the attractor network and moves the activity from one “attractor state” to another. This combined network “integrates” translational movement to calculate relative position. Grid-like activity results when periodic boundary conditions are applied to this model.
A second class of models makes use of the oscillatory properties of a major class of neurons in the medial entorhinal cortex, and suggests that if the information about the speed of movement is carried in the frequency of this oscillation, then the interference of this oscillation with another, reference oscillation will give information about relative position (O'Keefe and Burgess, 2005). One major difference in the practical implementation of this second class of models is that it does not allow the representation of a two-dimensional variable, such as position, without an intermediate step. The interference pattern of multiple (2 or more) oscillations can generate a grid pattern such as that exhibited by grid cells, however, the electrical properties of neurons do not allow them to carry multiple oscillations with different frequencies and phases (Remme et al., 2009). This means that each oscillation must be carried by a different neuron, and the summation carried out in stages.

One possible consequence of the interference of oscillations occurring in stages is that neurons that indeed represent one-dimensional position information exist. These neurons would fire in a “band-like” pattern when an animal traversed an environment, so that when the animal ran along one dimension of the environment, the neuron would alternate between rapid firing, and low or no firing, and in the orthogonal dimension the firing rate would remain the same (for a model using such cells see Mhatre et al., 2012). The summation of the inputs from 3 of these “band” cells, each with bands oriented at 60 degrees from one another, would result in the firing pattern already demonstrated by grid cells. Krupic et al. (2012) recently reported that the firing patterns of some medial entorhinal and para-subicular cells resemble those of these hypothesized “band cells.”

Unlike the oscillatory interference model, continuous attractor networks can naturally represent either a 1-dimensional variable or a 2-dimensional variable such as position. To represent a 1-dimensional variable such as head direction, the neural connectivity is arranged so that the neurons form a virtual line (or circle) with symmetrical connectivity along the line, and a 2-dimensional variable is represented with the neural connectivity arranged so that the neurons
form a virtual sheet with connectivity to all neighbors. This sheet is wrapped around both edges to form a torus to allow for the continuous repetition of firing fields shown by grid cells. Thus, the continuous attractor class of models favor representation of position immediately in 2-dimensions, with no intermediate 1-dimensional stage, and did not predict band cells (but is not incompatible with the existence of band cells). The oscillatory interference model would be greatly supported by the existence of band cells, but alternate forms of the model may also be implemented without band cells (Welday et al., 2011).

Thus, Krupic and colleagues’ finding is a potentially significant development in the field, and proof of the existence of band cells would further the understanding of path integration mechanisms. We hypothesized, however, that the appearance of spatial band-like patterns could instead be achieved with the summation of grid firing patterns. Fyhn et al. (2007) showed that MEC grid cells recorded on the same tetrode show the same spacing and orientation of their grids, but different phases (position offsets) of the firing fields. We hypothesized that considering spikes from two grid cells recorded from the same tetrode, with the same spacing and orientation but different offset, as one unit would result in a spatial firing pattern that resembled “spatial bands.” We simulated what would happen when clusters contaminated with spikes of a neighboring grid cell were analyzed in the same manner as in Krupic et al. (2012). Further, we review methods for single cell isolation and determining the contamination level of “single” units, and make some recommendations for analyses that would provide better proof that the firing patterns observed by Krupic et al. constitute well-isolated band cells.

Results

To determine if units contaminated by one or more grid cells could create firing patterns that appear band-like, we simulated firing of grid cells with different relative offsets, and combined them in different ratios. The firing patterns were simulated along a route followed by a rat during foraging for chocolate sprinkles
randomly distributed in a 1 m square box. An example firing pattern of two simulated cells is shown in Figure 1A. The firing peaks of these two cells were offset from each other by 50% of the grid spacing along the horizontal axis (one of the three major axes of the rhomboidal grid), and by 8.8% along the vertical axis (which represents a 10 degree shift from the major axis). The spikes from each cell are combined in a ratio of 2:1, generating a unit that is 33% contaminated from spikes of a second cell. This level of contamination by such a cell generated a firing pattern that showed many peaks, oriented approximately along several parallel lines, resembling bands (Fig. 1B). The autocorrelation of the spatial firing pattern of this “unit” appeared very band-like (Fig. 1C).

We calculated the Fourier spectrograms for the simulated contaminated “units” as in Krupic et al. (2012). This calculation is essentially decomposing the spatial firing pattern into band patterns at different scales and orientations (Krupic et al. figure 1F). Grid cells should show six peaks in this plot, corresponding to the orientation of the three major axes of the hexagonal grid, and the spacing between grids. This indicates that a grid pattern could be conceptualized as the summation of three band patterns of the same scale, oriented at 60 degrees from one another, but of course, says nothing of how the pattern was generated in the brain. For a band-like spatial firing pattern, the Fourier spectrogram would show only two peaks. As shown in Fig.1, however, the summation of two grid cells with the same spacing and orientation but an offset along (or almost along) one major axis will result in the Fourier components corresponding to that axis being stronger than the other two major axes (Fig. 1D).

