Supplementary material for “Measuring and interpreting neuronal correlations” by Cohen and Kohn

**Spike count correlation depends on measurement window in a correlated rate model**

We showed in the text that correlations arising from nearly synchronous (jittered) spikes depend on the length of the measurement window (Figure 3). Measurements of $r_{SC}$ depend on the measurement window even for very different correlation structures. For example, consider two neurons whose spiking responses are determined by independent Poisson processes conditioned on correlated underlying firing rates. The correlation between their spike count responses $n_1$ and $n_2$ computed over time $T$ is given by

$$r_{SC}(T) = \frac{\text{cov}(n_1, n_2)}{\sqrt{\text{var}(n_1) \text{var}(n_2)}} = \frac{\langle n_1 n_2 \rangle - \langle n_1 \rangle \langle n_2 \rangle}{\sqrt{\langle n_1^2 \rangle - \langle n_1 \rangle^2 \langle n_2^2 \rangle - \langle n_2 \rangle^2}} \quad (1)$$

The covariance and variances depend on three quantities: the joint distribution over spikes $P(n_1, n_2)$, the marginal distributions $P(n_1)$ and $P(n_2)$, and the conditional spike count distributions $P(n_1 | v_1)$ and $P(n_2 | v_2)$ where $v_1$ and $v_2$ are the correlated underlying firing rates. The joint and marginal distributions are given by

$$P(n_1, n_2) = \int dv_1 dv_2 P(n_1 | v_1)P(n_2 | v_2)P(v_1, v_2) \quad (2)$$

$$P(n_i) = \int dv_i P(n_i | v_i)P(v_i) \quad (3)$$

Because the conditional spike count distributions are Poisson, the conditional spike count distributions are given by

$$P(n_i | v_i) = \frac{(v_i T)^{n_i} e^{-v_i T}}{n_i!} \quad (4)$$

Equation (3) tells us that the mean spike counts are, as expected, given by

$$\langle n_i \rangle = \sum_{n_i} n_i \int dv_i P(n_i | v_i)P(v_i) = \int dv_i P(v_i) \sum_{n_i} n_i P(n_i | v_i) = \int dv_i P(v_i) v_i T = \langle v_i \rangle T \quad (5)$$

Given the joint distribution in Equation (2), we can compute

$$\langle n_1, n_2 \rangle = \sum_{n_1, n_2} n_1 n_2 \int dv_1 dv_2 P(n_1 | v_1)P(n_2 | v_2)P(v_1, v_2) \quad (6)$$

$$= \int dv_1 dv_2 P(v_1, v_2) \sum_{n_1, n_2} n_1 n_2 P(n_1 | v_1)P(n_2 | v_2)$$

$$= \int dv_1 dv_2 P(v_1, v_2) \sum_{n_1} n_1 P(n_1 | v_1) \sum_{n_2} n_2 P(n_2 | v_2)$$

$$= \int dv_1 dv_2 P(v_1, v_2) v_1 v_2 T^2 = \langle v_1 v_2 \rangle T^2.$$

Combining Equations (5) and (6) lets us compute the numerator of $r_{SC}$:

$$\text{cov}(n_1, n_2) = \langle n_1 n_2 \rangle - \langle n_1 \rangle \langle n_2 \rangle = \langle v_1 v_2 \rangle T^2. \quad (7)$$

To compute the variances of $n_1$ and $n_2$, we need to know $\langle n_i^2 \rangle$, which is given by
\[ \langle n_i^2 \rangle = \sum n_i^2 \int dv_i P(n_i \mid v_i)P(v_i) = \int dv_i P(v_i) \sum n_i^2 P(n_i \mid v_i). \quad (8) \]

For a Poisson distribution, the average of the square of the spike count is \( v_i^2 T^2 + v_i T \), so
\[ \langle n_i^2 \rangle = \int dv_i P(v_i)(v_i^2 T^2 + v_i T) = \langle v_i^2 \rangle T^2 + \langle v_i \rangle T. \quad (9) \]

Then
\[ \text{var}(n_i) = \langle n_i^2 \rangle - \langle n_i \rangle^2 = \text{var}(v_i) T^2 + \langle v_i \rangle T. \quad (10) \]

Combining Equations (7) and (10) gives
\[ r_{SC}(T) = \frac{\text{cov}(n_1, n_2)}{\sqrt{\text{var}(n_1) \text{var}(n_2)}} = \frac{\text{cov}(v_1, v_2) T^2}{\sqrt{(\text{var}(v_1) T^2 + \langle v_1 \rangle T) (\text{var}(v_2) T^2 + \langle v_2 \rangle T)}} \]
\[ = r_{SC}(\infty) \frac{T}{\sqrt{T + \frac{\langle v_1 \rangle}{\text{var}(v_1)} \left(T + \frac{\langle v_2 \rangle}{\text{var}(v_2)}\right)}}, \quad (11) \]

where \( r_{SC}(\infty) \) is the correlation between the underlying firing rates \( v_1 \) and \( v_2 \).

Because \( \frac{\langle v_i \rangle}{\text{var}(v_i)} > 0 \),
\[ \frac{T}{\sqrt{T + \frac{\langle v_1 \rangle}{\text{var}(v_1)} \left(T + \frac{\langle v_2 \rangle}{\text{var}(v_2)}\right)}} < 1. \]

Therefore, \( r_{SC}(T) < r_{SC}(\infty) \). As \( T \) becomes large compared to \( \frac{\langle v_i \rangle}{\text{var}(v_i)} \),
\[ \frac{T}{\sqrt{T + \frac{\langle v_1 \rangle}{\text{var}(v_1)} \left(T + \frac{\langle v_2 \rangle}{\text{var}(v_2)}\right)}} \to 1, \text{ so } r_{SC}(T) \to r_{SC}(\infty). \]

See Appendix 2 of Bair et al. (2001) for a similar derivation.

