Interactions between Intrinsic and Stimulus-Evoked Activity in Recurrent Neural Networks

Kanaka Rajan*
Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544.
(Dated: October 4, 2010)

To make optimal use of experimental data on the anatomy and physiology of cortical circuits and to account for the effects of plasticity and neuro-modulation, we must understand the relationship between the synaptic connectivity and neuronal properties of a network, and the resultant activity. Given detailed properties of a complex network, can we predict what it will do? If we know how plasticity mechanisms or modulators change those properties, can we predict how activity will change? These are difficult questions to answer. My thesis explored how knowledge of synaptic connectivity and neuronal properties can be used to predict the activity produced by a simplified network model, using a mathematical framework.

We studied firing rate networks of N interconnected neurons in which the current for a single neuron is obtained by summing the firing rates of all its presynaptic partners, weighted by the strength of the synapse between them. Specifically, the firing rate of neuron i is given by $r_i = R_0 + \phi(x_i)$. The description of the model closes when we define a transfer function (f-I curve, an example shown in bottom left panel of Fig. 1) that determines the firing rate of each neuron as a function of the total current it receives, such that $\phi(x) = R_0 \tanh(x/R_0)$ for $x \leq 0$, and $\phi(x) = (R_{\max} - R_0)\tanh\left[x/(R_{\max} - R_0)\right]$, otherwise.

The simplified construction of these models allowed us to focus on the synaptic connectivity of the network as the quantity of interest. Synaptic connectivity has a strong influence on the dynamics of neural networks [1]. Synaptic strengths between all pairs of neurons are typically represented in a synaptic connectivity matrix, which contains $O(N^2)$ elements. Our resolution for experimental measurements currently limits us to certain statistical properties of groups of synapses (e.g., paired recordings in layer 5 cortical neurons indicate that synaptic strengths may exist in a log-normal distribution [9]), rather than estimates of all individual synaptic strengths. Therefore, we drew the non-zero elements of synaptic matrices (like in Fig. 1) independently and randomly from Gaussian probability distributions with zero mean

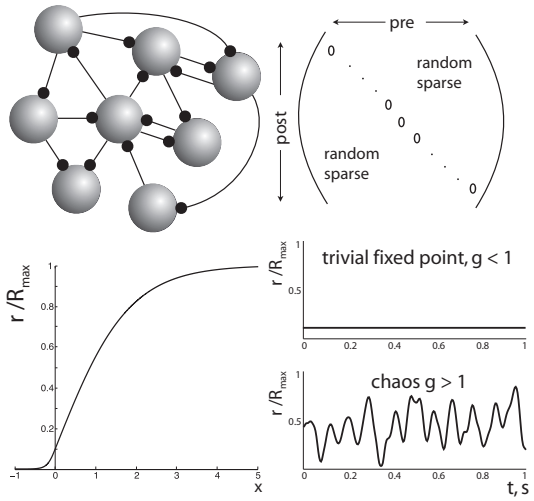

FIG. 1. Elements of a recurrently connected model neural network. **Top left panel:** Schematic of a model network of neurons each parameterized by a single firing rate. **Top right panel:** A synaptic connectivity matrix with the individual synaptic strength values for every neuronal pair randomly assigned from a Gaussian with 0 mean and variance of $O(1/N)$. **Bottom left panel:** The transfer function used in the network model. We use the value $R_0 = 0.1R_{max}$, reporting firing rates normalized by the maximum rate R_{max}. **Bottom right panel:** For no input ($I_i = 0 \ \forall\, i$), two patterns of spontaneous activity exist - trivial dynamics for $g < 1$ and chaos for $g > 1$.

and the variance given by $\left[J_{ij}^2\right]_J = g^2/N$, where the average runs over the Gaussian distribution of matrix elements. The synaptic input to neuron i was computed by multiplying the synaptic weight J_{ij} by the firing rate of presynaptic neuron j and summing over all j values. Neuron i could also generate an internal bias current b_i and receive an input term I_i representing the external drive. The activation variable for neuron i, x_i, is consequently determined by the equation, $dx_i/dt = -x_i + g\sum_j J_{ij} r_j - b_i + I_i$.