“Grid scores” for simulated units were calculated as in Sargolini et al. (2006) and Langston et al. (2010). This grid score is the maximum difference between the correlations of a circular sample of the autocorrelogram of the spatial firing plot at 60 and 120 degree rotations, versus 30, 90 and 150 degree rotations (see methods for details). Each method chooses a circular sample in different ways and calculates different values of grid scores, but the scores from
different methods are correlated (Langston et al., 2010). Krupic et al. (2012) used a third method (see Methods) to choose the circular sample. With each method, grid scores for randomized spike trains are also calculated, and any grid score above the 95-percentile score of the randomized spike trains is considered significant. We calculated a distribution of randomized grid scores for our simulated “units” using the Sargolini et al. (2006), peak-based grid score, and the Langston et al. (2010), best radius-based grid score, and found 95-percentile thresholds of 0.13 and 0.43, respectively. The grid score for the example contaminated "cell" in Fig.1 is -0.04 when calculated according to the method in Sargolini et al. (2006) and 0.33 when calculated as in Langston et al. (2010), both of which are below the respective significance level, and thus this “unit” would be considered a spatially periodic non-grid cell by Krupic et al. (2012). For the remainder of the simulations, grid scores calculated with the Sargolini et al. (2006), peak-based method are used, as these were more similar to the calculation Krupic et al. (2012) used, and the distribution of scores for randomized trains was more Gaussian with this measure.

Simulated “unit” contamination levels were varied between 10% and 50% to determine how much contamination is needed for the appearance of bands. An example of a unit increasingly contaminated by a cell with 50% offset is shown in figure 2. Bands appear in the autocorrelogram at a contamination of 25%, and the grid score becomes non-significant at a contamination of 33%. The Fourier spectrogram shows two main peaks at a contamination of 25%, and the other four peaks disappear by 40% contamination (data not shown). Units with the same contamination and offset parameters were simulated twice (with a different Poisson spike train) for each of two paths, and all four simulations at 40% contamination had non-significant grid scores. In one case, 25% contamination was enough for a non-significant grid score.

The relative offset of the primary grid cell and the contaminating grid cell is a major determinant of whether the resulting unit will appear grid-like, or band-
like, or some other pattern. We used contaminating grid cells with several offsets, and observed that offsets of up to 15% resulted in “units” with firing patterns and grid scores indistinguishable from those of clean units, while offsets of 35% or more resulted in lower, sometimes non-significant, grid scores, an appearance of “bandiness,” and decreased hexagonal symmetry in the Fourier transform (Figure 3). The three ‘prototypical’ offsets are little or no offset (lower left corner of Fig. 3A, and 3Bi), which would result in grid patterns and 6 peaks in the Fourier transform, 50% offset along a grid axis (and little or no offset in the orthogonal axis; lower right corner of Fig. 3A and figure 1), which would result in band-like patterns and 2 peaks in the Fourier transform, and 50% along a non-grid axis (30 degree shift from grid axis; upper right corner of Fig. 3A, and 3Biv), which would result in honey-comb like patterns, and 6 peaks in the Fourier transform, occurring at two spatial frequencies (if a large enough portion of the pattern is sampled). Examples of simulations with the most band-like patterns are shown in figures 1 (50% horizontal offset, 8.8% vertical offset) and 2 (45% horizontal offset, 0% vertical). The orientation and periodicity of the grid are the same for all grid cells recorded on the same tetrode (Fyhn et al., 2007; Stensola et al., 2012), and thus were not varied.

Even though many of our simulated “units” appeared band-like in the autocorrelogram, and showed only two peaks in the Fourier transform, passing the criteria for “spatially-periodic non-grid cell,” it was still evident that the spatial firing pattern of these units did not form clean bands. There were multiple peaks in the spatial firing plot, indicating multiple fields, rather than bands. An analysis of the firing rate along each of the bands formed by combining simulated grid cells, revealed repeating peaks of high firing rate (Fig. 4C,D), unlike a simulation of a band cell (Fig. 4E). Thus, we performed a similar analysis of the Krupic et al. data, to determine if their “band-like cells” resembled actual bands, with uniform firing rate along one axis, or a combination of many fields arranged in a band. The 9 best examples of band-like cells, compiled in Krupic et al.’s supplementary figure 10, all contained multiple peaks along each “band” (Fig. 5).
There are few reliable methods for estimating the contamination of a unit isolated from tetrode recordings (see discussion). The only method that does not rely on the same measures that are used for spike sorting is to check the number of spikes that occur within the refractory period of another spike. Hill and colleagues (2011) derived an equation for estimating the number of expected refractory period violations ($r$) of an isolated cluster with a given false positive (contamination) rate ($c$), which is as follows:

$$r = \frac{2(\tau_R - \tau_C)N^2(1-c)c}{T}$$

where $\tau_R$ is the length of the refractory period, $\tau_C$ is the censored period following a spike during which spikes are not detected by the recording system, $N$ is the number of spike events clustered as part of the unit, and $T$ is the total length of the recording during which spikes are detected. From this equation, we derived an estimate of the contamination of a unit:

$$c = \frac{1 - \sqrt{1 - \frac{4p}{F\tau_{2.5ms}}}}{2}$$

where $F$ is the firing rate of the unit ($N/T$), $r$ has been replaced with the proportion of refractory period violations ($p=r/N$), and ($\tau_R - \tau_C$) has been replaced with the typical values used in recordings in our laboratory, ($\tau_R = 2\text{ms}; \tau_C = 0.75\text{ms}$). This shows that the contamination rate of the unit depends not only on the proportion of refractory period violations, but also on the firing rate of the unit. Units with low firing rates are expected to show very few refractory period violations, even at high contamination rates. Laboratories vary widely in how they report cluster isolation quality and refractory period violations (see discussion), but often a threshold of maximum proportion of refractory period violations is used as a criterion, without considering the firing rate of the unit. As can be seen in figure 6A, even the conservative value of 0.2% of refractory period violations actually indicates an unacceptably high contamination rate when used for units with mean firing rates of less than 5Hz. The accepted proportion of refractory period
violations should actually be scaled with the firing rate of the cluster, depending on the desired maximum contamination rate according to the relation:

\[ p = (c - c^2) \times F \times 2.5 \text{ms} \]

which is plotted in figure 6B.

Another criterion used to evaluate refractory period violations is the \( R_{2:10} \) value proposed by Fee et al. (1996). This calculation involves comparing the rate of refractory period violations to the rate of spikes within \( \tau_C - 10 \text{ms} \) of another spike according to the equation:

\[
R_{2:10} = \frac{(10 \text{ms} - \tau_C)}{(\tau_R - \tau_C)} \times \frac{p}{F_{10}}
\]

If we assume that \( F_{10} / (10 \text{ms} - \tau_C) \) is similar to the firing rate of the unit, we can also use \( R_{2:10} \) to estimate the contamination rate as:

\[
c \approx 1 - \sqrt{1 - 4 \times R_{2:10}}
\]

This suggests that for a maximum contamination of 10%, only units with \( R_{2:10} \leq 0.09 \) should be used. Of course, the rate of spikes within 10 ms of another spike is higher than the baseline firing rate of a bursting neuron, and thus \( R_{2:10} \) provides an (often large) overestimate of the contamination rate of bursting cells.

Krupic et al. (2012) do not report the criterion of refractory period violations they used, and thus we cannot comment on the contamination levels of their units. Another important caveat is that the refractory period criterion is based on uncorrelated, Poisson spike trains. However, two grid cells with perfectly offset fields are expected to have a negative correlation. In this case, even the refractory period violation criterion does not aid in determining the isolation quality. This is a general issue for units that may have intrinsic non-zero correlations.
Discussion

Spike sorting methods

While tetrodes provide superior isolation of single units compared to most single channel electrodes, and hence facilitate extracting activity of many single neurons simultaneously, the isolation of single neurons is still difficult and for most units cannot be accomplished perfectly. This method involves detecting action potentials from extracellularly recorded electrical potentials. Band-pass filtering is used to detect signals in the frequency band of action potentials, and, most commonly, an amplitude filter is used to detect these spikes. To identify the spikes from single neurons, the waveform shapes and amplitudes on the four channels of the tetrode are used to classify different neurons (Gray et al., 1995). Different spike sorting algorithms use slightly different measures of the waveforms, but generally, each waveform is decomposed into 2 to 4 features that are most representative of the spike. Usually this is done with principal component analysis of the full waveforms of all recorded spikes on an electrode, and the first and second principal components are used for sorting. The first component usually corresponds to the amplitude of the waveform (Harris et al., 2000; e.g. Lewicki, 1998), which can also be replaced with a measure of the peak amplitude, the peak-to-trough height, or the energy of the waveform (the root mean square of all samples taken along the waveform). The second component corresponds to features of the second half of the spike waveform (Harris et al., 2000; e.g. Lewicki, 1998), as the spikes vary in width and amount of after-hyperpolarization. In algorithms where energy is used instead of the first principal component, the waveforms are energy normalized before PCA analysis, and thus the first PC corresponds to the waveform shape, similar to the second PC described earlier (e.g. Schmitzer-Torbert et al., 2005). Finally, these features of the waveforms are used (either manually, or with an automated algorithm) to sort spikes into clusters that roughly correspond to single neurons (called “units,” because single cell isolation cannot be completely confirmed). Automated algorithms often operate under the assumptions that clusters are Gaussian in shape (but see Fee et al., 1996; Quiroga et al., 2004; Takekawa et al., 2010),
spikes and the underlying “noise” are independent, and background noise is stationary. These assumptions are sometimes violated (in particular the Gaussian clusters assumption, see below), and thus a manual step, in which a user corrects the clusters generated, is often included. Manual correction of automated clustering is easiest if the automated algorithm over-clusters the data, and the experimenter then merges clusters that have similar spike shapes and cross-correlograms indicative of a single neuron.

