Accurately estimating neuronal correlation requires a new spike-sorting paradigm

Valérie Ventura[a,b,1] and Richard C. Gerkin[b,c]

[a]Department of Statistics, [b]Center for the Neural Basis of Cognition, and [c]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA 15213

Edited by Stephen E. Fienberg, Carnegie Mellon University, Pittsburgh, PA, and approved March 2, 2012 (received for review September 28, 2011)

Neurophysiology is increasingly focused on identifying coincident activity among neurons. Strong inferences about neural computation are made from the results of such studies, so it is important that these results be accurate. However, the preliminary step in the analysis of such data, the assignment of spike waveforms to individual neurons ("spike-sorting"), makes a critical assumption which undermines the analysis: that spikes, and hence neurons, are independent. We show that this assumption guarantees that coincident spiking estimates such as correlation coefficients are biased. We also show how to eliminate this bias. Our solution involves sorting spikes jointly, which contrasts with the current practice of sorting spikes independently of other spikes. This new "ensemble sorting" yields unbiased estimates of coincident spiking, and permits more data to be analyzed with confidence, improving the quality and quantity of neurophysiological inferences. These results should be of interest outside the context of neuronal correlations studies. Indeed, simultaneous recording of many neurons has become the rule rather than the exception in experiments, so it is essential to spike sort correctly if we are to make valid inferences about any properties of, and relationships between, neurons.

correlated spikes | clustering | point process history | statistical bias | statistical power

The behavior of neural systems is frequently explored by examining the temporal activity of the constituent neuronal cells. Individual cells are characterized by their spike times, descriptive statistics of which are used to characterize responses to stimuli, motor output, and other cognitive states related to memory, emotion, and planning (1). Neuron ensembles are also characterized by temporal relationships between cells, as measured by correlations between cell pairs. Understanding these relationships may be key to deciphering the ensemble neural code of the brain in vivo. For example, cells with correlated activities may be involved in perception (2) and memory formation (3). The correlation structure of cell ensembles also dictates how much information is passed to downstream brain areas (4, 5), and the probability of activating those areas (6). Measurements of correlation structure can also be used to decode subsequent activity (7, 8).

Measuring neuronal correlations requires simultaneously recording multiple cells, and technological developments have made this ensemble approach routine (9). Spike sorting, the assignment of spikes' waveforms to each cell that emitted them (10), is challenging. When waveforms do not cluster perfectly, some spikes will be incorrectly assigned to the wrong cell. And as the number of cells increases, or the signal-to-noise ratio of the recording decreases, assignments become increasingly ambiguous. Conventionally, a measure of "isolation quality," the discriminability of a cell's spike waveforms from those of other cells, determines which cells are worthy of further analysis (11). Cells of inferior isolation quality are typically discarded to preserve the integrity of subsequent analyses.

This results in an unfortunate trade-off. Accepting only the best isolated cells ensures the integrity of analyses but wastes data. For example, if half of recorded cells are excluded then the number of pairs available for computing correlations is reduced by $\frac{3}{4}$. Alternatively, we can retain poorly isolated cells at the risk of corrupting subsequent analyses and undermining the biological interpretation of the data. Ref. 12 previously showed that poor isolation yields biased firing rate estimates. Here, we show that neuronal correlation estimates are biased, unless isolation quality is perfect, or cells are Poisson, independent, and have equal firing rates. Hence, the trade-off between wasted data and mistaken inference may seem unavoidable. And as we move to the study of even higher-order relationships in neuronal networks (13), this trade-off worsens exponentially.

We are at an impasse because the "traditional" analytical framework treats spike sorting and data analysis serially: First, spikes are assigned to cells using waveform information, without regard for correlations between cells; second, these assignments are used to compute *those same previously ignored correlations*. However, we show that associations between cells contain information about spike identities, and that ignoring this information for spike sorting induces bias in correlation estimates. Refs. 12 and 14 had made a related point: When the firing rates of poorly isolated cells are modulated by covariates, these covariates contain spike identity information, and ignoring it for spike sorting induces bias in firing rate estimates. The solution is to sort spikes using all available information about spike identities, not just waveform information. Refs. 12 and 14 showed how to embed the spiking rates of independent cells into spike sorting. Here we show how to embed the joint spiking probabilities of dependent cells. This results in "ensemble sorting," whereby spikes are sorted jointly rather than one at a time. We show that ensemble sorting yields unbiased correlation estimates, even for poorly isolated cells. This enables a more efficient use of data, since more cells can be included in correlation analyses, without corrupting scientific interpretations.

Even though the scientific focus of this paper is the estimation of coincidence rates and correlations, it should be of interest to all areas of systems neuroscience. Indeed, ensemble analysis of many cells has become the rule rather than the exception in experimental neuroscience (15), making it essential to sort spikes with these analyses in mind if we are to correctly infer both the individual and the ensemble properties of cells.

Background and Results

Traditional spike-sorting algorithms sort spikes one at a time, so they ignore dependencies between spikes, and therefore between cells. This systematically biases subsequent estimates of neuronal correlation. We propose an "ensemble" approach to sort spikes jointly, which takes into account dependencies between spikes

Author contributions: V.V. and R.C.G. designed research; V.V. and R.C.G. performed research; V.V. contributed new analytic tools; R.C.G. analyzed data; and V.V. and R.C.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: vventura@stat.cmu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1115236109/-/DCSupplemental.

and cells. We show that subsequent correlation analyses are unbiased, and we illustrate our results on simulated data.

Spikes are Sorted One at a Time in Traditional Spike Sorting. This section presents a synopsis of spike sorting. Additional details can be found in the *Supporting Information, Sec. S1*. Consider an electrode that records I cells. When its bandpassed voltage exceeds a threshold, we record a snippet of data a around the threshold crossing. All spike-sorting methods assume implicitly or explicitly that suprathreshold waveform measurements a originate from a mixture distribution

$$f(a) = \sum_{i=1}^{I} \pi_i f_i(a), \qquad [1]$$

which simply states that, given a suprathreshold event, the probability that it originated from cell i is π_i, and if so, its measurement a arises from a distribution f_i centered around the true waveform of cell i. Eq. **1** is often visualized by overlaying the raw waveforms or by plotting their principal components against one another, which aims to reveal clusters that each corresponds to a different cell. Spike sorting consists of separating these clusters.

Many methods exist to separate clusters (10, 16). We focus on model-based clustering because it is optimal when the correct model is used (ref. 17, p. 350], and it can be extended to correct the deficiencies identified herein. Model-based clustering starts with applying Bayes rule to Eq. **1** to obtain the posterior probability that a spike with waveform a was emitted by cell i

$$P(X = i|a) = \frac{\pi_i f_i(a)}{f(a)}, \qquad [2]$$

where the denominator is Eq. **1** and the numerator is one of its summands. The spike is then assigned to the cell with identity x that maximizes $P(X = i|a)$,

$$x = \underset{i=1,\ldots,I}{\operatorname{argmax}}\ P(X = i|a). \qquad [3]$$

The most important feature of traditional spike sorting, with respect to our claims, is that spikes are sorted one at a time; i.e., independently of other spikes.

Spikes Should be Sorted Jointly: Ensemble Sorting. A standard probability result states that only Poisson spike trains are memoryless (18). In that case, spikes are independent, which means that the probability of a spike occurring at a particular time does not depend on the past—i.e., does not depend on the times of previous spikes or on which cells emitted these spikes. Conversely, this implies that if an electrode spike train is not Poisson, then the probability of a spike occurring at a particular time does depend on the times of previous spikes and on which cells emitted them. That is, *the history of a non-Poisson spike train contains spike identity information*. Current spike-sorting methods ignore that information since they sort spikes independently of other spikes. In Theorem 1.1, we show that this induces bias in estimates of neuronal correlations between cells.

This is of concern because electrode spike trains are seldom Poisson. Indeed, because only Poisson spike trains are memoryless, *spikes are independent if and only if all cells are Poisson and are mutually independent*. But phenomena such as refractory periods imply that spike timings are not independent; they depend on when and which cells spiked in the past. Refs. 19 and 20 handled this by prohibiting adjacent spikes from being assigned to the same cell. But more crucially in the context of studying neural correlations, the combined spike train of two or more dependent cells cannot be Poisson, even when individual cells are Poisson, since their spikes are dependent. Below, we develop the statistically optimal spike-sorter for non-Poisson electrode spike trains, and in Theorem 1.2, we show that this sorter yields unbiased estimates of neuronal correlations.

Assume that we want to sort jointly n spikes. We denote by $\underline{X} = (X_1, \ldots, X_n)$ the vector of their unknown identities; X_j can take values in $\{1, 2, \ldots, I\}$, where $X_j = i$ means that cell i emitted spike j. Spike identity information is contained in the waveform features vector $\underline{a} = (a_1, \ldots, a_n)$, and, as argued above, in the history of the electrode's spike train. This history is fully summarized by the vector of interspike intervals (ISIs), $\underline{s} = (s_1, \ldots, s_n, s_{n+1})$, where s_j is the time elapsed between spikes $(j-1)$ and j. Spike sorting aims to estimate $X_1, \ldots, X_n$. The optimal approach relies on calculating the I^n joint posterior probabilities of possible spike identities, given the variables that contain identity information:

$$P(X_1 = i_1, \ldots, X_n = i_n|\underline{a}, \underline{s}). \qquad [4]$$

Then the equivalent of single spike assignments in Eq. **3** consists of assigning the n spikes to the cells with identities

$$(x_1, x_2, \ldots, x_n) = \underset{(i_1,\ldots,i_n)}{\operatorname{argmax}}\ P(X_1 = i_1, \ldots, X_n = i_n|\underline{a}, \underline{s}). \qquad [5]$$

Notice that Eqs. **4** and **2**, and **5** and **3**, have parallel structures. Eqs. **2** and **3** apply to a single spike, while Eqs. **4** and **5** apply jointly to n spikes, and are further conditioned on the spike train's history $\underline{s}$. To illustrate Eqs. **4** and **5**, consider an electrode that records $I = 3$ cells (A, B, and C), and imagine that $n = 5$ spikes were observed. The spikes' joint identities can be one of $I^n = 3^5$ possibilities, with joint posterior probabilities (Eq. **4**) listed in Fig. 1. The largest is $P((X_1, X_2, X_3, X_4, X_5) = (A, A, B, A, A)|\underline{a}, \underline{s}) = 0.4$, so joint spike sorting (Eq. **5**) assigns the third spike to cell B and all others to cell A.

Ensemble sorting is fully determined by the joint posterior probability in Eq. **4**. To evaluate it, we use Bayes rule to rewrite

$$P(X_1 = i_1, \ldots, X_n = i_n|\underline{a}, \underline{s}) = \frac{\pi_{\underline{i}}(\underline{s}) f_{\underline{i}}(\underline{a}|\underline{s})}{f(\underline{a}|\underline{s})}, \qquad [6]$$

where the denominator is the sum of the numerator over all possible identities $\underline{i}$. Eq. **6** is the multivariate equivalent of the spike-sorting rule for a single spike in Eq. **2**, but with all quantities now functions of the spike train history $\underline{s}$. To use Eq. **6**, we need to specify $f_{\underline{i}}(\underline{a}|\underline{s})$ and $\pi_{\underline{i}}(\underline{s})$. The first term, $f_{\underline{i}}(\underline{a}|\underline{s})$, is the joint distri-

Possible Spike Identities		Probabilities
Bin 1	*Bin 2*	
A, A	B, A, A	0.40
A, A	B, B, C	0.20
B, A	B, A, A	0.13
B, A	B, B, C	0.10
C, A	B, A, A	0.07
C, A	B, A, B	0.06
C, A	B, A, C	0.04
Other Configurations		0
		Total = 1

Fig. 1. Hypothetical joint posterior probabilities (Eq. **4**) of the identities of five spikes observed on an electrode that records three cells, *A*, *B*, and *C*. For example, the probability that these spikes were emitted successively by cells *B*, *A*, *B*, *B*, and *C* is 0.10. Spikes are allotted by time into bins for the computation of coincidence measures.

bution of n observed waveforms $\underline{a}$ emitted by cells $\underline{i}$, and with ISIs $\underline{s}$. If waveforms are stationary, they do not depend on the ISIs, so $f_{\underline{i}}(\underline{a}|\underline{s}) = f_{\underline{i}}(\underline{a})$ reduces to the product of the marginals, $\prod_{j=1}^{n} f_{i_j}(a_j)$, where f_i is the waveform distribution of cell i that appeared in Eqs. **1** and **2**. Otherwise, a dependent waveform model must be used, e.g., ref. 20 to model waveform attenuations; this is described in *Supporting Information, Sec. S2.2*. The second term, $\pi_{\underline{i}}(\underline{s}) = P(X_1 = i_1, \ldots, X_n = i_n|\underline{s})$, is the probability that the n spikes were emitted by cells $\underline{i} = (i_1, i_2, \ldots, i_n)$, given the history $\underline{s}$ of the electrode spike train. If the cells are independent and Poisson, spikes do not depend on the past, so $\pi_{\underline{i}}(\underline{s}) = \pi_{\underline{i}} = \prod_{j=1}^{n} \pi_{i_j}$, where π_i is the proportion of spikes from cell i that appeared in Eqs. **1** and **2**. Otherwise, $\pi_{\underline{i}}(\underline{s})$ requires models for the cells firing rates, conditional on the activity of the cell ensemble, and the implied distribution of the ISIs. We discuss this further in the *Supporting Information, Sec. S2.2*, where we also derive the analytic expression for $\pi_{\underline{i}}(\underline{s})$ for the example of the illustration section.