Provided that the synaptic strengths are not too large and there is no drive, simulating such a network of randomly-connected model neurons results in every unit firing at the same constant background rate (which corresponds to a trivial fixed point in the dynamics of the network). Though the trivial fixed point is persistently active in the absence of external drive, inducing more complex biologically realistic temporal dynamics requires manipulating the synaptic strength distribution in different ways [1].

* **krajan@princeton.edu**; Thesis Supervisor: L. F. Abbott, Columbia University in the City of New York. Research of KR and LFA was supported by an NIH Director's Pioneer Award, part of the NIH Roadmap for Medical Research and the Swartz Foundation through the Swartz Center at Columbia University. KR is currently a Biophysics Theory Fellow at Princeton University.

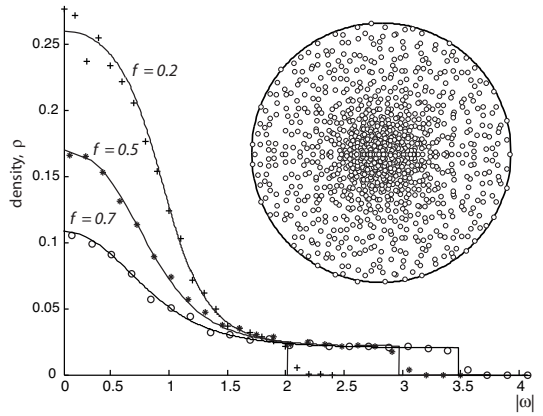

FIG. 2. Density ρ of eigenvalues as a function of position on the complex plane $|\omega|$, for $N = 1000$ for different fractions f of excitatory and inhibitory elements. The solid lines are the result of the analytic calculation and symbols are numerical results. The inset shows the spectrum of eigenvalues in the complex plane computed numerically for $f = 0.5$.

Eigenvalue Spectra of Random Matrices for Neural Networks

The dynamics of neural networks is influenced strongly by the spectrum of eigenvalues of the matrix describing the synaptic connectivity [1, 12, 13]. A classic result of random matrix theory is that, for large N, the eigenvalues of an $N \times N$ asymmetric random matrix lie uniformly within a unit circle in the complex plane [10], if the elements are chosen from a distribution with zero mean and variance $1/N$. This result does not apply to synaptic connectivity matrices because of the constraint that individual neurons are either excitatory or inhibitory. We computed eigenvalue spectra of large random matrices with fN excitatory and $(1-f)N$ inhibitory columns drawn from separate distributions with different means and equal or different variances [2]. First, we determined that the square of the radius of the circle containing the eigenvalues is N times the average of the variances of the excitatory and inhibitory distributions.

Second, we observed that the column-wise assignment of distributions has a dramatic effect on the actual spectrum of eigenvalues in the complex plane (inset of Fig. 2). We calculated the density of eigenvalues, ρ in the complex plane for such a matrix, averaged over the underlying probability distributions [2, 11]. When the location in the complex plane is represented by ω, ρ is determined by a potential, $\phi(|\omega|^2)$ through $\rho = \frac{1}{\pi}\left(|\omega|^2\phi'' + \phi'\right)$, where the primes denote derivatives with respect to $|\omega|^2$. Fig. 2 shows the results for different values of f. The value of f determines how many eigenvalues fall into the high-density central region, with higher central densities for smaller f values.

The eigenvalue distributions we calculated have several implications for neural network dynamics. Modifying the mean strengths of excitatory and inhibitory synapses has no effect on stability or small-fluctuation dynamics under balanced conditions. These are only sensitive to changing the widths of the distributions of excitatory and inhibitory synaptic strengths. If these widths are different, fewer eigenvalues will appear at the edge of the eigenvalue circle, meaning that there will be fewer slowly oscillating and long-lasting modes in the network dynamics. We concluded that having different cell types with different distributions of synaptic strengths has a large impact on network dynamics, and that the critical element to measure, and the critical element that may be modified by the modulatory and plasticity mechanisms that control neural circuit dynamics, are the variances of the synaptic strength distributions [3, 4].