Two types of problems are encountered when sorting spikes: incorrectly excluding spikes fired by the recorded neuron (false negatives, Type II error), and incorrectly including spikes not fired by the neuron in the cluster (false positives, Type I error). The reasons for false negatives are: exclusion of spikes when background noise in the recording causes the spike amplitude to drop below the detection threshold, exclusion of spikes when multiple recorded neurons fire in close temporal proximity, resulting in a compound spike waveform that is undetectable or unclassifiable (Harris et al., 2000; Lewicki, 1998), and incorrect classification of spikes from one cell to the cluster of another cell because of similarity in waveforms. The analysis of a unit with many false negatives will result in incorrect calculation of firing rate, and receptive fields will appear to have smaller amplitudes (Hill et al., 2011). Reasons for including false positives in a cluster are: the incorrect segregation of two or more cells with similar waveform shapes and amplitudes, and misclassification of composite spikes or noise in the recording as spikes from a single neuron. The analysis of a unit with many false positives also results in an incorrect calculation of firing rates, and, additionally, in an incorrect characterization of the receptive field (Hill et al., 2011), such as what is modeled in this paper.

The likelihood of spike sorting errors increases when there are large numbers of spike overlaps. Spike overlaps increase during population burst events such as sharp-waves (SPWs), or when two of the cells recorded on a tetrode are coupled, and therefore often spike in rapid succession. Additional problems occur when the assumptions of the spike sorting algorithms are violated. For example, spike waveform shapes and amplitudes are known to
change within a burst, violating the assumption that waveforms only vary with Gaussian noise (e.g. Harris et al., 2000). In addition, if the position of a tetrode drifts during a recording, the recorded spikes will change as well. These problems require users to manually adjust the results from automated clustering algorithms, but often cannot be assessed quantitatively.

Resulting spike clusters may include both false positive and false negative errors, and the larger the proportion of each, the noisier any subsequent analysis will be. It is therefore important to determine the quality of unit isolation, to be able to evaluate the significance of any analysis that assumes single cell isolation. In an experiment in which spike sorting from tetrode recordings was evaluated with knowledge of the true spike times of a cell that was also recorded intracellularly, manually cut clusters contained somewhere between 0 and 30% errors, depending on the recorded amplitude of the spike (which corresponds to the inverse square of the distance between the tetrode and the neuron) and the experience of the cluster cutter (Harris et al., 2000). Better results were achieved with a semi-automated algorithm, (in which the automated algorithm delineates clusters in multi-dimensional space, and then the human operator merges over-clustered units) up to a theoretical limit determined again by the strength of the signal, and the degree of violation of the assumptions of the algorithm the cell displays, including burstiness and overlapping spikes. Without simultaneous intracellular recordings, the true cell classification of spikes cannot be known, and thus any measures of isolation quality or errors are only estimates. The strength of the signal can by quantified by the signal to noise ratio (S/N), which determines how good the recording of a particular neuron will be, and can be used to ensure that spikes will not be excluded because they do not pass the threshold of detection due to noise in the recording (Lemon, 1984). This metric does not, however, quantify the proportion of undetected spikes, nor does it measure whether spikes from two cells have been classified as one unit (Joshua et al., 2007).

Unit isolation measures
One way to evaluate cluster isolation quality is to use the same measures used to separate clusters (e.g. waveform shape and amplitude). Harris et al. (2001) suggested a metric called isolation distance to calculate the distance between a cluster and all the other spikes recorded on the same tetrode (in the 8-dimensional space defined by the energy and first principal component of waveform shape on the 4 recording channels). This metric measures the distance from the center of an identified cluster that will contain all the spikes in the cluster as well as the same number of noise (non-cluster) spikes. When used on the Harris et al. (2000) tetrode recordings, which also included intracellular recordings indicating the true result, the isolation distance of a cluster was found to correlate well with the false positives included in that cluster and very slightly, but significantly, with the number of false negatives (Schmitzer-Torbert et al., 2005). While this measure allows the comparison of the isolation of units recorded on a single tetrode, it is very dependent on the distribution of the non-cluster spikes, and thus cannot be used to quantitatively compare values between tetrodes. For example, a cluster that is very close to one other smaller cluster, but far away from the rest of the spikes recorded on that tetrode, may earn a higher isolation score than a lower amplitude cluster (closer to the multi-unit noise) that is nonetheless farther removed from all other points. Therefore it is not possible to design a criterion value of isolation distance that can be compared between experiments. Schmitzer-Torbert and Redish (2004) designed a different measure of cluster isolation, L-Ratio, which discounts noise spikes that are distant from the center of the cluster, providing a more accurate account of the distribution of noise spikes surrounding a cluster. This measure was found to correlate pretty well both with true false positives and true false negatives (Schmitzer-Torbert et al., 2005). Joshua et al. (2007) went a step further when defining their “isolation score,” and quantified the local distances between points within the cluster and those outside. They also tested their score with a simulation of false positive and negative errors, which they used to suggest a threshold isolation score that would determine units suitable for further study. In addition to the isolation score, the authors designed separate calculations for
estimating false positives and false negatives, and determined that these estimates are likely to be accurate for units above the threshold isolation score. Their measures worked well in simulation, but have not been evaluated with data for which the true false positive and false negative rates are known (Harris et al., 2000), and thus cannot be adequately compared to the two above-mentioned measures. One fault in common between all above measures is that they only use features also used in automated spike sorting, and thus are subject to the same assumptions. For example, a cluster of a bursty cell may show a worse isolation score than two clusters made up of the high amplitude (early in the burst) and low amplitude (late in the burst) spikes from the same cell.