Neuronal Correlation Estimates are Biased. Several measures of correlated neuronal activity exist: For example, spike count correlations, coincidence rates, joint surprise (21), or information-theoretic measures capturing nonlinear associations; e.g., mutual information. To streamline the presentation, we focus in this section on coincidence rates, where by coincidence we mean co-occurrence within a time bin of width γ, not simultaneity. Results generalize to other measures, as shown in the illustration section, since coincidence rates are related arithmetically to correlations and, under certain assumptions, to other measures.

Let θ denote the true coincidence rate for a pair of cells. The usual estimate for θ is the proportion of bins in which both cells emitted at least 1 spike each; we call it $T_{\perp}$, where the "independent" subscript $\perp$ serves as a reminder that spikes were sorted one at a time. We call $T_{\cup}$ the same estimator calculated from spikes sorted jointly per Eq. **5**; the "union" sign $\cup$ symbolizes that spikes were sorted together. We do not however advocate the use $T_{\cup}$, but that of a related estimator, $\tilde{T}_{\cup}$, which is shown to have desirable properties in Thm.1.2: $\tilde{T}_{\cup}$ is the average over bins of the joint posterior probabilities that both cells spiked in these bins; $\tilde{T}_{\cup}$ preserves the uncertainty in determining spike identities when waveform clusters overlap, in contrast to the current practice of considering that the largest posterior probability (Eqs. **3**, **5**) provides the correct spike assignments. This is discussed further in *Supporting Information, Sec. S3*.

We illustrate the calculations of $T_{\cup}$ and $\tilde{T}_{\cup}$ using the data in Fig. 1. Imagine that we cut the spike train in two bins of width γ so that the first two spikes belong to the first bin, and the last three to the other bin. We want to estimate the coincidence rate θ of cells A and B in bins of such duration. We determined earlier that joint sorting assigns spike 3 (in bin 2) to cell B and all other spikes to cell A. Only the second bin contains spikes from cells A and B, so the estimate of θ is $T_{\cup} = 1/2$. To calculate $\tilde{T}_{\cup}$, we need the posterior probabilities that cells A and B spiked in each bin. Standard probability theory obtains these by summing the full joint posterior in Eq. **4** over the identities of all spikes outside of that bin. Hence, the posterior probability that cells A and B both spiked in the first bin is $P((X_1, X_2) = (A, B) \text{ or } (B, A)|\underline{a}, \underline{s}) = 0.13 + 0.1 = 0.23$. The posterior probability that cells A and B both spiked in the second bin is $P((X_3, X_4, X_5) = \text{any permutation of}(A, B, k), k = A, B, C|\underline{a}, \underline{s}) = 0.4 + 0.13 + 0.07 + 0.06 + 0.04 = 0.7$. This gives $\tilde{T}_{\cup} = (0.7 + 0.23)/2 = 0.465$.

We now state properties of traditional and ensemble coincidence rate estimates, $T_{\perp}$ and $\tilde{T}_{\cup}$.

Theorem 1.1. *$T_{\perp}$ is a biased and inconsistent estimate of θ, i.e., $E(T_{\perp}) \neq \theta$ even in large samples, unless spikes can be classified with perfect confidence, or cells are Poisson, independent, and have the same firing rates.*

Proofs are in the *Supporting Information, Sec. S4*. Although we proved Theorem 1.1 only for coincidence rates, any statistic computed from spike trains of two or more cells will also suffer from estimation bias under traditional spike sorting, if the distribution of that statistic depends on whether the spikes are in fact independent. This applies to all the common measures of association: correlation, coincidence rate, mutual information, and joint surprise (21), since they all depend on (and try to address) whether or not cells are independent. In practice, the bias will be greater the more waveform clusters overlap, the more cells deviate from Poisson, and the greater the cells' dependence, as illustrated in Figs. 2 and 3. Importantly, the bias does not disappear in large samples—i.e., $T_{\perp}$ is not consistent for θ—so recording more data does not help.

There are two cases when $T_{\perp}$ is unbiased. When waveform clusters do not overlap, spikes can be classified with no errors and subsequent analyses are valid. Otherwise $T_{\perp}$ is biased unless (*i*) cells are independent and Poisson, and their waveforms are independent, and (*ii*) cells have identical firing rates. Under (*i*), spikes and waveforms do not depend on the spike train history, so ignoring it for spike sorting, as traditional sorters do, is the correct assumption. But if (*ii*) is not met, $T_{\perp}$ suffers from another, unrelated, source of bias: When waveform clusters overlap, one can never be perfectly confident that a spike was emitted by a particular cell; we know that some spikes will be misclassified. Therefore, assigning each spike fully (Eq. **3**) to a neuron ignores our uncertainty about spike assignments. Ref. 12 showed that this biases firing rate estimates, unless cells have equal rates. Here, we also show that this biases estimates of neuronal correlations such as $T_{\perp}$; see Fig. 3. Note that $T_{\cup}$ also suffers from that source of bias, since it is calculated from spikes assigned fully (Eq. **5**). But $\tilde{T}_{\cup}$ uses the full posterior distribution of spike identities (Eq. **4**), which removes that source of bias, as shown below.

Theorem 1.2. *$\tilde{T}_{\cup}$ is unbiased; i.e., $E(\tilde{T}_{\cup}) = \theta$, and under weak conditions stated in the proof, its variability can be reduced arbitrarily by increasing the sample size, so $\tilde{T}_{\cup}$ is consistent for θ.*

Theorem 1.2 is valid regardless of how much waveform clusters overlap. In practice, if the overlap is substantial or if the sample is small, $\tilde{T}_{\cup}$ will be relatively variable so it may not match θ closely. However it matches in expectation, i.e., $E(\tilde{T}_{\cup}) = \theta$, and recording more data will help reduce its variability.

Illustrations: Correlation and Coincidence Rate Estimates for Two Dependent Cells. To understand the magnitude of the problem for experimental neuroscience, we estimate the spike count correlation ρ and coincidence rate θ between independent and dependent pairs of simulated Poisson and non-Poisson cells.

A simulation requires (*i*) a spiking model for individual cells, (*ii*) a dependence model for cell pairs, and (*iii*) a model for the waveforms. (*i*) We consider cells that deviate from Poisson according to parameters found empirically by ref. 22. They observed that the ISIs of cortical cells were well fit by gamma distributions with shape parameters k, and corresponding coefficient of variation $\text{CV} = 1/\sqrt{k}$; $k = \text{CV} = 1$ corresponds to Poisson spike trains, while $k > 1$ (CV < 1) models periodicity, and $k < 1$ (CV > 1) "burstiness." Ref. 22 found that k varied widely within cortical areas, with some cells having $k < 1$ and others $k > 1$; and across areas, with median values $k = 1.23$, 1.95, and 2.70 for MT/MST, LIP, and Area 5, respectively. In Fig. 1, we used $k = 0.5$, 1 and 2 to investigate a relevant range. (*ii*) Cells are correlated for various reasons, including common synaptic input, and reciprocal or directed connectivity. Here we consider directed connectivity; results are qualitatively similar under other models. We assume that cell A spikes at constant rate λ_A and cell B at rate λ_B, unless cell A spikes, in which case the rate of cell B is multiplied by β for the next 10 msec. The magnitude of β dictates the cells' degree of association; $\beta = 1$ models independence, $\beta > 1$ models an exci-

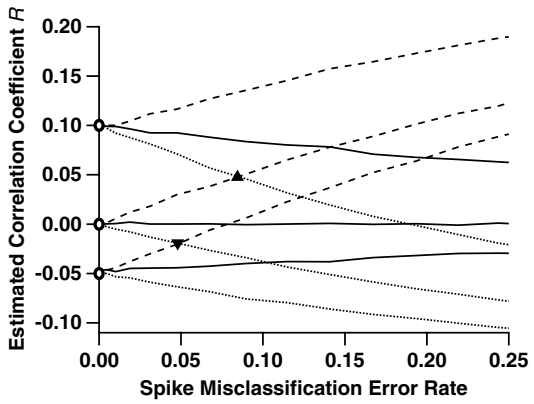

Fig. 2. Correlation estimates of true correlations −0.05, 0, and 0.10, as functions of the rate of misclassified spikes. Pairs of spike trains were simulated according to three different gamma interval processes: $k = 1$ (Poisson, solid lines); $k = 0.5$ (more bursty, dashed); and $k = 2$ (more periodic, dotted). These processes mimic the inter-spike-interval (ISI) distributions seen in experimental data (22), classically characterized by the coefficient of variation $CV = k^{-1/2}$. For each of the three cases, we simulated three different values of true correlation between spike trains: 0.1, 0, and −0.05, marked by the open circles along the *y* axis. Across the *x* axis we vary the rate at which a traditional spike-sorting procedure misclassifies spikes, the arithmetic mean of type 1 and type 2 errors. As this error rate increases, the estimation of correlation diverges from its true value. In the Poisson case, the bias is always towards an estimate of zero correlation. In the bursty case, it is always positive, and in the periodic case it is always negative. Also indicated are the rates of misclassification at which an uncorrelated ($\rho = 0$), bursty ($k = 0.5$), pair of cells has estimated correlation indistinguishable from a correlated ($\rho = 0.1$), periodic ($k = 2$), pair of cells (triangle); and the misclassification rate at which an uncorrelated ($\rho = 0$), periodic pair of cells becomes indistinguishable from an anticorrelated ($\rho = -0.05$), bursty pair of cells (inverted triangle).

tatory effect, and $\beta < 1$ an inhibitory one. For Fig. 2, we selected values of β to generate pairs of uncorrelated cells ($\rho = 0, \beta = 1$), and pairs of positively and negatively correlated cells, with true correlations $\rho = 0.1$ and -0.05. (*iii*) We assume that waveforms are independent, which is warranted when cells have low firing rates. Without loss of generality, we use one-dimensional waveform features that have Gaussian distributions with means 0 and μ for cells A and B, respectively, and unit variances (f_1 and f_2 in Eq. **1**). The overlap between f_1 and f_2 depends on μ, which determines the probability of spike misclassifications. (In the limit $\mu = 0$, clusters overlap completely, so waveforms contain no information that can be used to assign spikes.) In the two cell case, a spike misassignment is a type 1 error for one cell and a type 2 error for the other, so we define the spike misclassification rate to be the average of the rates of these two error types. For example, when $\lambda_A = 4$ Hz, $\lambda_B = 12$ Hz, and $\beta = 2$, then $\mu = 2$ induces a spike misclassification rate of 17.3%. In Fig. 2, we varied μ so that the spike misclassification rate would range from 0 to 25%.

We used this model to simulate the spike trains and waveform features of cell pairs, which we combined to produce electrode spike trains. We then sorted the spikes using traditional spike sorting (*Supporting Information, Sec. S5.2*), and estimated the correlation R between pairs of cells' spike counts in bins of length γ. Fig. 2 shows R against the spike misclassification rate, for $\lambda_A = \lambda_B = 10$ Hz and bin size $\gamma = 25$ ms; we vary these parameters in Fig. 3. The various curves correspond to various degrees of regularity: Poisson ($k = 1$), bursty ($k = 2$), and periodic ($k = 0.5$), and values of β that correspond to true correlations $\rho = 0$ (independent cells), 0.1 and −0.05. Fig. 2 shows that R is biased, unless cells are independent and Poisson ($\rho = 0, k = 1$), or perfectly isolated (zero misclassification rate). Otherwise, the bias increases with the rate of misclassified spikes. When cells are Poisson, R shrinks towards zero regardless of the sign and magnitude of the true correlation ρ; this is consistent with ref. 23; and also with ref. 24, who observed the same with their measure of association, the joint surprise. In the bursty ($k < 1$) and periodic ($k > 1$) cases, R has positive and negative bias, respectively. This is consistent with the findings of refs. 25 and 26 who show that misclassification of spikes from non-Poisson spike trains can lead to misestimation of correlations. Hence the sign and magnitude of the bias depend on the ISI distribution, and thus on the brain area the cells are recorded from. It is also possible to identify correlation when it does not exist; i.e., $\rho = 0$ but $R \neq 0$, and fail to identify it when it does exist; i.e., $\rho \neq 0$ but $R = 0$. Importantly, the bias is sufficiently large that sets of cell pairs whose true correlations ρ are very different from one another, may have the same estimated correlations R once the rate of misclassification reaches 5–10% (Fig. 2, triangles). Since a misclassification rate of 5% or more is not rare (11) (also see *Supporting Information, Sec. 5.1*), our results cast some doubts on the correlation values reported in the neuroscience literature.

