

**Supplemental Data for:**

Goldberg et al., Neuron 42, pp. 489-500

Bias in the Distribution of the Spike-Triggered SI

To demonstrate the relationship between the bias in the distribution of the spike-triggered SI ( $\rho_\psi^{spike}$ , defined in the manuscript) and the correlation between the SI ( $\rho_\psi$ ) and the spiking of a neuron, we assume that this spiking is described by a time-varying Poisson process with a time-varying rate,  $m(t)$ . We define an indicator variable  $S(t)$  that is equal to unity if the neuron emits a spike in a short interval  $\Delta t$  around time  $t$ . Formally,  $\Pr[S(t) = 1 | m(t)] = m(t)\Delta t$ . We denote the conditional distribution of  $\rho_\psi$  given  $S$  as  $P(\rho | S)$ , where for brevity we have dropped the index  $\psi$ . Thus, the distribution of  $\rho^{spike}$  over time is  $P(\rho | S = 1)$ . Calculating the mean of this distribution, denoted  $B$ , yields

$$B \equiv \int \rho P(\rho | S = 1) d\rho = \frac{1}{\Pr[S = 1]} \sum_{S=0,1} \int S \rho P(\rho | S) \Pr[S] d\rho = \tag{S1}$$

$$\frac{1}{\langle m \rangle_t \Delta t} \sum_{S=0,1} \int S \rho P(\rho | m) \Pr[S | m] P(m) dm d\rho = \frac{\langle m \rho \rangle_t}{\langle m \rangle_t}$$

where  $\langle \dots \rangle_t$  denotes a temporal average. If we assume that the distribution of  $\rho$  has zero mean, and that the standard deviation of the underlying rate  $m(t)$  is on the order of its mean, then by replacing the latter with the former in equation (S1) and dividing this equation by the standard deviation of the SI, denoted  $\sigma$ , we arrive at an approximate expression for the bias in the distribution of  $\rho^{spike}$

$$B/\sigma \sim c.c.(m, \rho) \tag{S2}$$

where *c.c.* stands for correlation coefficient. The results of the network simulations indicate that this approximation is reasonable.

In the single state scenario the firing rate of one neuron is related to the value of one degree of freedom (DOF) in the map, whereas the SI is a spatial filtering of  $N$  DOF. Therefore the *c.c.* between the two is expected to be on the order of  $1/\sqrt{N}$  leading to a bias of the same magnitude in the distribution of  $\rho^{spike}$ . In contrast, in the

multiple state scenario the fluctuations in the spontaneous map reside mostly in a low-dimensional space, so that the firing rates of all neurons are highly correlated with each other and with the SI, resulting in a large bias.

### Analytical Solution of the Ring Model

**Phase diagram of the ring model:** we consider the case in which the distribution of POs is perfectly uniform and there are no spatial correlations in the LGN input. After plugging equations (1) and (4) into equation (3) (all in the Experimental Procedures), we take the number of columns  $N$  to infinity, and average over the noise for every orientation  $\theta$ . This leads to the following deterministic dynamical equation for the average firing rate at time  $t$  of the columns with this PO, denoted  $m(\theta, t)$

$$\tau_0 \frac{d}{dt} m(\theta, t) = -m(\theta, t) + \sigma_n f \left( \left[ 2\lambda \int_{-\pi/2}^{\pi/2} \frac{d\varphi}{\pi} \cos 2(\theta - \varphi) m(\varphi, t) + T \right] / \sigma_n \right) \quad (\text{S3})$$