Stimulus-dependent Suppression of Chaos in Recurrent Neural Networks

Increasing the variance of synaptic strengths within randomly connected neural networks induces complex autonomous activity by destabilizing multiple eigenvalues (this instability is eventually pulled back from explosion by the nonlinearity of the f-I curve). We can get nontrivial fixed points, complicated oscillatory behavior and even chaotic spontaneous activity from the network, simply by progressively increasing variability of the synaptic strengths across the network [4, 14–16].

How does an external input interact with ongoing dynamics in the different activity regimes as we move from weak to stronger synapses? Sensory response in neural systems is a result of a complex interplay between activity driven by external stimuli and internally generated activity, so the question is indeed pertinent. To model sensory stimuli and study evoked responses, we introduced oscillatory inputs. The amplitude of the input was always the same for each neuron, but in most of our studies we introduced a neuron-specific phase factor, θ_i, chosen randomly from a uniform distribution between 0 and 2π, so that the input to neuron i is $I\cos(2\pi f t + \theta_i)$. In visually responsive neurons, this input mimics a population of simple cells driven by a drifting grating of temporal frequency f, with the different phases arising from offsets in spatial receptive field locations.

We saw that as synapses get stronger, spontaneous activity increased in complexity, and the amplitude of response to the input grew. Indeed, supercritical networks exhibit complex spontaneous activity and equally complex activity in the presence of the input. To quantify this effect, we computed the power at the frequency of the stimulus in the network units that do not receive the input directly. This power grows smoothly as a function of synaptic variance and the variance at which the signal power is the highest corresponds to chaotic spontaneous activity in the network [3].

How is the stimulus decoded from complex evoked activity, knowing only that the signal power is extremely high in this regime? To explore these results analytically and more systematically, we developed dynamic mean-field equations appropriate for large N networks with the same connectivity statistics as our model [4, 14]. The the-

ory is based on the observation that the total recurrent synaptic input onto each network neuron can be approximated as Gaussian noise, such that $dx_i^1/dt = -x_i^1 + \eta_i$. The temporal correlation of this noise can be calculated self-consistently from the autocorrelation function of the network, $C(\tau) = \langle \phi(x_j(t))\phi(x_j(t+\tau)) \rangle$ (averaged over J and θ) because the recurrent synaptic input is generated and received by the same population of neurons.

We found that inputs not only drive network responses, they also actively suppress ongoing activity, ultimately leading to a phase transition in which chaos is completely eliminated (bottom panel of Fig. 3). The critical input intensity at the phase transition is a non-monotonic function of stimulus frequency, revealing a resonant frequency at which the input is most effective at suppressing chaos even though the power spectrum of the spontaneous activity peaks at zero and falls exponentially. Our analysis predicts that the variance of neural responses should be most strongly suppressed at frequencies matching the range over which many sensory systems operate [4–8].

To study the implications of the phase transition further, we divided network responses into signal and noise components by splitting the full response variance into chaotic and driven oscillatory components, i.e., $C(0) = \sigma^2_{\rm chaos} + \sigma^2_{\rm osc}$. The chaotic variance, $\sigma^2_{\rm chaos}$, is defined as the difference between the full variance, $C(0)$ and the variance, $\sigma^2_{\rm osc}$ due to the periodic oscillations. We call $\sigma_{\rm osc}$ the signal amplitude and $\sigma_{\rm chaos}$ the noise amplitude. In the frequency domain, $\sigma^2_{\rm osc}$ measures the total power in the network activity at the stimulus frequency and its harmonics, whereas $\sigma^2_{\rm chaos}$ measures the residual power.