Our next simulation serves several purposes. We use the same simulation model, but (*i*) we consider only Poisson cells to investigate more specifically the bias induced by true neuronal correlation rather than by non-Poisson spiking. (*ii*) We investigate the effects of more model parameters. (*iii*) We estimate the coincidence rate θ of cell pairs rather than their correlation ρ, to illustrate that all association measures are biased under traditional spike sorting. (*iv*) We implement ensemble sorting to confirm that $\tilde{T}_\cup$ is indeed unbiased for θ, as proved in Thm.1.2. Detailed descriptions of the traditional and ensemble sorters are in the *Supporting Information, Sec. 5*.

Fig. 3 shows plots of the relative biases of $T_\perp$ and $\tilde{T}_\cup$ against the spike misclassification rate, the degree of dependence between cells via β, the spike rate of cell A, and the relative spike rate of cell B compared with cell A. A *y* axis value of 0 implies an

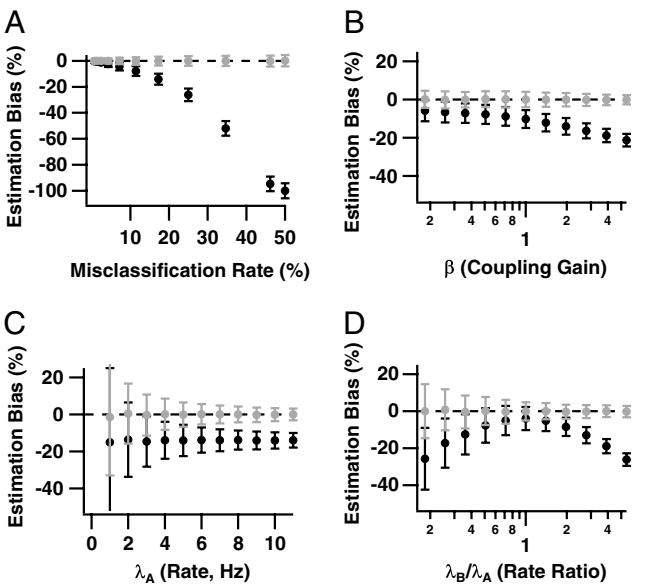

Fig. 3. Relative biases of coincidence rate estimates from traditional and ensemble sorting, $(T_\perp - \theta)/\theta$ (black) and $(\tilde{T}_\cup - \theta)/\theta$ (gray), as functions of (*A*) the spike misclassification rate due to overlapping waveform clusters, (*B*) the coupling strength β, (*C*) the spike rate λ_A, and (*D*) the rate ratio λ_B/λ_A. The *x* axes of *B* and *D* are on a logarithmic scale. In each panel, parameters held constant are $\mu = 2$ (17.3% misclassified spikes), $\lambda_A = 4$ Hz, $\lambda_B/\lambda_A = 3$, and $\beta = 2$. Ensemble sorting (gray) always yields unbiased estimates. Traditional sorting (black) never does, except when cells are independent and have equal rates (Theorem 1.1), or when waveform clusters are fully separated (Theorem 1.1 and Fig. 3*A*, zero error rate). Otherwise, the bias of $T_\perp$ increases as the waveform clusters overlap (Fig. 3*A*), the coupling strength increases (Fig. 3*B*), and the firing rates differ (Fig. 3*D*).

unbiased coincidence rate estimate; -1 means that all coincidences are missed; 1 means that the estimate is twofold too large. Fig. 3 shows that the ensemble coincidence rate estimate $\tilde{T}_{\cup}$ is unbiased for all configurations of the model parameters. The traditional estimate $T_{\perp}$ never is, and its bias appears to be a complex function of the model parameters. Fig. 3*B* displays clearly the two sources of bias—full spike assignments (Eq. **3**) and independent sorting – that plague $T_{\perp}$. The cells are independent when $\beta = 1$, yet $T_{\perp}$ underestimates θ; this is the effect of full spike assignments, which we know (12) becomes more severe as the cells' firing rates diverge from one another (here $\lambda_B/\lambda_A = 3$). As $|\beta - 1|$ increases, which strengthens the association between cells, the bias arising from independently sorting correlated spikes combines with the first source of bias. Consider also Fig. 3*D*, which plots relative bias versus λ_B/λ_A for a fixed degree of association $\beta = 2$. When firing rates are equal ($\lambda_B/\lambda_A = 1$), and when, as here, the waveform feature distributions are identical apart from a shift in their mean, it is easy to show that traditional spike sorting misclassifies the same average number of spikes for both cells. In that case, full assignments (Eq. **3**) do not bias the coincidence rate estimate, so that the negative bias of $T_{\perp}$ at $\lambda_B/\lambda_A = 1$ must be due to ignoring the dependence between cells when spike sorting. Then as $|\lambda_B/\lambda_A - 1|$ increases, that bias is compounded with an increasingly large bias from full spike assignments. Considering Fig. 3 overall, we see that the bias of $T_{\perp}$ is largest when the cells have overlapping waveform clusters, have very different firing rates, and are strongly correlated.

Next, we illustrate that biased estimates not only corrupt scientific inference and interpretation, they also degrade our ability to detect correlations. We consider again the many datasets we simulated to produce Fig. 3*B*, and for each, we test the null hypothesis that the cells are independent at the 0.05 significance level. This consists of comparing the observed coincidence rate against the rate expected under independence, and calculating a p-value. We then estimate the power of the test—i.e., the probability that it detects correlated activity—by the proportion of times that a significant p-value (<0.05) is obtained.

Fig. 4*A* shows the power functions of the two tests based on $T_{\perp}$ and $\tilde{T}_{\cup}$ as functions of $\beta \geq 1$, the degree of positive association between the cells, for $\lambda_A = 4$ Hz, $\lambda_B/\lambda_A = 3$, and $\mu = 2$ (17.3% misclassified spikes). Results are similar for other model settings. We see that ensemble sorting provides uniformly more power. For example, when $\beta = 2$, the power to detect that the cells are dependent is approximately 35% using $T_{\perp}$ and 90% using $\tilde{T}_{\cup}$.

Note that when the null hypothesis of independence is true ($\beta = 1$), both tests reject the null hypothesis approximately 5% of the time. Hence, while the test based on $T_{\perp}$ lacks power to detect existing associations, it does have the correct rate of false positive (spurious) detections. This should not be surprising here: Traditional spike sorting assumes that spikes are independent, which matches the data in this illustration at $\beta = 1$, since we simulated Poisson cells. When cells are not Poisson, we know from Thm.1.1 and Fig. 2 that $T_{\perp}$ is biased even when cells are independent, which necessarily biases the rate of spurious detections. In contrast, the test based on $\tilde{T}_{\cup}$ retains the correct detection rate since $\tilde{T}_{\cup}$ is unbiased.

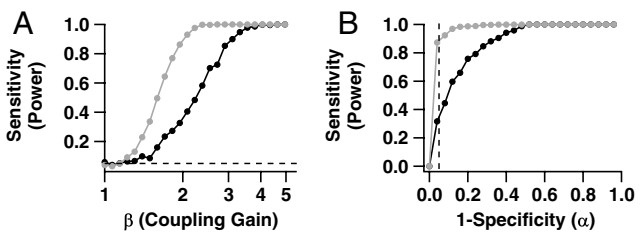

Fig. 4. Bias in coincident spike rate estimates translates into loss of power to detect correlated cells. (*A*) Sensitivity (power) of tests for coincident spiking as function of coupling strength β. Other model parameters are held at $\mu = 2$, $\lambda_A = 4$ Hz, and $\lambda_B/\lambda_A = 3$. The light and dark curves correspond respectively to the tests based on ensemble and traditional estimates, $\tilde{T}_{\cup}$ and $T_{\perp}$, of the coincidence rate θ. The test based on ensemble sorting is more powerful. (*B*) Receiver-operating characteristic (ROC) curves show that for any choice of the significance level α, the sensitivity of the coincident spiking test is higher when using ensemble sorting. The vertical dashed line is at the most commonly used significance level of 5%.

Another common summary of the power of a test is the receiver-operating characteristic (ROC) curve, which plots power (sensitivity, the probability of rejecting a null hypothesis when it is false) versus significance level (1—specificity, the probability of rejecting a null hypothesis when it is true). A desirable test has high sensitivity for all specificity. Fig. 4*B* shows the ROC curves of the two tests based on $T_{\perp}$ and $\tilde{T}_{\cup}$. The vertical dashed line is at the commonly used significance level of 5%. Again, the test based on $\tilde{T}_{\cup}$ has uniformly better power.

Discussion

Several authors have reported that spike misclassifications lead to misestimation of associations between cells. For example, refs. 23 and 24 found that positive correlations were underestimated by both overly liberal and overly conservative spike-sorting procedures, refs. 25 and 26 noted that cells deviating from Poisson also impacted correlation estimates. In this paper, we proved that correlation estimates calculated from spike trains sorted using current methods are biased, unless cells are isolated, or they are Poisson, independent, and have equal rates. Here and previously (12, 14) we argued that this bias occurs because the standard paradigm—spike sorting first, statistical analyses second—exploits the information that spike sorting provides about the quantities of interest (e.g., correlations here, tuning curves in refs. 12 and 14), but ignores the reverse that these quantities also provide information about spike identities. The correct approach is to consider these relationships simultaneously rather than sequentially. Our solution is "ensemble sorting," which consists of sorting spikes using waveforms as usual, while also accounting for the interdependent spiking of the cell ensemble. We proved that resulting correlation estimates are unbiased, even when spikes cannot be sorted with perfect confidence.

Ensemble sorting involves significant conceptual and practical shifts for spike sorting: Spikes must be sorted jointly, and waveforms must be retained in case some later analysis requires resorting spikes under a different joint spiking model. This is more cumbersome than the current practice of spike sorting once and never considering the waveforms again. But our simulations motivate a serious consideration of ensemble sorting. Indeed, unless the analyst retains only perfectly sorted cells, the size of the biases can be scientifically significant and corrupt inferences about coincident spiking. Bias can also translate into a sizable loss of power to detect existing correlations in data. In contrast, ensemble sorting yields unbiased estimates even for poorly isolated cells, which permits the analysts to retain more cells for analysis. (One might still exclude some poorly isolated cells when the variances of estimates are too high.) Therefore, experimental investigations will be more efficient: Fewer experiments are needed to accurately infer the degree of correlated activity in a neural system; and when the number of experiments is fixed, these inferences are more accurate.

We stress that the bias discussed in this paper is unrelated to the biases that can arise from implementation issues such as choosing poor models for waveform features, having difficulties estimating the spike sorter from data, and determining the number of cells recorded by electrodes. Our results prove that if spikes are sorted independently of other spikes, the bias will not vanish even if an otherwise optimal spike sorter is applied with the true models. But if the correct models are used with ensemble sorting, correlation estimates will be unbiased. The bias discussed in this

paper arises because current spike sorters ignore some information about spike identities. Our results should therefore be of interest to all experimentalists, since simultaneous recording of many cells has become the rule rather than the exception in experiments (15). Ref. 12 showed that firing rate estimates are biased. Here we showed that coincident spiking estimates are biased. Estimates of other statistical quantities are also presumably biased, and it would be worth investigating when these biases are scientifically significant.

In this paper, we formulated the concept and framework for ensemble sorting, which relies on the posterior probabilities of spike identities in Eq. **6**. To obtain Eq. **6**, one must specify a model for potentially nonstationary waveforms, and a model for the cells' firing rates, conditional on the previous activity of the cell ensemble—i.e., conditional on when past spikes were emitted and on which cells emitted them. Such conditional rate models are routinely used in the analysis of neural data (27). Finally, one must also determine the ISI distribution induced by the conditional rate model. When spikes are independent, this distribution is exponential. Otherwise it must be derived analytically based on the properties of the assumed joint spiking model. The *Supporting Information, Sec. S5.3* contains that derivation for the unidirectional coupling two-cell model application in Figs. 3 and 4, and shows that the ISI distribution involves mixtures of truncated and shifted exponential distributions. We have not provided the form of the ISI distribution for a network involving an arbitrary number of cells, but we are confident we can do this in future work using the same framework.