At this limit the sum over columns becomes an integral over orientations. The activation function  $f(\cdot)$ , appearing in equation (S3), is given by

$$f(x) = \int_{-\infty}^x \Phi(u) du; \quad \Phi(x) = \int_{-\infty}^x \frac{dz}{\sqrt{2\pi}} e^{-z^2/2}$$

where  $\Phi(\cdot)$  is the effective gain function of each column. We define the order

parameter of the model as  $m_2(t) = \left| \int_{-\pi/2}^{\pi/2} \frac{d\theta}{\pi} e^{-i2\theta} m(\theta, t) \right|$ . This parameter is the absolute

value of the fundamental Fourier mode in orientation space of the profile of firing rates in the network. Calculating the Fourier mode of equation (S3) yields a dynamical equation for this order parameter

$$\tau_0 \frac{d}{dt} m_2(t) = -m_2(t) + \sigma_n \int_{-\pi/2}^{\pi/2} \frac{d\theta}{\pi} f \left( \left[ 2\lambda m_2(t) \cos 2\theta + T \right] / \sigma_n \right) \cos 2\theta \quad (\text{S4})$$

The  $H$  regime in the phase diagram corresponds to the case where the steady-state solution  $m_2 = 0$  of equation (S4) is stable. The  $M$  regime corresponds to a stable solution where  $m_2 > 0$ . In the  $I$  regime  $m_2$  has no finite stable solution. The transition between the  $H$  and  $M$  regions is along the contour where the homogeneous solution ( $m_2 = 0$ ) loses its stability, given by  $T/\sigma_n = \Phi^{-1}(1/\lambda)$  where  $\Phi^{-1}(\cdot)$  is the inverse

gain function. The transition between the  $M$  and  $I$  regimes is given by the line  $\lambda = 2$ , and represents the line where the solution  $m_2 > 0$  diverges.

**Autocorrelation of the SI:** we calculated analytically the auto-correlation function (ACF) of the SI  $\rho$ , defined as  $ACF(\Delta) \equiv \langle \rho(t)\rho(t+\Delta) \rangle_t$ , in the linear regime of the ring model for large  $N$ . Neglecting the small fluctuations in the norm of the spontaneous maps, we obtain

$$ACF(\Delta) = \frac{\sigma_n^2}{N} \left[ \frac{\lambda(2-\lambda)}{(1-\lambda)\tau_0} \frac{\tau^{-1} e^{-|\Delta|/\tau} - \tau_\lambda^{-1} e^{-|\Delta|/\tau_\lambda}}{\tau^{-2} - \tau_\lambda^{-2}} + e^{-|\Delta|/\tau} \right]; \quad \tau_\lambda \equiv \frac{\tau_0}{1-\lambda} \quad (S5)$$

which means that: (a) the time scale of the fluctuations of the SI is determined by the larger of the two time scales: that of the noise  $\tau$  or that of the network  $\tau_\lambda$ ; and, (b) the temporal variance of the SI is given by

$$ACF(0) = \frac{\sigma_n^2}{N} \left[ 1 + \frac{\lambda(2-\lambda)}{(1-\lambda)^2 \left(1 + \frac{\tau_\lambda}{\tau}\right)} \right] \quad (S6)$$

This expression demonstrates that the variance can be increased by increasing the cortical gain  $\lambda$  (Figure 4B, inset) and for  $\lambda > 0$  also by increasing the time constant of the noise relative to that of the network.

#### Distribution of the SI in the Case of a Spherical Manifold

If we assume that the attractor manifold encodes  $d$  features (angles) and has the geometry of the surface of a  $d + 1$ -dimensional sphere then, at the limit of large  $N$ , the distribution of the SI between the network-state and a state of orientation, is given by (Figure 5H)

$$P_d(\rho) = \frac{(1-\rho^2)^{\frac{d}{2}-1}}{\int_{-1}^1 (1-x^2)^{\frac{d}{2}-1} dx} \quad (S7)$$

#### Measurement Noise

In our analysis, the only source of noise in the system is neural noise, arising presumably from the LGN. It may be argued that measurement noise could both change the shape and reduce the width of the distribution of the SI thereby blurring

the distinction between the predictions of the single and the multiple state scenarios. However, measurement noise is likely to induce only weak correlations between neighboring pixels in the optical maps. As a result, the contribution of this noise to the numerator of the SI is small, on the order of  $1/\sqrt{P}$  where  $P$  is the number of pixels.