The signal amplitude increases linearly with the strength of the input (I) over the range considered in Fig. 4. The noise amplitude has a more complex nonlinear dependence, reflecting the presence of the phase transition we saw in Fig. 3, in which a sufficiently strong input completely suppresses the chaotic component of the response. An interesting feature to note is that there is no clear signature of this chaotic-to-periodic transition in the signal amplitude. When plotted as a function of input frequency for fixed I, the signal amplitude shows relatively weak frequency dependence below about 4 Hz and then rolls off at higher frequencies (right panel of Fig. 4). This is a result of the low-pass filtering property of the network. The noise amplitude has a more interesting dependence. Between 0 and 3 Hz, the noise amplitude drops steeply and vanishes for frequencies between 3 and 7 Hz, rising again above 7 Hz. This double transition is a consequence of the non-monotonicity of the phase transition curves in Fig. 3. There is no apparent indication of these transitions in the signal amplitude.

Variability in cortical responses is sometimes described by adding stochastic noise linearly to a deterministic response (top panel of Fig. 3). Our results indicate that the interaction between intrinsically generated *noise* and responses to external drive is highly nonlinear. Near the onset of chaos, complete noise suppression can be achieved with relatively low amplitude inputs, weaker,

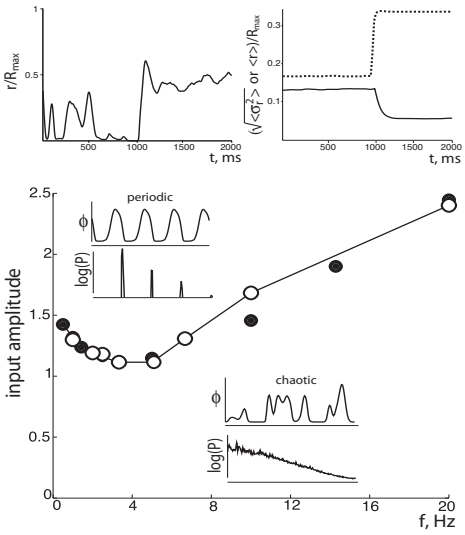

FIG. 3. **Top panel:** Adding external drive to a model that exhibits chaotic spontaneous activity results in a sharp drop in variance at stimulus onset, but with only partial suppression of response variability in the presence of additional stochastic noise. Firing rates and response variability normalized by the maximum firing rate showing an increase in average firing rate (dashed trace) and a decrease in response variability (solid trace) in the presence of external input (note that this variability does not go to zero). **Bottom panel:** Phase transition curve showing the critical input amplitude that divides regions of periodic and chaotic activity as a function of input frequency. Inset traces show representative firing rates for the regions indicated; along with the logarithm of the power spectrum of the activity across the network. A comparison of the phase transition curve computed analytically by mean-field theory (open circles) and by numerically simulating a network (filled circles) of $N = 10,000$ shows excellent agreement.

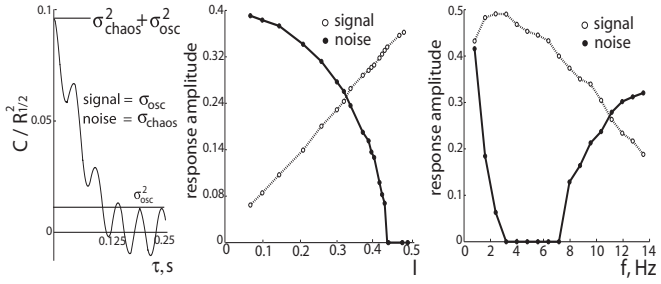

FIG. 4. Signal and noise amplitudes as a function of input strength and temporal frequency. Open circles denote the noise amplitude and filled circles the signal amplitude. **Left panel:** An example mean-subtracted autocorrelation function where the horizontal lines indicate the definitions the signal and noise amplitudes. Parameters used for this figure are $I/I_{1/2} = 0.4$, $g = 1.8$ and $f = 20Hz$. **Middle panel:** Signal and noise amplitudes for $f = 20$ Hz and $g = 1.5$ as a function of input amplitude. The transition from chaotic to non-chaotic regimes occurs at $I_c = 0.44$. **Right panel:** Signal and noise amplitudes for fixed input strength ($I = 0.2$) and varying the temporal frequency of the external drive. In the region between 3 and 7 Hz, responses of the network are completely free of chaotic noise. Unlike the top panel of Fig. 3, there is no additive stochastic noise in the system.

for example, than the strength of the internal feedback. Thus, suppression of spontaneously generated *noise* in neural networks does not require stimuli so strong that they simple overwhelm fluctuations through saturation.