Once all its components are specified, Eq. **6** must also be fitted to data. This requires that models for waveforms and for ensemble spiking be specified together. The traditional approach requires the same components, but they are used separately: The waveform model is used for spike sorting first, the joint spiking model for statistical analyses second. The important point is that ensemble sorting requires no more model assumptions than the traditional approach. Moreover, ensemble sorting makes consistent assumptions across spike sorting and data analysis—the traditional approach does not—which is why the former alone yields unbiased analyses. Since Eq. **6** is a mixture distribution, we are confident that estimation can be done by maximum likelihood via an expectation-maximization (EM) algorithm. We implemented a related algorithm (14), as well as the algorithm needed to estimate the ensemble sorter for the connectivity model of the illustration section; the detailed implementation will be reported elsewhere.

We have not addressed any computational issues in this paper. They include choosing appropriate waveform and spiking models, choosing initial values of the EM algorithm, and determining the number of neurons. The same issues arise in traditional spike sorting and analyses. The methods proposed here are conceptual; they are not designed to resolve implementation issues. When we implement the proposed procedures to data in future work, we will choose currently available models for waveforms and joint spiking, we will fit the joint spike-sorting rule (Eq. **6**) using an EM algorithm, we will determine the number of cells using penalized likelihood, and test the goodness of model fits by penalized likelihood or likelihood ratio tests (14).

Finally, the results in this paper concern cells recorded on the same electrode. But if cells recorded on different electrodes do not behave independently, these electrodes may contain information about the spike identities on other electrodes, and ignoring that information will bias correlation estimates. To see this, assume that cells A and B are recorded on an electrode, cell C on another, that A is independent of B and C, but that B and C always spike together. Assume that A and B fire at equal rates and have the same spike waveforms. If we sort the electrodes separately, we will assign spikes to A and B at random, with spike misclassification 50%; then estimates of correlation between A and C, and between B and C, are inflated and deflated, respectively. But if we sort the electrodes jointly, we will (correctly) assign to B all the spikes that occur concurrently with a spike from C, and to A all the other spikes, with spike misclassification rate 0%; then correlation estimates are unbiased. This is related to refs. 28 and 23, who report that for imperfect spike classification, the degree of true correlation of cells recorded on different electrodes will influence estimates of the degree of correlation of cells recorded on the same electrode.

ACKNOWLEDGMENTS. V.V. was supported by National Institutes of Health Grants 2R01MH064537 and 1R01EB005847. R.C.G. was supported by National Institutes of Health Grant F32DC010535.

1. Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1999) *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA).
2. Engel AK, Singer W (2001) Temporal binding and the neural correlates of sensory awareness. *Trends Cogn Sci* 5:16–25.
3. Bi GQ, Poo MM (2001) Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu Rev Neurosci* 24:139–66.
4. Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370:140–143.
5. Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7:358–366.
6. Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. *Nat Rev Neurosci* 2:539–550.
7. Pillow JW, et al. (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454:995–999.
8. Graf ABA, Kohn A, Jazayeri M, Movshon JA (2011) Decoding the activity of neuronal populations in macaque primary visual cortex. *Nat Neurosci* 14:239–245.
9. Stevenson IH, Kording KP (2011) How advances in neural recording affect data analysis. *Nat Neurosci* 14:139–142.
10. Lewicki MS (1998) A review of methods for spike sorting: The detection and classification of neural action potentials. *Network* 9:R53–78.
11. Schmitzer-Torbert N, Jackson J, Henze D, Harris K, Redish AD (2005) Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* 131(1):1–11.
12. Ventura V (2009) Traditional waveform based spike sorting yields biased rate code estimates. *Proc Natl Acad Sci USA* 106:6921–6.
13. Staude B, Grün S, Rotter S (2010) Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference. *Front Comput Neurosci Jul 2* 4. pii:16.
14. Ventura V (2009) Automatic spike sorting using tuning information. *Neural Comput* 21:2466–2501.
15. Brown EN, Kass RE, Mitra PP (2004) Multiple neural spike trains analysis: state-of-the-art and future challenges. *Nat Neurosci* 7:456–461.
16. Quiroga R-Q (2007) Spike sorting. *Scholarpedia* 2:3583.
17. Wasserman L (2004) *All of Statistics* (Springer, New York).
18. Daley DJ, Vere-Jones D (1988) *Introduction to the Theory of Point Processes* (Springer, New York).
19. Fee MS, Mitra PP, Kleinfeld D (1996) Variability of extracellular spike waveforms of cortical neurons. *J Neurophysiol* 76:3823–3833.
20. Pouzat C, Delescluse M, Voit P, Diebolt J (2004) Improved spike-sorting by modeling firing statistics and burst-dependent spike amplitude attenuation: a Markov chain Monte Carlo approach. *J Neurophysiol* 91:2910–2928.
21. Grün S (2009) Data-driven significance estimation for precise spike correlation. *J Neurophysiol* 101:1126–1140.
22. Maimon G, Assad JA (2009) Beyond Poisson: Increased spike-time regularity across primate parietal cortex. *Neuron* 62:426–40.
23. Gerstein GL (2000) Cross-correlation measures of unresolved multi-neuron recordings. *J Neurosci Methods* 100:41–51.
24. Pazienti A, Grün S (2006) Robustness of the significance of spike synchrony with respect to sorting errors. *J Comput Neurosci* 21:329–42.
25. Ecker AS, et al. (2010) Decorrelated neuronal firing in cortical microcircuits. *Science* 327:84–587.
26. Cohen MR, Kohn A (2011) Measuring and interpreting neuronal correlations. *Nat Neurosci* 14:811–9.
27. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (Feb 2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *J Neurophysiology* 93:1074–1089.
28. Bedenbaugh P, Gerstein GL (1997) Multiunit normalized cross correlation differs from the average single-unit normalized correlation. *Neural Comput* 9:1265–75.

Supporting Information

Ventura and Gerkin 10.1073/pnas.1115236109

SI Text

S1 Traditional Spike Sorting. To keep our presentation clear, we made simplifying assumptions. Eq. **1** in the main text

$$f(a) = \sum_{i=1}^{I} \pi_i f_i(a). \qquad [S1]$$

suggests that each spike belongs to one and only one cell i, since the number of summands is equal to the number of cells I. However, we could add some components to that mixture to account for spikes emitted by several cells firing approximately together. Similarly, in the main text we assumed that coincident spikes never occurred exactly together, so that waveforms never overlapped. This assumption could easily be relaxed, but would require additional notation without adding illustrative value. The results presented here are general, and so this distinction does not matter.

We also assumed for simplicity that neurons had constant firing rates. In the case when neurons' firing rates are modulated by covariates c such as environmental parameters or experimental time, c contains information about spikes identities, and using constant π_i's in Eq. **1** effectively consists of ignoring that information (1). This in turn yields a suboptimal spike sorter, and biased estimates of tuning curves. This presumably also biases estimates of correlated activity, although we did not provide proof of that in this paper, because it was not our primary focus.

If cells' firing rates are modulated by covariates, the optimally efficient approach is to obtain the joint posterior probabilities of spike identities, given all the variables that contain information about these identities; i.e., $P(\underline{X} = \underline{i}|\underline{a}, \underline{s}, c)$ instead of $P(\underline{X} = \underline{i}|\underline{a}, \underline{s})$ in Eq. **4**. For the traditional spike sorter, Ventura (2) showed that calculating $P(X = i|a, c)$ consists of applying the same spike sorting procedure outlined for traditional sorting in the main text, but with the constant π_i in Eq. **1** replaced by functions $\pi_i(c)$ that depend on the cells' firing rates. We will address the inclusion of rate modulating covariates c for ensemble sorting in future work.

S2 Ensemble Sorting. ***S2.1 Relationship to traditional sorting.*** In the main text, we pointed to similarities and differences between traditional and ensemble sorting. For example, we noted that the posterior probabilities in Eqs. **4** and **2**, and the spike assignments in Eqs. **5** and **3**, had the same structures. Eqs. **2** and **3** apply to a single spike, while Eqs. **4** and **5** apply jointly to n spikes, and are further conditioned on the spike train's history $\underline{s}$. Here we show that ensemble sorting reduces to traditional sorting when neurons are Poisson and independent, and their waveforms are independent.

Ensemble sorting is based on Eq. **6** in the main text:

$$P(X_1 = i_1, \ldots, X_n = i_n|\underline{a}, \underline{s}) = \frac{\pi_{\underline{i}}(\underline{s}) f_{\underline{i}}(\underline{a}|\underline{s})}{f(\underline{a}|\underline{s})}.$$

where $\pi_{\underline{i}}(\underline{s}) = P(X_1 = i_1, \ldots, X_n = i_n|\underline{s}, N = n)$ is the probability that the n spikes were emitted by cells with identities $\underline{i} = (i_1, i_2, \ldots, i_n)$, $f_{\underline{i}}(\underline{a}|\underline{s}) = f(\underline{a}|\underline{i}, \underline{s})$ is the joint distribution of their waveform measurements, and

$$f(\underline{a}|\underline{s}) = \sum_{i_1=1}^{I} \sum_{i_2=1}^{I} \cdots \sum_{i_n=1}^{I} \pi_{\underline{i}}(\underline{s}) f_{\underline{i}}(\underline{a}|\underline{s})$$

is the unconditional joint distribution of the waveforms (unconditional, although all probabilities and densities are conditioned on $\underline{s}$).

To relate ensemble to traditional sorting, consider the situation where the n spikes and their waveforms are independent. Then the history of the electrode spike train $\underline{s}$ contains no information about spike identities, so that conditioning on $(\underline{a}, \underline{s})$ is equivalent to conditioning only on $\underline{a}$. Then the denominator of Eq. **6**, $f(\underline{a}|\underline{s})$, reduces to

$$\begin{aligned} f(\underline{a}) &= \sum_{i_1=1}^{I} \cdots \sum_{i_n=1}^{I} \frac{P(X_1 = i_1) \ldots P(X_n = i_n)}{\sum_{j_1=1}^{I} \cdots \sum_{j_n=1}^{I} P(X_1 = j_1) \ldots P(X_n = j_n)} \\ &\quad \times f_{i_1}(a_1) \ldots f_{i_n}(a_n) = \left(\sum_{i_1=1}^{I} \frac{P(X_1 = i_1)}{\sum_{j_1=1}^{I} P(X_1 = j_1)} f_{i_1}(a_1) \right) \\ &\quad \cdots \left(\sum_{i_n=1}^{I} \frac{P(X_n = i_n)}{\sum_{j_n=1}^{I} P(X_n = j_n)} f_{i_n}(a_n) \right) \\ &= \prod_{k=1}^{n} \left(\sum_{i=1}^{I} \frac{P(X_k = i)}{\sum_{j=1}^{I} P(X_k = j)} f_i(a_k) \right) = \prod_{k=1}^{n} \left(\sum_{i=1}^{I} \pi_i f_i(a_k) \right) \\ &= \prod_{k=1}^{n} f(a_k), \end{aligned}$$

which is the product of the marginal distributions of the waveforms in Eq. **1** in the main text, as we should expect under independence. It is equally easy to show that, under independence, $P(X_1 = i_1, \ldots, X_n = i_n|\underline{a}, \underline{s})$ reduces to the product of the marginals, $\prod_{j=1}^{n} P(X_j = i_j|a_j)$, which implies that ensemble sorting reduces to traditional sorting as described in the main text.

S2.2 Waveform and joint spiking models. To specify the spike sorting scheme for a group of $n \geq 1$ spikes, we need to specify Eq. **6** in the main text:

$$P(\underline{X} = \underline{i}|\underline{a}, \underline{s}) = \frac{f(\underline{a}|\underline{i}, \underline{s}) P(\underline{X} = \underline{i}|\underline{s})}{f(\underline{a}|\underline{s})},$$

which are the joint probabilities that the spike IDs have value $\underline{i} = (i_1, \ldots, i_n)$, given all waveforms $\underline{a} = (a_1, \ldots, a_n)$ and ISIs $\underline{s} = (s_1, \ldots, s_n, s_{n+1})$. We specify the three components of Eq. **6** below.

