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 $T_\cup$ 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 $\tilde{R}_U$. 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

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

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

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

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

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

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

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

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

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

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

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

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

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

Putting the result together, we have

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

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

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

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

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

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

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

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

$$f_{S_j|X_{j-1}, S_{j-1}, X_{j-2}, \ldots, X_1, S_1}(s) = (\lambda_1 + \lambda_2)e^{-(\lambda_1 + \lambda_2)s} \quad \text{if } C_{j-1} > \nu,$$
$$= (\lambda_1 + \beta\lambda_2)e^{-(\lambda_1 + \beta\lambda_2)s} I_{\{s < \nu - C_{j-1}\}}$$
$$+ (\lambda_1 + \lambda_2)e^{-(\lambda_1 + \lambda_2)s} e^{-\lambda_2(\beta - 1)(\nu - C_{j-1})} I_{\{s > \nu - C_{j-1}\}}, \quad \text{if } C_{j-1} \le \nu. \qquad \textbf{[S15]}$$

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.

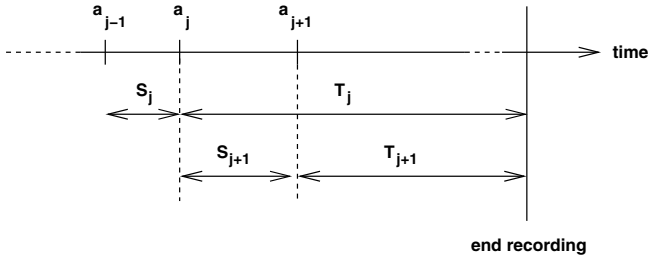

Fig. S1. Partial spike train (*Top*) with vertical ticks showing 3 consecutive spikes. These spikes have waveform measurements a_{j-1}, a_j, a_{j+1}. The interspike intervals (ISIs), S_j and S_{j+1}, and the length of time between each spike and the end of recording, T_j and T_{j+1}, are also indicated.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