However, the denominator of the SI that measures the overall variance of the image could increase substantially depending on the noise level. These considerations imply that the main effect of measurement noise is to decrease the width of the SI without significantly affecting its shape. Indeed, simulating our network with spatially white non-neuronal noise demonstrates that the shape of the distribution is hardly affected (Figure S1). As for its width, when the standard deviation of this noise equals that of the signal, the width of the distribution decreases by 30%. In particular, in the multiple state regime of the ring model, the SI distribution extends up to 0.7 instead of almost 1 but its shape remains essentially the same as in figure 4G.

#### Distribution of the SI during the Evoked State in the 2D Model

In the manuscript we focused on the statistics of the ongoing activity. However in the studies of Grinvald and colleagues the statistics of images that were evoked by oriented drifting gratings were analyzed, as well (Tsodyks et al., 1999; Kenet et al., 2003). This analysis showed that the distribution of the SI in the evoked state was substantially biased, as compared with the spontaneous SI. Additionally, the images in the evoked and spontaneous states were similar in terms of the maximal value of the SI and the amplitude of spatial modulation. This similarity is consistent with the multiple state hypothesis, which asserts that during the ongoing activity the system wanders between attractor states that are similar to the evoked states. However, it is less clear whether the scenario of a single state can exhibit such a similarity between the spontaneous and evoked activity.

To study this question we simulated our 2D model in the presence of the orientation modulated term in the LGN input and computed the SI with the average evoked map. The statistics of the evoked images depend on the strength of the orientation modulated input  $L\varepsilon$ . For weak inputs the evoked maps are similar to the spontaneous maps, in terms of the maximal SI and spatial modulation depth. But in contrast with the empirical results, in this regime the bias in the SI is very small. When the inputs are strong, the bias is considerable but the maximal evoked SI is substantially larger than the maximal spontaneous SI. Finally, for moderate input

strengths the bias is substantial and the evoked and spontaneous images are similar both in their maximal SI (see Figure S2) and in their spatial modulation depth (data not shown). In this regime the orientation modulated input is strong enough to drive the neurons in the anti-preferred orientation below threshold. This effectively reduces the gain of the intra-cortical connections and reduces the width of the SI. In conjunction with the bias in the SI, this leaves the maximal SI approximately unchanged. In conclusion, the empirical similarity between the statistics of the evoked and spontaneous images is consistent with the single state regime with a moderate strength of orientation modulated inputs.

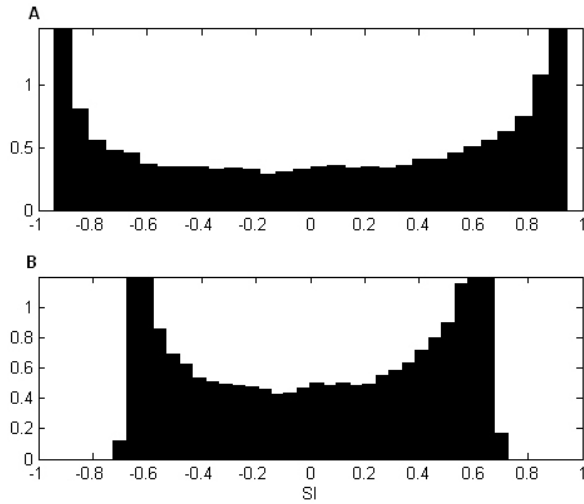
#### Kohonen Self-Organized Feature Mapping

In their recent paper, Kenet et al. contend to present novel evidence that the spontaneous optical imaging maps in V1 represent switching between cortical states (Kenet et al., 2003). This was done by applying to the ongoing maps an unsupervised feature mapping (FM) algorithm that produced a set of 40 templates that were highly correlated (at a level of  $\sim 0.8$ ) with maps evoked by oriented stimuli. Moreover, ordering the templates along a ring they concluded that the set of templates are representations of the underlying cortical states. They demonstrated that the time-series of the transition between the templates is smoother than and exhibits a time scale twice as long (82 ms) as a time-series generated using surrogate FMs (40 ms). If this analysis is valid, then not only does it reveal the presence of cortical states but it also shows that these states resemble the evoked states of orientation.