A method to detect false positive errors in spike sorting that is completely independent of the sorting method involves counting the spikes that occur during the refractory period following another spike in the same cluster. The spikes that occur during the refractory period cannot belong to the same cell, and thus the proportion of spikes showing a low (2-3ms) inter-spike interval can be used to estimate the contamination of a cluster, under the appropriate conditions. This value has been used in two main ways: calculating just the proportion of spikes that occur in the refractory period (Takehara-Nishiuchi and McNaughton, 2008), or the ratio of spikes in the refractory period (2 ms within another spike) to spikes within 10 ms of another spike (R2:10; Fee et al., 1996). Unlike the proportion of spikes in the refractory period, R2:10 is an indicator of contamination which is independent of the firing rate of the unit (see Results and below), however, bursting neurons increase their firing rate in the 10 ms following a previous spike, and thus this value is higher than the overall firing rate for those neurons. Harris et al. (2000) recommended comparing firing in the refractory period to the asymptotic value of the autocorrelogram (which is the firing rate), but did not suggest a quantitative criterion.

Krupic et al. (2012) do not report what criterion for refractory period violations was used in their study, but hippocampal electrophysiology studies in our lab typically exclude any units with a percentage of spikes in the refractory period greater than between 1 and 0.2%. The false positive rate can be
calculated from the percentage of spikes occurring in the refractory period based on the expected probability that a rogue spike occurs around the spikes of the clean unit (Hill et al., 2011). From Hill and colleagues’ equation, the false positive rate (c) depends on the proportion of spikes occurring in the refractory period (p) and the firing rate of the cluster (F) (see Results and Figure 6). For a cell that fires at 10Hz, 0.2% spikes in the refractory period represents a 5% false positive rate, but for a cell that only fires at 1Hz (which is the average rate reported in the Krupic at al. study), 0.2% spikes in the refractory period would mean a false positive rate of higher than 50% (see Figure 6). From our experience, most low firing rate clusters would not show this high of a rate of refractory period violations, but this calculation illustrates that the criterion, needs to be adjusted for cells with low firing rates. Further, this calculation makes the assumption that contaminating spikes occur independently of spikes from the unit of interest, which is definitely violated when the contamination comes from another place-related cell. Two grid cells with non- or only partially overlapping fields (such as those we show in Fig. 1 and 2, which would sum to a “band-like” response) should show very little spike timing overlap, and thus in that case this calculation is a severe underestimate of the contamination rate. On the other hand, during SPWs spike timing overlap is higher than expected by chance, and so when epochs containing many SPWs are used for spike sorting, this calculation may be an overestimate of the false positive rate.

Additional methods of estimating false positives and false negatives include estimating the spikes missed as a result of the detection methods, and measuring the overlap between each pair of clusters (Hill et al., 2011). The proportion of spikes that have not been detected based on a threshold spike amplitude (or any other detection method) can be calculated by plotting the distribution of the cluster around the detection threshold, fitting a Gaussian to the distribution, and calculating the proportion of spikes that would fall under the threshold but have not been detected (Hill et al., 2011). This problem should be reduced, however, by using only cells with a high enough S/N and setting an appropriate detection threshold. The other causes of undetected spikes are
spikes that occurred during the detection of another spike (either because of spike overlap, or because of the censored period following detection of a spike). The expected proportion of spikes missed in this way can be easily calculated (under the assumptions of Poisson spiking that is uncorrelated between neurons) from the firing rate of the cell, the total number of detected events and the length of each censored period (Hill et al., 2011). Hill and colleagues (2011) suggest estimating the total number of false positives and false negatives by combining these measurements of undetected spikes, the measurement of false positives from refractory period violations, and a measurement of the probability of misclassifying each spike into another identified cluster. As complete as this calculation attempts to be, it is still only an estimate, since the statistics of spike timing are not strictly Poisson (and all the calculations dependent on expected rates of spikes assume Poisson statistics). Nonetheless, Hill et al.’s calculation captures the idea that no one measure gives a complete picture of the isolation quality of a unit and that strict criteria should be set and reported for inclusion of “well-isolated single units” in an analysis.

Conclusions

While methods to determine cluster isolation quality are still imperfect, the reporting of the use of such metrics is even more so. A literature search of the spike sorting metrics used in prominent laboratories in the in vivo electrophysiology hippocampal field indicated that very few papers actually report which specific criteria were used for evaluation of spike sorting quality. Many papers report qualitative criteria (e.g. “Only units with clear refractory periods are included”), or which measure was used (e.g. isolation distance was calculated), without stating the criterion value(s), and many do not report which measures were used at all. As more laboratories use these techniques and more spike sorting methods are introduced, this standard will have to change, so that readers and reviewers can better evaluate the results and conclusions of a paper. Promising new spike sorting methods are being developed (e.g. Takekawa et al., 2010), which have yet to be tested with novel in vivo data, and
there is a push for more automated methods (for review, see Rey et al., 2015). This makes it even more important to discuss isolation quality and come to some consensus on criteria.