**Overly stringent spike sorting criteria reduces measured \( r_{SC} \)**

We showed through simulations that overly stringent spike sorting criteria reduces measurements of correlations (main text, Figure 5) and provided an analytical description of this phenomenon. Here we derive this relationship.

Consider two neurons with spike count responses \( n_1 \) and \( n_2 \), whose spikes are discarded with probability \( p_1 \) and \( p_2 \). The spikes kept from each neuron during spike sorting are given by \( m_1 \) and \( m_2 \), where \( m_i = n_i (1 - p_i) \).

The random discarding of spikes is independent for the two neurons, so \( \text{cov}(m_1, m_2) \) is simply a function of \( \text{cov}(n_1, n_2) \). Specifically,
\[
\text{cov}(m_1, m_2) = (1 - p_1)(1 - p_2)\text{cov}(n_1, n_2). 
\] (12)

As above, \(<m_i>\) depend on the conditional probabilities \(P(m_i|n_i)\). Because \(m_i\) has a binominal distribution over \(n_i\) spike counts with probability of success \((1-p_i)\),

\[
P(m_i | p_i) = \sum_{n_i} P(m_i | n_i, p_i)P(n_i | p_i). 
\] (13)

Then

\[
\langle m_i \rangle = \sum_{m_i} m_i P(m_i | p_i) 
= \sum_{n_i} P(n_i | p_i) \sum_{m_i} m_i P(m_i | n_i, p_i).
\] (14)

The inner sum is equal to the average value of \(m_i\) for a particular value of \(n_i\) and \(p_i\), which is equal to \((1 - p_i)n_i\), so

\[
\langle m_i \rangle = \sum_{n_i} P(n_i | p_i)n_i(1 - p_i) = (1 - p_i)\langle n_i \rangle. 
\] (15)

We can also use Equation (13) to compute \(<m_i^2>\):

\[
\langle m_i^2 \rangle = \sum_{m_i} m_i^2 P(m_i | p_i) 
= \sum_{n_i} P(n_i | p_i) \sum_{m_i} m_i^2 P(m_i | n_i, p_i).
\] (16)

As above, the inner sum equals the average value of \(m_i^2\) for a particular value of \(n_i\) and \(p_i\), so we can use the definition of variance to obtain

\[
\langle m_i^2 \rangle = \sum_{n_i} P(n_i | p_i) \left[ \text{var}(m_i) + \langle m_i \rangle^2 \right] 
\] (17)

Using the mean of the binominal distribution over \(n_i\) spike counts with probability of success \((1-p_i)\) and the formula for \(<m_i>\) from Equation (15), we can calculate

\[
\langle m_i^2 \rangle = \sum_{n_i} P(n_i | p_i) \left[ n_i p_i (1 - p_i) + n_i^2 (1 - p_i)^2 \right] 
\] (18)

\[
= \langle n_i \rangle p_i (1 - p_i) + \langle n_i^2 \rangle (1 - p_i)^2 
= \langle n_i \rangle p_i (1 - p_i) + \text{var}(n_i)(1 - p_i)^2 + \langle n_i \rangle^2 (1 - p_i)^2 
\]

Equations (15) and (18) let us calculate \(\text{var}(m_i)\):

\[
\text{var}(m_i) = \langle m_i^2 \rangle - \langle m_i \rangle^2 
= \langle n_i \rangle p_i (1 - p_i) + \text{var}(n_i)(1 - p_i)^2 + \langle n_i \rangle^2 (1 - p_i)^2 - \langle n_i \rangle^2 (1 - p_i)^2 
= \langle n_i \rangle p_i (1 - p_i) + \text{var}(n_i)(1 - p_i)^2 
\]

Nature Neuroscience: doi:10.1038/nn.2842
We can then use (12) and (19) to compute $r_{SC}$:

$$r_{SC-oversort} = \frac{\text{cov}(m_1, m_2)}{\sqrt{\text{var}(m_1) \cdot \text{var}(m_2)}}$$

$$= \frac{\text{cov}(n_1, n_2)}{\sqrt{\text{var}(n_1) \cdot \text{var}(n_2)}} \left[ \left(1 + \frac{p_1 \langle n_1 \rangle}{(1 - p_1) \cdot \text{var}(n_1)}\right) \left(1 + \frac{p_2 \langle n_1 \rangle}{(1 - p_2) \cdot \text{var}(n_2)}\right) \right]$$

$$= \frac{r_{SC-original}}{\sqrt{\left[ \left(1 + \frac{p_1 \langle n_1 \rangle}{(1 - p_1) \cdot \text{var}(n_1)}\right) \left(1 + \frac{p_2 \langle n_1 \rangle}{(1 - p_2) \cdot \text{var}(n_2)}\right) \right]}}$$

By definition, $0 \leq p_i \leq 1$, $\langle n_i \rangle \geq 0$, and $\text{var}(n_i) \geq 0$. Therefore, the denominator in Equation (20) is always greater than 1, so $r_{SC-oversort}$ will always underestimate $r_{SC-original}$. 