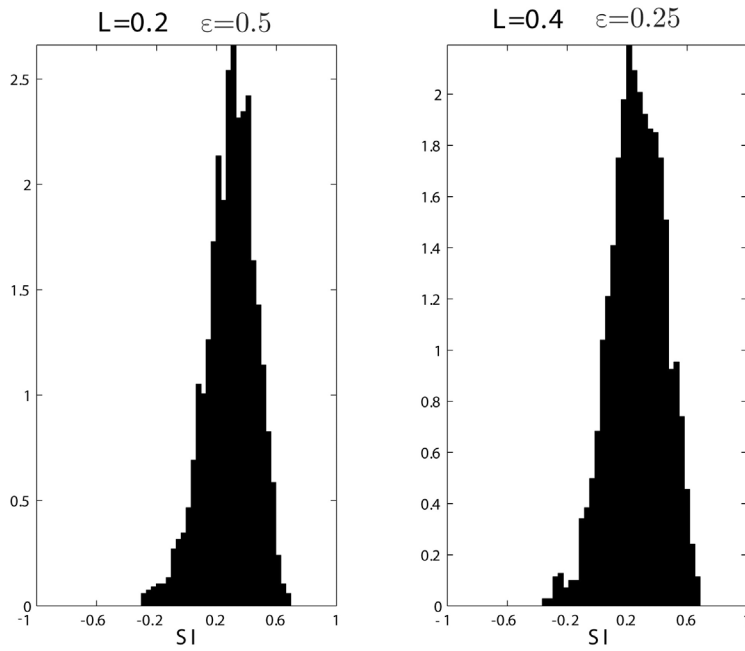
In order to test these interpretations we have applied the same FM algorithm, with comparable training parameters, to the spontaneous maps generated by the network dynamics in the single state regime of the 2D model for  $\lambda = 0.6$ . The fluctuating spontaneous maps in this regime are high-dimensional and Gaussian with an increased variance within the subspace of evoked maps. Each of the resulting 40 templates in our simulation attained a correlation of 0.8 with one of the evoked maps in the model (Figure S3, red circles in figure depict correlation of templates with a single evoked map), and were arranged in a manner corresponding to a monotonic rotation in orientation space. For comparison, we ran the algorithm also in the absence of cortical interactions ( $\lambda = 0$ ) in the 2D model. The resulting templates attained a correlation of 0.2 in this case (black points). We conclude that the FM algorithm detects the dimensions in the data with maximal variance even when this variance is small compared with the total variance.

We calculated the two time series of node numbers for the two cases of (1) moderate cortical feedback ( $\lambda = 0.6$ ), and of (2) no cortical feedback ( $\lambda = 0$ ) in the 2D model. We found time scales of 50 ms for  $\lambda = 0.6$  and 30 ms for  $\lambda = 0$  (calculated from the autocorrelation functions), which are comparable with the empirical time scales (compare Figure S4 to Figure 4D in Kenet et al.). In light of these findings, we conclude that the results of the FM algorithm are consistent with the single state hypothesis, and that this method is not appropriate for discovering multiple intrinsic dynamic states.

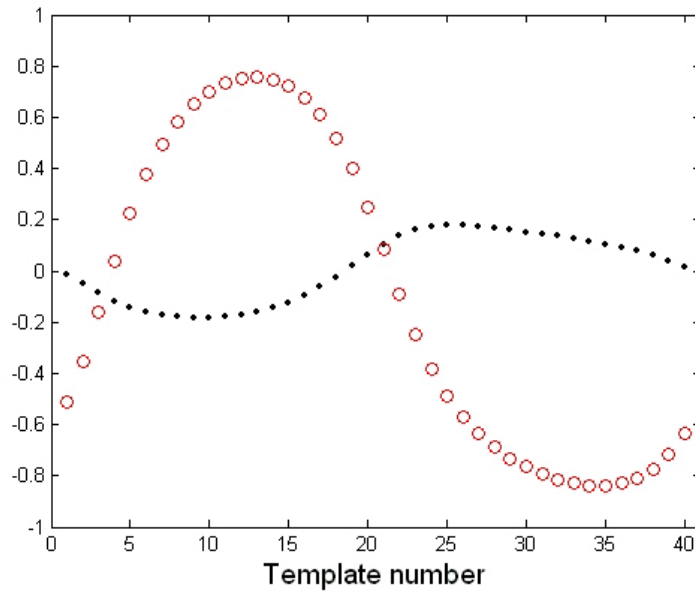
**Figure S1.** The effect of measurement noise on the SI distribution. (A) The distribution of the SI in the ring-model without noise. (B) The distribution of the SI with external white noise added to the spontaneous maps. The variance of the external noise is equal to the spatial variance of the “neural” map.