Neuronal population analyses don’t necessarily require perfect cell isolation (e.g. Davidson et al., 2009), but when making a statement about the receptive fields of single cells, extreme caution should be used to determine that the property in question is true of even the best isolated single units. We have shown in this paper that, given the current knowledge of grid cells, contamination of grid cell clusters would result in multi-peaked spatial receptive fields that could be considered band-like. Because of the difficulty in ascertaining single cell isolation, and the simulation that we show here, we believe that Krupic et al. (2012) did not provide definitive proof of “band-like” cells in either of the two regions they recorded from (MEC and PaS). In order to demonstrate that such cells exist, they would have to show several more rigorous analyses to prove that the units they claim are band cells are very well isolated, as well as that their Fourier transform analysis is reliable at identifying band cells. Krupic et al. (2012) claim that showing the mean isolation distance and L-ratio measures for their “grid” and “band-like” cells proves that the two classes are equally well isolated. We show in figure 3, however, that contaminated clusters could also be classified as grid cells. We posit that poorly isolated units also contributed to the mean isolation score for the “grid cell” class, and suggest that the mean values of those measures are not definitive proof that the classes are equal, as it does not show the distribution of isolation metrics for each recorded unit. Showing this distribution, or the correlation between isolation and gridness score would better illustrate whether any of the best-isolated units contribute to the finding of band-like cells. Krupic et al.’s observation that grid cells have greater between-session stability and than spatially-periodic non-grid cells (their Fig 2) is consistent with our hypothesis that their band-like clusters are less likely to be well-isolated.

Finally, we show that Krupic and colleagues’ raw data (spatial firing rate plots) show multiple peaks along the axis of each of their identified “bands” (Fig. 5), putting into question the reliability of their Fourier transform analysis for
identifying bands. This analysis attempts to fit differently oriented and spaced band patterns to the spatial firing plot of units (see Krupic et al.'s figure 1F), and then interprets the presence of less than 3 peaks (which indicate grid cells) in the resulting Fourier plot as evidence of band-like firing. We found that the presence of spikes from other grid cells reduces the number of peaks in this spatial Fourier transform (figure 2), and that this measure becomes unreliable with increasing spacing of fields when only a few fields are observed in the environment (data not shown). Krupic et al. used data from units with few and very noisy fields in their analysis of the proportion of grid and band-like cells, without accounting for how unreliable the measurements on these units are. Their observation that spatially periodic non-grid cells show less precise Fourier component orientation tuning than grid cells (their figure 3) is consistent with our suggestion that many of their band-like cells are poorly isolated and/or sparsely spiking, resulting in a less reliable Fourier analysis. When they initially discovered grid cells, Fyhn et al. (2004) compared the waveforms of spikes in each of the regularly repeating fields, to ensure that each field was in fact a result of firing of the identical cell, and that their finding was not due to pure isolation (their supplementary figure 3). Krupic et al. (2012) could show a similar analysis for the multiple peaks observed in the spatial firing plots of their “band-like” cells, to convince us that our simulation results do not apply to their data. Therefore, we call on Krupic and colleagues to show better evidence of cluster isolation quality, and analyze their best-isolated units for “bandiness” with a measure that includes a reliability estimate.

Methods

Grid Cell Simulations:
Activity of simulated grid cells was generated using position data from a rat randomly foraging in a 1m x 1m box during hippocampal recordings. The position of the rat was determined from a circular set of lights on the head stage connected to a hyperdrive implanted over the rat’s head and recorded on an
overhead camera. For each time frame of the video (60 frames/s), a circle was fit
to the active pixels in the video, and the center of this circle was counted as the
rat position. Noise in the video was removed by considering only pixels within a
small radius of the position in the previous time frame.

For each simulated grid cell, an independent hexagonal grid pattern was
defined in space and the number of spikes generated by the cell was calculated
based on the rat's location for each point in time. The probability a simulated cell
would generate a spike depended on the distance from the nearest vertex at that
point in time. Specifically, the probability of a spike at time $t$, $p_t$, was:

$$p_t = Ke^{-x/\tau}$$

where $K$ was the maximum probability of generating a spike at a given location
($K = 0.012$ to $K = 0.12$ per video frame, frame duration 16.7ms), $x$ was the
distance to the nearest grid vertex, and $\tau$ was 20% of the distance between
vertices. $\tau$ was set to 20% to approximate the field width observed in grid cells
recorded in MEC.

Contaminated “units” were created by combining spikes from two
simulated cells with varying spiking probabilities. For example, spikes from a
simulated grid cell with $K=0.12$ were combined with spikes from a simulated cell
with $K=0.06$ to create a 33% contaminated cell. Grid patterns of combined cells
had the same orientation and periodicity but different spatial phase.

Spike trains generated as described above were then used for standard
grid cell analyses. The position of the rat as well as the locations where each
spike was fired was plotted (Fig. 1A); spatial firing rate maps were computed
(Fig. 1B), and used to generate spatial autocorrelations (Fig. 1C); and gridness
scores were calculated.