- The denominator, $f(\underline{a}|\underline{s})$, is evaluated as the sum over all $\underline{i}$ of the numerator.
- The first term in the numerator, $f_{\underline{i}}(\underline{a}|\underline{s})$, is the joint distribution of n observed waveforms $\underline{a}$ fired by cells $\underline{i}$, and with ISIs $\underline{s}$. It is determined by the properties of the spike waveforms. For example, if waveforms are independent and stationary, they do not depend on the history of the electrode spike train. Then $f_{\underline{i}}(\underline{a}|\underline{s}) = f_{\underline{i}}(\underline{a})$, which further reduces to the product of the marginals, $\prod_{j=1}^{n} f_{i_j}(a_j)$, where f_i is the distribution of waveforms emitted by cell i that appeared in Eqs. **1** and **2** in the main text. Typically, f_i are assumed to be Gaussian distributions.
  If waveforms are attenuated after short ISIs, then the condition on $\underline{s}$ cannot be dropped. However, waveforms are independent conditional on $\underline{s}$ and $\underline{i}$, so we can write $f_{\underline{i}}(\underline{a}|\underline{s}) = f(\underline{a}|\underline{i}, \underline{s}) = \prod_{j=1}^{n} f(a_j|\underline{i}, \underline{s})$, and design appropriate

attenuation models for the individual components. To illustrate this, imagine that an electrode records $I = 2$ cells, and that we recorded only two spikes with identities (i_1, i_2), waveforms (a_1, a_2), and ISIs (s_1, s_2, s_3). Imagine that s_1 is long enough so that the first observed waveform a_1 is not attenuated, but that s_2 is short enough that the second waveform would be attenuated in the event that the same cell emitted both spikes. This means that $f(a_1|\underline{i}, \underline{s}) = f(a_1|i_1) = f_{i_1}(a_1)$, $i_1 = 1, 2$, does not depend on the past. Similarly, $f(a_2|\underline{i} = (1, 2), \underline{s}) = f_2(a_2)$ and $f(a_2|\underline{i} = (2, 1), \underline{s}) = f_1(a_2)$ do not depend on the past either, given that the two spikes were emitted by different cells. However, $f(a_2|\underline{i}, \underline{s})$ depends on s_2 when $\underline{i} = (1, 1)$ or $(2, 2)$. A possible model was proposed by Pouzat et al. (3): They used the amplitudes of the waveforms as features, assumed that they were normally distributed with different means for each cell, and used an exponential relaxation to model attenuation of waveform amplitude. In our small example, it means that $f(a_1|\underline{i}, \underline{s}) = N(\mu_{i_1}, \sigma^2)$, $f(a_2|\underline{i} = (i_1, i_2), \underline{s}) = N(\mu_{i_2}, \sigma^2)$ when $i_1 \neq i_2$, and $f(a_2|\underline{i} = (i_1, i_2), \underline{s}) = N(\mu_{i_2}(1 - ae^{-bs_2}), \sigma^2)$ when $i_1 = i_2$, where a and b are positive constants.

- The other term we must specify in Eq. **6** is $\pi_{\underline{i}}(\underline{s}) = P(X_1 = i_1, \ldots, X_n = i_n|\underline{s})$, the probability that the n spikes were fired by cells $\underline{i} = (i_1, i_2, \ldots, i_n)$, given that the electrode history is $\underline{s}$. It is determined by the cells' ensemble spiking. The simplest case is that of independent and Poisson cells. Then the spikes' identities do not depend on the past, so that $\pi_{\underline{i}}(\underline{s}) = \pi_{\underline{i}} = \prod_{j=1}^{n} \pi_{i_j}$, where π_i is the proportion of spikes from cell i that appeared earlier in Eqs. **1** and **2** in the main text. Note that when spikes and waveforms are independent, ensemble sorting reduces to traditional spike sorting, as shown in the previous section.

  When spikes are not independent, we must specify a joint spiking model, and derive $\pi_{\underline{i}}(\underline{s})$ given that model. We first rewrite $\pi_{\underline{i}}(\underline{s}) = P(\underline{X} = \underline{i}|\underline{s})$ as $P(\underline{X} = \underline{i}, \underline{s})/f(\underline{s})$, where $f(\underline{s})$ is obtained as the sum over $\underline{i}$ of $P(\underline{X} = \underline{i}, \underline{s})$, and we condition iteratively to obtain

$$\begin{aligned} P(\underline{X} = \underline{i}, \underline{s}) = & f(s_{n+1}|X_n = i_n, s_n, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, s_1) \\ & \cdot P(X_n = i_n|s_n, X_{n-1} = i_{n-1}, s_{n-1}, \ldots, X_1 = i_1, s_1) \\ & \cdot f(s_n|X_{n-1} = i_{n-1}, s_{n-1}, X_{n-2} = i_{n-2}, \ldots, X_1 = i_1, s_1) \\ & \vdots \\ & \cdot P(X_1 = i_1|s_1) \cdot f(s_1). \end{aligned} \quad \textbf{[S2]}$$

  We thus need to determine

$$P(X_j = i_j|s_j, X_{j-1} = i_{j-1}, s_{j-1}, \ldots, X_1 = i_1, s_1), \quad j = 1, \ldots, n$$

  the probability that spike j has identity i_j, given the IDs of all previous spikes and the ISIs up to spike j. This probability is given by the conditional spiking rate function of neuron with identity i_j, given the previous spikes of the neuron ensemble. Such conditional rate models are routinely used in the analysis of neural data (4). See also the analytic expression for $\pi_{\underline{i}}(\underline{s})$ in Supporting Information, Sec. S5.3 for the illustration in the main text.

  We also need to determine

$$f_{S_j|X_{j-1}, S_{j-1}, X_{j-2}, \ldots, X_1, S_1}(s), \quad j = 1, \ldots, n+1$$

  the ISI distribution given past ISIs and spike identities. When spikes are independent, this reduces to an exponential distribution. Otherwise it must be derived analytically based on the properties of the assumed joint spiking model. Supporting Information, Sec. S5.3 contains that derivation for the illustration in the main text, which shows that the solution involves mixtures of truncated and shifted exponential distributions. The general case will be obtained similarly.

S3 Estimation of Coincident Spiking. The main conceptual contribution of this paper is the idea of ensemble sorting: Because the history of an electrode spike train contains information about spikes identities, spikes should be sorted jointly rather than one at a time, otherwise estimates of coincident spiking will be biased. But there is another, unrelated, source of bias that afflicts statistical estimates calculated from sorted data, not just estimates of coincident spiking. This is discussed below.

S3.1 Full versus probabilistic spike assignments. Traditional sorting assigns each spike fully to one cell. When model based spike sorting is used, as we do in this paper, each spike is typically assigned fully to the one neuron that has the highest probability of having emitted the spike (Eq. **3**). Then the probabilities that the other neurons emitted that spike (Eq. **2**) are discarded for subsequent analyses. But when waveform clusters overlap, one can never be perfectly confident that a spike was emitted by a particular cell; we know that some spikes will be misclassified. Therefore, assigning each spike fully to a neuron ignores our uncertainty about spike assignments. Ventura (1) showed that this induces bias in estimates of firing rates, unless neurons have equal firing rates. Ignoring that uncertainty biases other estimates. For example in Fig. 2*B*, it is clear that estimated and true coincidence rates do not match in expectation; i.e., $E(T_\perp) \neq \theta$, even when neurons are Poisson and independent ($\beta = 1$).

To remove that source of bias, one must preserve the uncertainty about spike assignments. We do that by making partial assignments of spikes to neurons according to the posterior distribution of spike identities, instead of using only the largest posterior probability. For example, consider an electrode that records three neurons. Then each spike could have been emitted by one of the three neurons. Imagine that a spike with waveform a was observed, and that the posterior probabilities that this spike was emitted by neuron $i = A$, B, and C are $P(X = i|a) = 0.1, 0.3$, and 0.6, respectively. Typically, the spike is assigned to the neuron that is most likely to have emitted it; i.e., neuron C in this case. But neurons A and B also have relatively high probabilities of having emitted that spike. Probabilistic ("soft" in Ventura's, refs. 1 and 2, terminology) assignments consist of allocating 0.1, 0.3, and 0.6 partial spikes to the three neurons, respectively. Similarly, imagine that we want to estimate the spike count correlation between two neurons. Typically, spike counts from full spike assignments are used for that calculation. Let N be the number of spikes from cell A in some bin: We cannot be certain that this number is correct, because some spikes will be misclassified even by the best spike sorter when waveform clusters overlap. Instead of N, we use the posterior expected number of spikes from cell A, because that number summarizes the uncertainty in spike assignments. The estimators we recommend are based on that principle. We illustrate this below for the coincidence rate θ and the spike count correlation ρ.

S3.2 Estimates of the coincidence rate θ. In the main text, we explained how to calculate $T_\perp$ and $\tilde{T}_\cup$ without using mathematical formulas. These formulas are much less intuitive than the description we gave, but we need them to prove Theorems 1.1 and 1.2.

Let θ denote the probability that two cells both spike in time bins of length γ. The usual estimate for θ is the observed proportion of bins in which these cells emitted at least 1 spike each. This is written as

$$T_{\perp} = M^{-1} \sum_{m=1}^{M} \mathbf{I}\{(x_{m1}, \ldots, x_{mn_m}) \in \mathscr{C}_m\}, \qquad \textbf{[S3]}$$

where $\mathbf{I}\{A\} = 1$ if A is true and 0 otherwise, M is the number of bins, m indexes them, n_m is the number of spikes in bin m, and $\mathscr{C}_m = \{(i_{m1}, \ldots, i_{mn_m})$ such that $(i_{ml}, i_{mk}) = (1, 2)$ for some $(l, k) \in \{1, \ldots, n_m\}^2\}$ is the set of all joint spike identities that code for cells 1 and 2 both spiking in bin m. Note that we modified our notation: Previously, x_k denoted the identity of spike k, $k = 1, \ldots, n$; x_{mk} now denotes the identity of the k th spike in bin m, $k = 1, \ldots, n_m$. Eq. **S3** also applies when spikes are sorted jointly per Eq. **5**, which yields the estimate we called $T_{\cup}$ in the main text.

The estimator we recommend, $\tilde{T}_{\cup}$, is the average over bins of the posterior probabilities of observing at least one spike from each cell:

$$\tilde{T}_{\cup} = M^{-1} \sum_{m=1}^{M} P((X_{m1}, \ldots, X_{mn_m}) \in \mathscr{C}_m | \underline{a}, \underline{s}) \qquad \textbf{[S4]}$$

where $P((X_{m1}, \ldots, X_{mn_m}) \in \mathscr{C}_m | \underline{a}, \underline{s})$ is the joint posterior probability function of the identities of the n_m spikes in bin m, obtained by summing the full joint posterior in Eq. **6** over all variables, except $(X_{m1}, \ldots, X_{mn_m})$. Eqs. **S3** and **S4** have the same structure; Eq. **S3** uses spikes that were each assigned fully to a neuron, whereas Eq. **S4** preserves the uncertainty in determining spike identities.

To make sense of Eqs. **S3** and **S4**, consider the spike train sketch in Fig. 1 of the main text. We observed $N = 5$ spikes with waveforms $\underline{a}$ and history $\underline{s}$ on an electrode that records $I = 3$ cells. The spikes' joint identities can be one of $I^n = 3^5$ possibilities, with posterior probabilities (Eq. **6** in the main text) given in Fig. 1. These values are for illustrative purposes only, and did not result from actually applying our spike sorter to data; see the illustration in the main text and sec. S5 below for an application of our spike sorter to simulated data. To estimate θ, we cut the recording time into $M = 2$ bins of width γ, as shown in Fig. 1. The first two spikes belong to the first bin, the other three to the second bin.

We first calculate $T_{\cup}$. The largest joint posterior probability for the 5 recorded spikes is $P((X_1, X_2, X_3, X_4, X_5) = (A, A, B, A, A)|\underline{a}, \underline{s}) = 0.4$, so joint spike sorting (Eq. **5**) assigns the third spike to cell B and all others to cell A. Only the second bin contains spikes from cells A and B, so $T_{\cup} = 1/2$. We now verify that Eq. **S3** gives the same answer. Bin 1 has $n_1 = 2$ spikes, $\mathscr{C}_1 = \{(A, B), (B, A)\}$ are the possible identities that code for a coincidence, and the identities of the first two spikes are $(x_{11}, x_{12}) = (A, A)$. Since $(x_{11}, x_{12}) \notin \mathscr{C}_1$, the contribution of bin 1 to $T_{\cup}$ is $\mathbf{I}\{(x_{11}, x_{12}) \in \mathscr{C}_1\} = 0$. Bin 2 has $n_2 = 3$ spikes, and $\mathscr{C}_2 = \{$all permutations of$(A, B, k), k = A, B, C\}$. The identities of the three spikes in bin 2, $(x_{21}, x_{22}, x_{23}) = (B, A, A)$, are in $\mathscr{C}_2$, so the contribution of bin 2 to $T_{\cup}$ is $\mathbf{I}\{(x_{21}, x_{22}, x_{23}) \in \mathscr{C}_2\} = 1$. Hence $T_{\cup} = 1/2$, as needed.