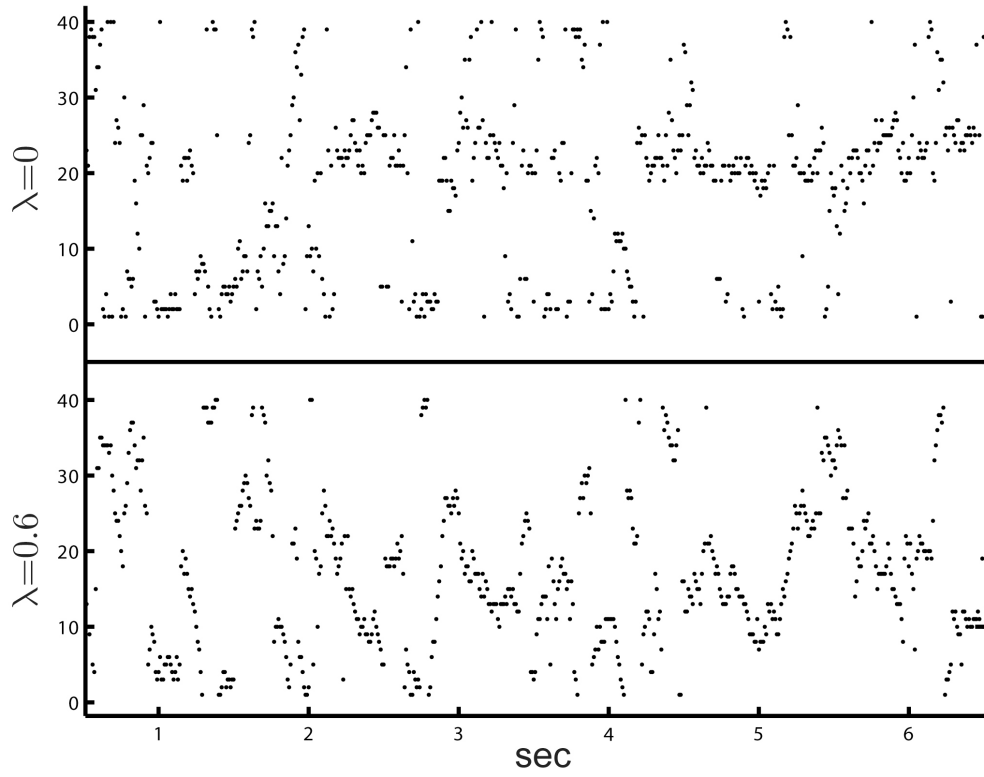
**Figure S2.** The distribution of the SI in the evoked state of the 2D model. Two examples of the distribution for moderate values of  $L$  and  $\varepsilon$ . The rest of the parameters are as in Figure 7 in the manuscript.



**Figure S3.** Correlation of an evoked map with the 40 templates derived in the single state of the 2D model. Red circles,  $\lambda = 0.6$ ; black points,  $\lambda = 0$ .



**Figure S4.** Time series of Kohonen Template numbers in the single state regime of the 2D model. Upper panel,  $\lambda = 0$ ; lower panel,  $\lambda = 0.6$ .



### **Supplemental References**

Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., and Arieli, A. (2003).

Spontaneously emerging cortical representations of visual attributes. *Nature* 425, 954–956.

Tsodyks, M., Kenet, T., Grinvald, A., and Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* 286, 1943–1946.