Gridness Score:

Grid scores were calculated as in Sargolini et al. (2006) and Langston et al.
(2010). The autocorrelogram of the smoothed and occupancy normalized spatial
firing plot was calculated, and a circular sample centered on the central peak
(with the central peak removed) was used to calculate the grid score. This score
is the maximum difference between the correlations of the circular sample at 60
and 120 degree rotations versus 30, 90 and 150 degree rotations. The original
“gridness” score designed by Sargolini et al. (2006) used a circular sample of the
autocorrelrogram defined by the first six peaks (outside of the central peak).
Because cells that are less grid-like do not have six clear peaks at the same
spacing in the autocorrelrogram, Krupic et al. (2012) found the first peak, and
expanded the circle to 2.5 times that distance. Langston et al. (2010) used a
different method to calculate grid scores for poorly defined grid cells of young
rats, which was to calculate a grid score for each circular sample between 10cm
and 10cm less than the width of the box, and use the maximum. Each method
calculates different values of grid scores, but they are correlated (Langston et al.,
2010). To calculate the significance of the grid score, the timing of the spikes is
offset by random values (greater than 20s), and the resulting spike trains (and
the semi-random positions of the animal at those spike times) is used to calculate
a distribution of grid scores. Any grid score above the 95-percentile value of the
grid scores from randomized spike trains is considered significant. Using the
same method, we calculated a distribution of randomized grid scores with our
simulated “units” for the Sargolini et al. (2006), peak-based grid score, and the
Langston et al. (2010), best radius-based grid scores, and found 95-percentile
thresholds of 0.13 and 0.43, respectively.

Fast Fourier Transform:
The periodicity of the spatial responses of clusters was estimated using a two-
dimensional Fourier transform of the unsmoothed firing rate map (Krupic et al.,
2012). The unsmoothed map was represented in a 64x64 array (corresponding
to 1.5 cm² bins) with the mean value subtracted from each bin. The array was
then zero-padded out to 256x256, to increase spatial resolution, and then the
Fourier transform applied. The two-dimensional Fourier spectogram was shifted
such that low frequencies were at the center and higher frequencies were in the
periphery.
Grid cell recording:

Extracellular recording was performed in a Long Evans rat using a chronically implanted 12-tetrode 'hyperdrive' (see e.g. Navratilova et al., 2012). All animal procedures were performed in accordance with the animal care guidelines issued by the government of Belgium, and approved by the institutional welfare body at KU Leuven. The hyperdrive was implanted 3.6mm laterally in the right hemisphere, 0.2-0.3 mm anterior to the transverse sinus, and at an angle of 10° to the sagittal plane in the anterior direction. Tetrodes were gradually lowered to 2.5-3mm below brain surface over 1-2 weeks. Rat was trained to forage for chocolate and was kept at 85% of free-feeding body weight. Recording was performed in an open planar area, with area of movement limited only by the length of recording wires. A single boundary existed at the edge of the foraging area. Foraging radius was approximately 1.7 meters.

'Bandiness' analysis:

To examine the band-like response of cells reported by Krupic et al. (2012), the smoothed position maps and corresponding Fourier plots (Krupic et al., 2012, Supplementary Fig. 10) were rotated until lines in the autocorrelogram were horizontal. Using image manipulation software (www.gimp.org), a rectangular section was taken from the rotated map at each location corresponding to a periodic band. A reverse mapping was made between the color of the position map and the map intensity. Using the reverse mapping, the average intensity of each band was plotted as a function of position along the band.

Acknowledgements: We would like to thank E. Moser for sharing Matlab code used in his laboratory for calculation of gridness scores. This work was supported by a Polaris Award from Alberta Innovates Health Solutions, by a Discovery grant from the National Sciences and Engineering Research Council of Canada, and by an operating Grant from Neuroelectronics Research Flanders.
References


Figure legends

**Figure 1**: Numerical simulation of spatially periodic bands arising from overlap of the grid fields of two poorly isolated or possibly physiologically coupled grid cells. 

**A**: Position data from a rat foraging randomly in an environment is used to create a simulated spike distribution from two grid cells (red and blue). The spikes from cell 1 (red) are combined with 50% of the spikes from cell 2 (blue), to generate a “unit” containing 33% contamination. 

**B**: Smoothed and occupancy normalized firing rate distributions of the spikes from the simulated cells and the combined contaminated “unit.” The firing rate color scale ranges from 0 Hz (blue) to 5 Hz (dark red) in all images. 

**C**: Spatial autocorrelogram of the smoothed firing rate distributions. 

**D**: 2-D Fourier transform of the clean and contaminated units, calculated as in (Krupic et al., 2012).

**Figure 2**: Increasing contamination from a second grid cell increases the appearance of a band-like pattern. Contaminations of 10, 25, and 33% are shown. 

**Left**: Smoothed firing rate plots contain several bumps, corresponding to firing fields of two cells. The firing rate color scale ranges from 0 Hz (blue) to 5 Hz (dark red) in all images. 

**Middle**: Autocorrelograms also contain multiple bumps. 

**Right**: As a result of the contamination, two or more peaks in the Fourier transform become weaker. Grid scores for these simulated units decrease from 0.39 and 0.27 (significant gridness) to 0.049 (below the 95% percentile of shuffled data).