Next, we make sense of Eq. **S4** to calculate $\tilde{T}_{\cup}$. The contribution to $\tilde{T}_{\cup}$ of the first bin is $P((X_{11}, X_{12}) \in \mathscr{C}_1 | \underline{a}, \underline{s}) = P((X_{11}, X_{12}) = (A, B)$ or $(B, A)|\underline{a}, \underline{s}) = 0.13 + 0.1 = 0.23$. The contribution of the second bin is $P((X_{21}, X_{22}, X_{23}) \in \mathscr{C}_2 | \underline{a}, \underline{s}) = P((X_{21}, X_{22}, X_{23}) =$ any permutation of$(A, B, k), k = A, B, C|\underline{a}, \underline{s}) = 0.4 + 0.13 + 0.07 + 0.06 + 0.04 = 0.7$.
Hence $\tilde{T}_{\cup} = (0.7 + 0.23)/2 = 0.465$.

S3.3 Estimates of the spike count correlation ρ. Perhaps the most common measure of association in the activities of two cells is the spike count correlation ρ. Fig. 2 in the main text shows that the sample correlation is biased when spike identities are determined by traditional spike sorting. Here, we define an unbiased estimator of ρ based on the joint posterior probabilities of spike identities.

The true correlation ρ between the spike counts of two neurons in bins of specified lengths is

$$\rho = \frac{E(N_A N_B) - E(N_A)E(N_B)}{\sigma(N_A)\sigma(N_B)}, \qquad \textbf{[S5]}$$

where $N_X, X = A, B$ denotes the number of spikes emitted by cells A and B in a bin, $E(N_X)$ denotes its true mean, and $\sigma^2(N_X) = E(N_X^2) - E^2(N_X)$ its true variance.

The usual estimate of ρ is the sample correlation, which is obtained by replacing the true expectations in Eq. **S5** by their estimates in the data. That is,

$$R = \frac{\overline{N_A N_B} - \bar{N}_A \bar{N}_B}{\sqrt{\overline{N_A^2} - \bar{N}_A{}^2}\sqrt{\overline{N_B^2} - \bar{N}_B{}^2}}, \qquad \textbf{[S6]}$$

where the estimates of the various expectations are

True value	Replaced by Estimate	Sample mean over bins of...
$E(N_X)$	$\bar{N}_X = M^{-1}\sum_{m=1}^{M} N_{X,m}$	The spike count for cell X
$E(N_{X_i}N_{X_j})$	$\overline{N_{X_i}N_{X_j}} = M^{-1}\sum_{m=1}^{M}(N_{X_i,m}N_{X_j,m})$	The product of spike counts for cells X_i and X_j
$E(N_X^2)$	$\overline{N_X^2} = M^{-1}\sum_{m=1}^{M} N_{X,m}^2$	The squared spike count for cell X

For the example in Fig. 1 in the main text, the most likely set of spike identities assignments—i.e., the set that maximizes the joint posterior probability in Eq. **6**—is for the 3rd spike to be assigned to cell B and all 4 other spikes to cell A. Then the first bin (Fig. 1) contains two spikes from cell A and none from cells B and C, and the second bin contains 2 spikes from cell A, 1 from cell B, and none from cell C.

Therefore, $E(N_A)$ is estimated by $\bar{N}_A = 2^{-1}(2 + 2) = 2$
$E(N_B)$ is estimated by $\bar{N}_B = 2^{-1}(0 + 1) = 1/2$
$E(N_A N_B)$ is estimated by $\overline{N_A N_B} = 2^{-1}(2 \times 0 + 2 \times 1) = 1$
$E(N_A^2)$ is estimated by $\overline{N_A^2} = 2^{-1}(2^2 + 2^2) = 4$
$E(N_B^2)$ is estimated by $\overline{N_B^2} = 2^{-1}(0^2 + 1^2) = 1/2$
which yields $R = 0$ in this example.

The estimator we recommend is

$$\tilde{R}_{\cup} = \frac{E(N_A N_B|\underline{a}, \underline{s}) - E(N_A|\underline{a}, \underline{s})E(N_B|\underline{a}, \underline{s})}{\sigma(N_A|\underline{a}, \underline{s})\sigma(N_B|\underline{a}, \underline{s})} \qquad \textbf{[S7]}$$

where the expectations are calculated with respect to the full joint probabilities of spike assignments. This estimator has the same form as the usual estimator R, but all spike counts are replaced by their expectation under the full joint posterior distribution of spike identities (Eq. **6** in the main text). For example, for $X = A$ or B,

$$\bar{N}_X = M^{-1}\sum_{m=1}^{M} N_{X,m} \quad \text{is replaced by}$$

$$E(N_X|\underline{a},\underline{s}) = M^{-1}\sum_{m=1}^{M} E(N_{X,m}|\underline{a},\underline{s}) \qquad \textbf{[S8]}$$

where

$$E(N_{X,m}|\underline{a},\underline{s}) = \sum_{k=0}^{n_m} k \times P(N_{X,m} = k|\underline{a},\underline{s}) \qquad \textbf{[S9]}$$

is the posterior expected number of spikes from cell X in bin m. The summation is from 0 to n_m, the total number of spikes in the bin, allowing for the possibility that cell X spiked $k = 0, 1, \ldots$, up to n_m spikes, and $E(N_{X,m}|\underline{a},\underline{s})$ is the sum of these k's, weighted by the posterior probabilities of observing k spikes from cell X; i.e., $P(N_{X,m} = k|\underline{a},\underline{s})$. Basic laws of probability obtain $P(N_{X,m} = k|\underline{a},\underline{s})$ by summing the full posterior probabilities in Eq. **6** (main text) over all configurations of joint spike identities that have a total of k spike identities belonging to cell X in bin m.

Consider the example in Fig. 1 (main text) again to illustrate the calculation of $\tilde{R}_{\cup}$. There are two bins; the first contains the first two spikes, so $n_m = 2$, and the possible number of spikes from cell A are $k = 0$, 1, and 2. According to the listed posterior probabilities, the chance of getting no spikes from cell A in bin 1 is

$$P(N_{A,1} = 0|\underline{a},\underline{s}) = 0,$$

the chance of observing one spike in bin 1 from cell A is

$$\begin{aligned} P(N_{A,1} = 1|\underline{a},\underline{s}) &= P((X_1, X_2) = (A, B) \quad \text{or } (A, C) \quad \text{or } (B, A) \\ &\quad \text{or } (C, A)|\underline{a},\underline{s}) \\ &= 0 + 0 + (0.13 + 0.1) + (0.07 + 0.06 + 0.04) \\ &= 0.4, \end{aligned}$$

and the chance of observing two spikes from cell A in bin 1 is

$$P(N_{A,1} = 2) = P((X_1, X_2) = (A, A)|\underline{a},\underline{s}) = (0.4 + 0.2) = 0.6.$$

Bin 2 has a total of three spikes, so the possible values of $N_{A,2}$, the spike count for neuron A in bin $b = 2$, are $k = 0$, 1, 2, and 3, with posterior probabilities calculated using the same process. We collected the posterior probabilities for $N_{A,1}$ and $N_{A,2}$ in the following tables:

$N_{A,1}$	0	1	2
probability	0	0.4	0.6

and

$N_{A,2}$	0	1	2	3
probability	0.3	0.1	0.6	0

This yields (Eq. **S9**)

$$E(N_{A,1}|\underline{a},\underline{s}) = (0 \times 0) + (1 \times 0.4) + (2 \times 0.6) = 1.6$$

$$E(N_{A,1}^2|\underline{a},\underline{s}) = (0^2 \times 0) + (1^2 \times 0.4) + (2^2 \times 0.6) = 2.8$$

$$E(N_{A,2}|\underline{a},\underline{s}) = (0 \times 0.3) + (1 \times 0.1) + (2 \times 0.6) = 1.3$$

$$E(N_{A,2}^2|\underline{a},\underline{s}) = (0^2 \times 0.3) + (1^2 \times 0.1) + (2^2 \times 0.6) = 2.5$$

so that Eq. **S8** evaluates to

$$\begin{aligned} E(N_A|\underline{a},\underline{s}) &= \frac{1.6 + 1.3}{2} = 1.45, \quad \text{and} \\ E(N_A^2|\underline{a},\underline{s}) &= \frac{2.8 + 2.5}{2} = 2.65. \end{aligned} \qquad \textbf{[S10]}$$

Using the same process, we calculate that the posterior probabilities for the spike count of cell B in bins 1 and 2 are

$N_{B,1}$	0	1	2
probability	0.77	0.23	0

and

$N_{B,2}$	0	1	2	3
probability	0	0.64	0.36	0

which yields

$$(N_{B,1}|\underline{a},\underline{s}) = 0 \times 0.77 + 1 \times 0.23 + 2 \times 0 = 0.23$$

$$E(N_{B,1}^2|\underline{a},\underline{s}) = 0 \times 0.77 + 1^2 \times 0.23 + 2^2 \times 0 = 0.23$$

$$E(N_{B,2}|\underline{a},\underline{s}) = 0 \times 0 + 1 \times 0.64 + 2 \times 0.36 + 3 \times 0 = 1.36$$

$$E(N_{B,2}^2|\underline{a},\underline{s}) = 0 \times 0 + 1^2 \times 0.64 + 2^2 \times 0.36 + 3^2 \times 0 = 2.08$$

so that

$$\begin{aligned} E(N_B|\underline{a},\underline{s}) &= \frac{0.23 + 1.36}{2} = 0.795 \quad \text{and} \\ E(N_B^2|\underline{a},\underline{s}) &= \frac{0.23 + 2.08}{2} = 1.155. \end{aligned} \qquad \textbf{[S11]}$$

It remains to calculate $E(N_A N_B|\underline{a},\underline{s}) = M^{-1}\sum_{m=1}^{M} E(N_{A,m}N_{B,m}|\underline{a},\underline{s})$. For the first bin, we have

$$\begin{aligned} P(N_{A,1}N_{B,1} = 0|\underline{a},\underline{s}) &= P(N_{A,1} = 0 \\ &\quad \text{or } N_{B,1} = 0|\underline{a},\underline{s}) = P(N_{A,1} = 0|\underline{a},\underline{s}) + P(N_{B,1} = 0|\underline{a},\underline{s}) \\ &= 0 + 0.77 = 0.77. \end{aligned}$$

$$P(N_{A,1}N_{B,1} = 1|\underline{a},\underline{s}) = P(N_{A,1} = 1 \quad \text{and} \quad N_{B,1} = 1|\underline{a},\underline{s})$$

cannot be reduced to the product of the marginal probabilities calculated earlier, because the spike counts may not be independent. Instead we must use the joint probabilities, which gives

$$\begin{aligned} P(N_{A,1}N_{B,1} = 1|\underline{a},\underline{s}) &= P((X_1, X_2) = (A, B) \\ &\quad \text{or } (B, A)|\underline{a},\underline{s}) = 0.13 + 0.10 = 0.23 \end{aligned}$$

$$P(N_{A,1}N_{B,1} \geq 2|\underline{a},\underline{s}) = 0$$

This gives $E(N_{A,1}N_{B,1}|\underline{a},\underline{s}) = 0.23$. Similarly, for the second bin, we have

$$P(N_{A,2}N_{B,2} = 0|\underline{a},\underline{s}) = 0.3 + 0 = 0.3$$

$$P(N_{A,1}N_{B,1}=1|\underline{a},\underline{s})=P(N_{A,1}=1 \quad \text{and} \quad N_{B,1}=0|\underline{a},\underline{s})=0.04$$

$$P(N_{A,1}N_{B,1}=2|\underline{a},\underline{s})=P(N_{A,1}=1 \quad \text{and} \quad N_{B,1}=2|\underline{a},\underline{s})+P(N_{A,1}=2 \quad \text{and} \quad N_{B,1}=1|\underline{a},\underline{s})=0.4+0.13+0.07+0.06=0.66$$

$$P(N_{A,1}N_{B,1}\geq 3|\underline{a},\underline{s})=0$$

Hence $E(N_{A,2}N_{B,2}|\underline{a},\underline{s})=1\times 0.04+2\times 0.66=1.36$, which gives

$$E(N_AN_B|\underline{a},\underline{s})=\frac{0.23+1.36}{2}=0.795. \qquad \textbf{[S12]}$$

Putting Eqs. **S10**, **S11**, and **S12** together, we have

$$\tilde{R}_{\cup}=\frac{0.795-1.45\times 0.795}{\sqrt{2.65-1.45^2}\sqrt{1.155-0.795^2}}\approx -0.67,$$

which is quite different from the usual estimate $R=0$. This illustrates why using the biased, traditional estimate R can have profoundly poor consequences for inferences about correlation, as we observed in Fig. 2 of the main text.

S4 Theorem Proofs. ***S4.1 Proof of Theorem 1.2.*** Before we prove Theorem 1.2, it is useful to point out that if the true identities of the spikes were known, we would estimate the coincidence rate with

$$T_{\text{true}}=M^{-1}\sum_{m=1}^{M}I\{\underline{x}_m^{\text{true}}\in\mathscr{C}_m\},$$

which has the same form as $T_{\perp}$ in Eq. **S3**, but involves the true spike identities $\underline{x}_m^{\text{true}}=(x_{m1},\ldots,x_{mn_m})^{\text{true}}$ rather than the identities determined by traditional spike sorting, $\underline{x}_m=(x_{m1},\ldots,x_{mn_m})$. The expectation of T_{true} is θ, since it uses the true spike identities. It is written as

$$\theta=E(T_{\text{true}})=M^{-1}\sum_{m=1}^{M}P(\underline{X}_m\in\mathscr{C}_m) \qquad \textbf{[S13]}$$

since the expectation of an indicator function is the probability of its argument. Note that this calculation depends neither on the waveforms $\underline{a}$, nor on the spike train history $\underline{s}$.

To prove Theorem 1.2, we will show that the expectations of T_{true} and $\tilde{T}_{\cup}$ are equal. The expectation of $\tilde{T}_{\cup}$ is

$$E(\tilde{T}_{\cup})=M^{-1}\sum_{m=1}^{M}\left(\sum_{\underline{i}_m\in\mathscr{C}m}E(P(\underline{X}_m=\underline{i}_m|\underline{a},\underline{s}))\right),$$

where $P(\underline{X}_m=\underline{i}_m|\underline{a},\underline{s})$ is the sum over all $\underline{i}$ except $\underline{i}_m$ of the joint posterior of $\underline{X}$ given $(\underline{a},\underline{s})$ in Eq. **6** in the main text, and its expectation is with respect to the joint distribution of the random variables inside the probability; i.e., $(\underline{a},\underline{s})$. Hence

$$\begin{aligned}
E=E(P(\underline{X}_m=\underline{i}_m|\underline{a},\underline{s}))&=\int_{\underline{s}}\int_{\underline{a}}\left(\sum_{\underline{i}\setminus\underline{i}_m}P(\underline{X}=\underline{i}|\underline{a},\underline{s})\right)f(\underline{a},\underline{s})d\underline{a}d\underline{s}\\
&=\int_{\underline{s}}\left[\int_{\underline{a}}\left(\sum_{\underline{i}\setminus\underline{i}_m}P(\underline{X}=\underline{i}|\underline{a},\underline{s})\right)f(\underline{a}|\underline{s})d\underline{a}\right]f(\underline{s})d\underline{s}\\
&=\int_{\underline{s}}\left[\int_{\underline{a}}\left(\sum_{\underline{i}\setminus\underline{i}_m}\frac{\pi_{\underline{i}}(\underline{s})f_{\underline{i}}(\underline{a}|\underline{s})}{f(\underline{a}|\underline{s})}\right)f(\underline{a}|\underline{s})d\underline{a}\right]f(\underline{s})d\underline{s}\ \text{by Eq.6 in the main text}\\
&=\int_{\underline{s}}\left[\int_{\underline{a}}\left(\sum_{\underline{i}\setminus\underline{i}_m}\pi_{\underline{i}}(\underline{s})f_{\underline{i}}(\underline{a}|\underline{s})\right)d\underline{a}\right]f(\underline{s})d\underline{s}\\
&=\int_{\underline{s}}\left[\sum_{\underline{i}\setminus\underline{i}_m}\pi_{\underline{i}}(\underline{s})\int_{\underline{a}}f_{\underline{i}}(\underline{a}|\underline{s})d\underline{a}\right]f(\underline{s})d\underline{s}\\
&=\int_{\underline{s}}\left[\sum_{\underline{i}\setminus\underline{i}_m}\pi_{\underline{i}}(\underline{s})\right]f(\underline{s})d\underline{s}\ \text{since a density function integrates to 1}\\
&=\int_{\underline{s}}\left[\sum_{\underline{i}\setminus\underline{i}_m}P(\underline{X}=\underline{i}|\underline{s})\right]f(\underline{s})d\underline{s}\ \text{by definition of }\pi_{\underline{i}}(\underline{s})\\
&=\int_{\underline{s}}[P(\underline{X}_m=\underline{i}_m|\underline{s})]f(\underline{s})d\underline{s}=P(\underline{X}_m=\underline{i}_m).
\end{aligned}$$

Putting the result together, we have

$$\begin{aligned}
E(\tilde{T}_{\cup})&=M^{-1}\sum_{m=1}^{M}\left(\sum_{\underline{i}_m\in\mathscr{C}m}P(\underline{X}_m=\underline{i}_m)\right)\\
&=M^{-1}\sum_{m=1}^{M}P(\underline{X}_m\in\mathscr{C}_m),
\end{aligned}$$

which is Eq. **S13**. Hence $E(\tilde{T}_{\cup})=\theta:\ \tilde{T}_{\cup}$ is unbiased for the true coincidence rate. QED.

It is not possible to prove in general that the variance of $\tilde{T}_{\cup}$ tends to zero in large samples. This is because the summands of $\tilde{T}_{\cup}$ are not independent, since they all depend on the same vector of random variables $(\underline{a},\underline{s})$. However, if the firing rates of the cells depend on a finite rather than on the infinite past, then $\tilde{T}_{\cup}$ is the average of terms whose correlation decreases with time. Provided the correlation wanes off sufficiently fast, a central limit theorem applies to $\tilde{T}_{\cup}$ so that $\text{Var}(\tilde{T}_{\cup})$ tends to zero when the sample size increases, which guarantees consistency. This is the case in the illustration in the main text (Fig. 3), where groups of spikes separated by more than the coupling length of 10 ms are independent (See section S5.3).

S4.2 Proof of Theorem 1.1. It is useful to introduce another estimator for θ, $\tilde{T}_{\perp}$, which is obtained from spikes sorted one at a time (traditional sorting), but which uses the posterior of spikes identities in Eq. **2** rather than the maximum posterior probability in Eq. **3**. The usual estimator for θ, $T_{\perp}$, suffers from two sources of bias: the bias that originates from sorting spikes independently, and the bias that originates from using full spike assignment; $\tilde{T}_{\perp}$ only suffers from the former source, which we now prove.

If spikes are sorted one at a time, then

$$\tilde{T}_{\perp}=M^{-1}\sum_{m=1}^{M}\left(\sum_{\underline{i}_m\in\mathscr{C}m}\left(\prod_{k=1}^{n_m}P(X_{mk}=i_{mk}|a_{mk})\right)\right),$$

which is obtained by replacing the joint probability in Eq. **S4** by the product of the marginals from traditional sorting (Eq. **2** in the main text). Following basic rules of probability, the expectation of $\tilde{T}_{\perp}$ is calculated by integrating over the random variables, the $\underline{a}_m$,

with respect to their true distribution, $f(\underline{a}_m)$. This gives

$$E(\tilde{T}_\perp) = M^{-1} \sum_{m=1}^{M} \left(\sum_{\underline{i}_m \in \mathscr{C}_m} \int_{\underline{a}_m} \prod_{k=1}^{n_m} \left(\frac{\pi_{i_{mk}} f_{i_{mk}}(a_{mk})}{f(a_{mk})} \right) f(\underline{a}_m) d\underline{a}_m \right),$$

which does not reduce to θ in Eq. **S13**, hence $\tilde{T}_\perp$ is biased. But if the cells and their spikes are independent, the joint distribution reduces to the product of its marginals; i.e., $f(\underline{a}_m) = \prod_{k=1}^{n_m} f(a_{mk})$. In that case,

$$\begin{aligned} E(\tilde{T}_\perp) &= M^{-1} \sum_{m=1}^{M} \left(\sum_{\underline{i}_m \in \mathscr{C}_m} \prod_{k=1}^{n_m} \int_{a_{mk}} \pi_{i_{mk}} f_{i_{mk}}(a_{mk}) da_{mk} \right) \\ &= M^{-1} \sum_{m=1}^{M} \left(\sum_{\underline{i}_m \in \mathscr{C}_m} \prod_{k=1}^{n_m} \pi_{i_{mk}} \right), \end{aligned}$$

and $\prod_{k=1}^{n_m} \pi_{i_{mk}} = \prod_{k=1}^{n_m} P(X_{mk} = i_{mk}) = P(\underline{X}_m = \underline{i}_m)$ iff the spikes are independent, so that $E(T_\perp)$ reduces to Eq. **S13**. QED.

The usual estimate of θ, $T_\perp$, suffers from the additional source of bias that originates from using full spike assignments, so it will be biased even when spikes and their waveforms are independent. To prove this formally, one would need to calculate its expectation

$$E(T_\perp) = M^{-1} \sum_{m=1}^{M} P(\underline{X}_m^{\text{full}} \in \mathscr{C}_m | \underline{a})$$

with respect to the joint distribution of the waveforms $\underline{a}$, where $\underline{X}_m^{\text{full}}$ is the random variable of full spike assignments, which is different from the true spike identities $\underline{X}_m$. This calculation is very difficult because it involves evaluating the integral of $f(\underline{a})$ over the values of $\underline{a}$ that are such that $\{(\arg\max_{\underline{i}_m} P(\underline{X}_m = \underline{i}_m | \underline{a})) \in C_m\}$. We avoid the general proof and instead refer to the illustration in the main text (Fig. 3) for counterexamples to the claim that $T_\perp$ is unbiased.

There are exceptions. First, if spikes can be classified perfectly, then $T_\perp = T_{\text{true}}$ since the true spike identities are known. Second, if spikes are independent, then sorting spikes one at a time is equivalent to ensemble sorting, the "correct," unbiased sorting procedure. But unlike $\tilde{T}_\perp$, $T_\perp$ is not unbiased despite cells being independent, unless neurons have equal firing rates, because it uses full spike assignments.

S5 Supplement for the Illustrations.

S5.1 Additional comments for Fig. 1—Tolerated percentage of misclassified spikes. Figs. 2 and 3 in the main text show that the spike count correlation and coincidence rate estimates can be substantially biased for physiologically plausible true values of neuronal correlations, even for relatively low rates of misclassified spikes. Here we discuss the types of misclassification rates one can expect from sorting experimental data.

To know for certain the typical spike misclassification error rate in a neurophysiology experiment would require knowing the ground truth (which cells actually emitted which spikes). There is at least one dataset which contains this information (5), and the rate of misclassification has been compared against two easily computed metrics, L-ratio and Isolation Distance, which are available to any analyst for use with their own, uncertain data. The values of these metrics are sometimes reported in papers which make use of ensemble recordings (for a recent example, Allen et al. ref. 6, where the Isolation Distance ≈ 18).

Consider a large number of cells of equal firing rate recorded on a tetrode. A common approach is to spike sort using the first three principal components from each of the four wires on the tetrode, yielding 12 spike features. Suppose that each cluster is Gaussian, and evenly spaced in this multidimensional feature space. The isolation distance of a cluster is defined (5) as the square of the radius from the cluster's center for which exactly half of the spikes contained within that radius belong to that cluster. In this configuration, if each cluster is separated from its nearest neighbors by $\mu = 3$, with μ defined as in the main text, then the isolation distance is $D^2 \approx 25$. In other words, if each cell's waveform distribution is about three standard deviations away from its nearest neighbors in each feature dimension, then the isolation distance is approximately 25. This isolation distance is typically considered to be adequate by experimentalists, and yet it implies a pairwise misclassification rate (considering only nearest neighbor clusters in a single feature dimension) of approximately 6.7%. While this may not seem large, it is more than enough to lead to mistaken inferences about correlation, according to the results in Fig. 2. Furthermore, the total misclassification rate due to "pollution" from all neighboring clusters in all dimensions will be much higher.

Alternatively, we can use the experimental data from Schmitzer-Torbert et al. (5) to directly estimate the impact of a given isolation distance on misclassification rate, and indirectly on bias of correlation estimates. In the Schmitzer-Torbert et al. (5) data, it is reported that an Isolation Distance of 25 corresponds to a Type 1 error rate of 1% and a Type 2 error rate of 10%, for a mean misclassification rate of 5.5%. There is great variability around this mean, and the misclassification rate even for cells with more impressive Isolation Distances may still be quite large. As can be seen in Fig. 2 of the main text, this rate of misclassification results in a nontrivial amount of bias in correlation estimates, which will significantly impact inference and interpretation of scientific questions.

S5.2 Traditional spike sorting. The application in the main text compares ensemble sorting to traditional sorting. To apply the latter, we follow the implications of Eq. **3**. All we need is to specify Eq. **1**, the distribution of a spike waveform (WF), since spike assignments (Eq. **3**) depend on it.