**Figure 3**: Varying the offset and the direction of the offset of the contaminating grid cell results in varying “gridness” scores, a varying appearance of “bandedness,” and different numbers of peaks in the spatial Fourier transform. 

**A**: Schematic of mixtures of two grid cells with different offsets from each other. 

The component of the offset along the x-axis (one of the axes of the grid pattern) is expressed in percent of grid spacing, and the offset along the orthogonal axis is expressed as an angle from the x-axis. (The columns correspond to 10%, 30%...
and 50% x-axis offsets, and the rows to approximately 29% 18% and 5% y-axis offsets).

**B**: Examples of spatial firing (top), spatial autocorrellograms (middle) and spatial Fourier transforms (bottom) of units with 33% contamination from cells with offsets i-v, shown in panel A. The firing rate color scale ranges from 0 Hz (blue) to 5 Hz (dark red) in all spatial firing rate images. A. The prototypical examples are i (almost perfect grid) and iv (perfect honey comb). Unit ii is an example of a contamination of medium offset along a non-grid axis (30% x-(grid)-axis offset, and 17.3% y-axis offset), which distorts the grid somewhat, but the gridness score is still significant. Unit v contains a contamination with large offset along a non-grid axis (50% x-axis offset, 18.2% y-axis offset), which results in a more zigzag, rather than band-like pattern, but the grid is distorted enough to make the gridness score non-significant, and only one axis on the Fourier transform shows significant peaks. The Fourier transforms also show artifacts relating to the enclosure walls, and the fact that only a small portion of the spatial pattern is sampled.

**C**: “Gridness” score as a function of offset for all simulated units containing 33% contamination (red circles). Simulated “clean” grid cells are also plotted (blue circles). The size of the circle indicates the average gridness score of all simulations at that offset and 33% contamination. Gridness scores that are not significantly different from random are also marked with an asterisk. Each simulation was repeated for two different paths through the environment, and 1-4 times with different Poisson distributions of spikes. (Thus each point is the average of 2-8 simulations). All average grid scores of contaminated and clean units are plotted repeatedly, to illustrate the pattern of the grid. This figure illustrates that contamination of a grid cell (with vertices at the blue circles) with spikes from another grid cell (with vertices at one set of red circles), results in non-significant gridness scores at certain grid field offsets, but significant gridness at other offsets.
Figure 4: Comparison of spiking distributions for a grid cell, simulated grid cells with various contaminations, and a hypothetical band cell. 

A. Spiking activity of a recorded grid cell. White line indicates the position of the animal over time. Red shows the locations when this cell spiked. The overlaid rectangle outlines a slice that was analyzed for band-like behavior. Bottom panel shows the relative spiking response as a function of location along the long axis of the rectangle (number of spikes divided by occupancy). Red line is the firing rate smoothed with a 4-6cm kernel.

B. A simulated grid cell. Simulated spikes were generated based on position, using the same tracking data as recorded in A. Analysis rectangle is the same as in A.

C. The summed spikes of two simulated grid cells. Spikes were generated for two grids, both aligned with the long axis of the analysis rectangle, with vertices 50% offset. 67% of spikes come from one grid cell and 33% are from the other.

D. The summed spikes of three simulated grid cells. Three grids with vertices aligned along the analysis rectangle and offset 33% and 67% from one another were simulated. 50% of spikes come from one grid cell, and 33% and 17% from the other two.

E. A simulated band cell. Spikes were generated in the same way as B-D, except instead of distance from the vertex of hexagonal grids; the distance from 3 lines aligned with the analysis rectangle was used. The result shows a band of spiking activity (top) with no periodicity and smaller variance to mean ratio (bottom) than the cases in which apparent bands are the result of grid superposition.

Figure 5: Periodic firing rate modulation within bands recorded by Krupic et al. suggests the summation of multiple fields. 

A: An analysis of firing rate along the bands was conducted on best 8 examples from Krupic et al. (2012). Cells with band-like behavior were taken directly from their figure S10 and rotated so the axis of the ‘band’ in the firing rate plot (i) and spatial correlogram (ii) was horizontal. Each band was extracted from the firing rate maps (iii), and the firing rate (intensity) along this band was estimated by converting the color map (iv).
If these activity patterns represented bands from hypothesized band cells, the intensity of activity should be largely uniform across the length of the band. Instead, the activity within each purported band is variable, and usually periodic. The maximum firing rate (in Hz) from each firing rate plot in Krupic’s figure S10 is displayed on the y-axis of our converted intensity plot.

**Figure 6:** Assuming Poisson firing statistics and uncorrelated tuning curves of neurons, the false positive rate (contamination rate) of a unit can be estimated from the number of spikes occurring during the refractory period following another spike (see text). **A:** False positive rate as a function of firing rate of the cluster, given an accepted percentage of 0.2% and 0.02% spikes in the refractory period (p). Note that a contamination rate of 0.5 (50%) indicates a cluster containing 50% true spikes, and 50% spikes from other neurons. **B:** The acceptable percentage of spikes in the refractory period should vary based on firing rate of the unit, and the experimenter’s accepted false positive rate, for example 5% or 10%, as plotted here.