A single spike with WF feature a has distribution $f(a) = \pi_1 f_1(a) + \pi_2 f_2(a)$, where π_i is the probability that cell i spiked, and f_i is the distribution of WF features for cell i. By assumption, f_1 and f_2 are univariate Gaussian distributions with means 0 and μ, and unit variances (with no loss of generality); also $\pi_1 = \lambda_1/(\lambda_1 + \Lambda_2)$ and $\pi_2 = \Lambda_2/(\lambda_1 + \Lambda_2)$, since a single spike was emitted by one of the two cells, with probabilities proportional to their overall firing rates, λ_1 and Λ_2.

We explain Λ_2. Recall that cell 2 has the stochastic rate $\lambda_2(t) = \lambda_2 \times g(C_t)$, where C_t is the time elapsed since the last spike from cell 1; cell 2 has an altered firing rate $\beta\lambda_2$ for a duration $\nu = 10$ ms after cell 1 spikes, and rate λ_2 otherwise. Hence its overall rate is the expectation of $\lambda_2 \times g(C_t)$ with respect to C_t. Since cell 1 spikes according to a Poisson process with rate λ_1, C_t is distributed as an exponential random variable with mean λ_1^{-1}. A trivial calculation yields $\Lambda_2 = E(\lambda_2 \times g(C_t)) = \lambda_2 \times (\beta - \beta e^{-\lambda_1 \nu} + e^{-\lambda_1 \nu})$. We can verify that $\Lambda_2 \geq \lambda_2$ if $\beta \geq 1$, and $\Lambda_2 < \lambda_2$ if $\beta < 1$, since $\Lambda_2 = \lambda_2 \times (1 + (\beta - 1)(1 - e^{-\lambda_1 \nu}))$ and $(1 - e^{-\lambda_1 \nu}) \geq 0$. If $\nu = 0$ or $\beta = 1$, in which cases cell 1 has no effect on cell 2 spiking, then $\Lambda_2 = \lambda_2 = \lambda_2(t)$.

S5.3 Joint sorting for the two correlated Poisson cell simulation. To specify the spike sorting scheme for a group of $n \geq 1$ spikes, we need to specify Eq. **6** in the main text:

$$P(\underline{X} = \underline{i} | \underline{a}, \underline{s}) = \frac{f(\underline{a} | \underline{i}, \underline{s}) P(\underline{X} = \underline{i} | \underline{s})}{f(\underline{a} | \underline{s})},$$

which are the joint probabilities that the spike IDs have value $\underline{i} = (i_1, \ldots, i_n)$, given all waveforms $\underline{a} = (a_1, \ldots, a_n)$ and ISIs $\underline{s} = (s_1, \ldots, s_n, s_{n+1})$.

We assume in this example that WFs are independent and stationary, so $f(\underline{a}|\underline{i}, \underline{s})$ simplifies to $\prod_{j=1}^{n} f_{i_j}(a_j)$, as argued in sec. S2.2, where, in our illustration, f_1 and f_2 are univariate Gaussian distributions with means 0 and μ, and unit variances.

Next, we need $P(\underline{X} = \underline{i}|\underline{s})$. In the main text, we showed that this amounts to calculating

$$P(X_j = i_j | s_j, X_{j-1} = i_{j-1}, s_{j-1}, \ldots, X_1 = i_1, s_1), \qquad j = 1, \ldots, n$$

the probability that spike j has identity $i_j = 1$ or 2, given the IDs of all previous spikes and the ISIs up to spike j, and

$$f_{S_j|X_{j-1}, S_{j-1}, X_{j-2}, \ldots, X_1, S_1}(s), \qquad j = 1, \ldots, m+1$$

the density of ISI S_j, given the past up to spike $j-1$.

The past that is relevant to the probability of X_j is fully contained in the random variable $C_j = C(s_1, X_1, \ldots, s_j, X_j)$, the time elapsed between spike j and the last spike from cell 1; for example, $C(X_1 = 2, X_2 = 1, X_3 = 1, X_4 = 1) = 0$ and $C(X_1 = 2, X_2 = 1, X_3 = 2, X_4 = 2) = s_3 + s_4$. If $C_{j-1} + s_j > \nu$ ($\nu = 10$ ms is the coupling duration), then the last spike from cell 1 happened more than ν ago, so that the firing rate of cell 2 is λ_2. This implies that the spike with ID X_j was fired by cell 1 with probability $\lambda_1/(\lambda_1 + \lambda_2)$, and by cell 2 with probability $\lambda_2/(\lambda_1 + \lambda_2)$. Conversely, if $C_{j-1} + s_j < \nu$, $X_j = 1$ and 2 with probabilities $\lambda_1/(\lambda_1 + \beta\lambda_2)$ and $\beta\lambda_2/(\lambda_1 + \beta\lambda_2)$, respectively. To summarize, we have

$$\begin{aligned}
&P(X_j = i | s_j, X_{j-1} = i_{j-1}, s_{j-1}, \ldots, X_1 = i_1, s_1) = \lambda_i/(\lambda_1 + \lambda_2), i = 1, 2,\\
&\text{if } C_{j-1} > \nu,\\
&P(X_j = 1 | s_j, X_{j-1} = i_{j-1}, s_{j-1}, \ldots, X_1 = i_1, s_1) = \lambda_1/(\lambda_1 + \beta\lambda_2)\\
&\text{if } C_{j-1} \le \nu,\\
&P(X_j = 2 | s_j, X_{j-1} = i_{j-1}, s_{j-1}, \ldots, X_1 = i_1, s_1) = \beta\lambda_2/(\lambda_1 + \beta\lambda_2)\\
&\text{if } C_{j-1} \le \nu. \qquad \textbf{[S14]}
\end{aligned}$$

We also split the calculation in two cases to calculate

$$f_{S_j|X_{j-1}, S_{j-1}, X_{j-2}, \ldots, X_1, S_1}(s), \qquad j = 1, \ldots, n+1.$$

If $C_{j-1} > \nu$, the effect of the last spike from cell 1 on the firing rate of cell 2 has ceased, so the waiting time S_j to the next spike is exponentially distributed with rate $(\lambda_1 + \lambda_2)$. If $C_{j-1} \le \nu$, cell 2 has rate $\beta\lambda_2$ for a duration $\nu - C_{j-1}$, and rate λ_2 after that, so that S_j has the mixture distribution

$$\begin{aligned}
f_{S_j|X_{j-1}, S_{j-1}, X_{j-2}, \ldots, X_1, S_1}(s) = p \cdot f_{S_j|S_j < \nu - C_{j-1}}(s) + (1-p)\\
\cdot f_{S_j|S_j \ge \nu - C_{j-1}}(s),
\end{aligned}$$

where $p = P(S_j < \nu - C_{j-1}) = 1 - e^{-(\lambda_1 + \beta\lambda_2)(\nu - C_{j-1})}$, the density $f_{S_j|S_j < \nu - C_{j-1}}(s)$ is the truncated exponential density $(\lambda_1 + \beta\lambda_2)e^{-(\lambda_1 + \beta\lambda_2)s}/(1 - e^{-(\lambda_1 + \beta\lambda_2)(\nu - C_{j-1})})$ when $s < \nu - C_{j-1}$, and 0 otherwise, and the density $f_{S_j|S_j \ge \nu - C_{j-1}}(s)$ is the shifted exponential density $(\lambda_1 + \lambda_2)e^{-(\lambda_1 + \lambda_2)(s - (\nu - C_{j-1}))}$ if $s > \nu - C_{j-1}$, and 0 otherwise. Putting everything together, we have

$$\begin{aligned}
f_{S_j|X_{j-1}, S_{j-1}, X_{j-2}, \ldots, X_1, S_1}(s) &= (\lambda_1 + \lambda_2)e^{-(\lambda_1 + \lambda_2)s} \quad \text{if } C_{j-1} > \nu,\\
&= (\lambda_1 + \beta\lambda_2)e^{-(\lambda_1 + \beta\lambda_2)s} I_{\{s < \nu - C_{j-1}\}}\\
&+ (\lambda_1 + \lambda_2)e^{-(\lambda_1 + \lambda_2)s} e^{-\lambda_2(\beta - 1)(\nu - C_{j-1})} I_{\{s > \nu - C_{j-1}\}}, \quad \text{if } C_{j-1} \le \nu.
\end{aligned}$$

[S15]

These densities are then evaluated at observed ISI values s_j to obtain $P(\underline{X} = \underline{i}|\underline{s})$.

The derivation of Eq. **S15** shows that the ISIs do contain information about the spike IDs $\underline{X}$, since their distributions depend on $\underline{X}$ through the C_j's. To understand this more intuitively, imagine that λ_1 is relatively low, that $\beta \gg 1$, and that ν is fairly long, so that cell 2 is very likely to spike shortly after cell 1 spikes. Imagine that we observe a spike, whose ID X_j we must determine. Then the odds of cell 1 having fired that spike decreases as the next ISI S_{j+1} increases, since a spike from cell 1 almost guarantees a spike from cell 2 shortly afterwards. When sorting a group of n spikes with IDs $X_1, \ldots, X_n$, it is therefore important to include the ISIs $S_1, \ldots, S_n$ and the extra ISI S_{n+1} in the calculation, as we have done, since S_j contains information about the IDs of past spikes. Failing to do so will bias spike sorting and subsequent correlation estimates.

There are two outstanding issues with the calculations of Eqs. **S14** and **S15**, which arise because the electrode recording is finite. First, consider a spike j that occurs within ν of the beginning of the recording. Then C_j, the time elapsed since the last spike of cell 1, is unknown when spike j and its predecessors have IDs $X_1 = X_2 = \ldots = X_j = 2$. Although it is relatively easy to derive the proper correction for this left truncation problem, we will ignore it here for simplicity, and instead discard the beginning of the spike train until an ISI larger than ν is observed. Second, we need a right truncation correction for Eq. **S15**. Indeed, given the past up to spike $j-1$, the duration S_j until we observe the next spike is not an ISI, but the minimum between and ISI and the remaining duration T_j of the experiment. Although it is true that only the last observed ISI s_{n+1} is truncated, we cannot assume that we know the future when we derive the density of the random variable S_j in Eq. **S15**, since the calculation is not conditioned on the future. Hence Eq. **S15** must be rescaled by the integral from 0 to T_j of Eq. **S15**. For spikes that are far from the end of the experiment, this correction will be negligible, but it may matter for the last few ISIs.

Finally, we can verify that the calculations of Eqs. **S14** and **S15** do not depend on the IDs of the spikes that are not included the group of n spikes we are sorting together, provided that s_1 and s_{n+1} are larger than $\nu = 10$ ms. This means that, in our particular application, groups of spikes that are separated by more than ν are necessarily independent, so that we can sort such groups independently, rather than ensemble sort the whole spike train at once. To proceed, we thus consider each new spike and check if adjacent spikes are within ν. If not, we sort the spike alone and move onto the next spike. Otherwise we collect all spikes that are less than ν away from their immediate neighbors, and we sort them jointly, as described above. Then we move onto the next spike.

1. Ventura V (2009) Traditional waveform based spike sorting yields biased rate code estimates. *Proc Natl Acad Sci USA* 106:6921–6.
2. Ventura V (2009) Automatic spike sorting using tuning information. *Neural Comput* 21:2466–2501.
3. Pouzat C, Delescluse M, Voit P, Diebolt J (2004) Improved spike-sorting by modeling firing statistics and burst-dependent spike amplitude attenuation: A Markov chain Monte Carlo approach. *J Neurophysiol* 91:2910–2928.
4. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *J Neurophysiol* 93:1074–1089.
5. Schmitzer-Torbert N, Jackson J, Henze D, Harris K, Redish AD (2005) Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* 131:1–11.
6. Allen K, Fuchs EC, Jaschonek H, Bannerman DM, Monyer H (2011) Gap junctions between interneurons are required for normal spatial coding in the hippocampus and short-term spatial memory. *J Neurosci* 31:6542–6552.

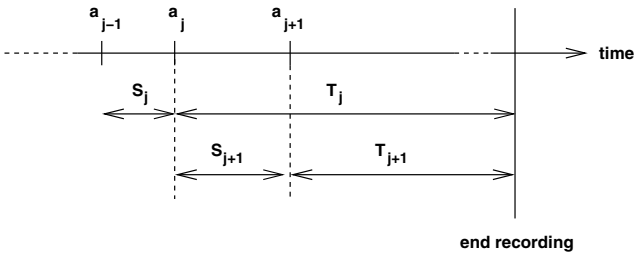

Fig. S1. Partial spike train (*Top*) with vertical ticks showing 3 consecutive spikes. These spikes have waveform measurements a_{j-1}, a_j, a_{j+1}. The interspike intervals (ISIs), S_j and S_{j+1}, and the length of time between each spike and the end of recording, T_j and T_{j+1}, are also indicated. All of these quantities are used in ensemble sorting, as each contributes information about the provenance of each spike.