Neuronal selectivity to stimulus features is often studied by determining how the mean response across experimental trials depends on various stimulus parameters. The presence of nonlinear interactions between stimulus-evoked and spontaneous fluctuating activity indicates that response components that are not locked to the temporal modulation of the stimulus may also be sensitive to stimulus parameters. In general, our results suggest that experiments studying the stimulus-dependence of the noise component of neural responses provide important insights into the nature and origin of activity fluctuations in neuronal circuits, as well as their role in neuronal information processing.

Inferring Stimulus Selectivity from Spatial Structure of Neural Network Dynamics

In the non-chaotic regime, we have seen that the temporal structure of network responses is largely determined by the input; they both oscillate at the same frequency, although the network activity includes harmonics not present in the input. The input does not, however, exert nearly as strong control on the spatial structure of the network response. The phases of the firing-rate oscillations of network neurons are only partially correlated with the phases of the inputs that drive them, and they are strongly influenced by the recurrent feedback. Spontaneous activity is a useful indicator of recurrent effects, because it is completely determined by network feedback.

How are the spatial patterns of spontaneous and evoked population responses related? We studied the impact of connectivity on the spatial pattern of fluctuations in the input-generated response, by comparing the distribution of evoked and intrinsically generated activity across the different units of a neural network [8]. We developed a complementary approach to Principal Component Analysis (PCA) in which separate high-variance directions are typically derived for each input condition [17]. We analyzed subspace angles to compute the difference between the shapes of trajectories corresponding to different network states, and the orientation of the low-dimensional subspaces that driven trajectories occupy within the full space of neuronal activity. We concluded that the absence of a detailed spatial map of afferent inputs and cortical connectivity does not limit our ability to design spatially extended stimuli that evoke strong responses [3, 4, 8].

Our analysis shows that even in cortical areas where underlying connectivity does not exhibit systematic topography [18], dissecting spatial patterns in neuronal fluctuations yields important insight into intrinsic network dynamics and stimulus selectivity. As a first pass example, we can see how the projections of network activity onto the PC directions fluctuate more rapidly for higher components revealing the interaction between spatial and temporal structure of the chaotic fluctuations (Fig. 5, top right panels).

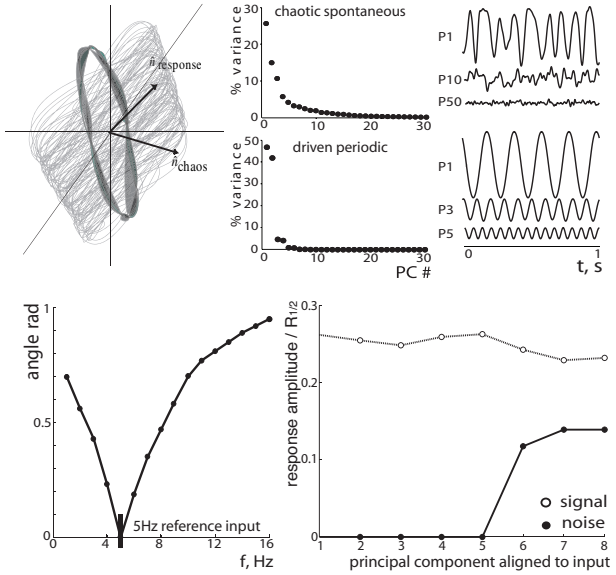

FIG. 5. Spatial pattern of network responses. **Top left panel:** Schematic of the angle between the subspaces defined by the first 2 components of the chaotic activity (grey) and a 2D description of the periodic orbit (black curve). **Top right panels:** PCA of the chaotic spontaneous state and non-chaotic driven state reached when an input of sufficiently high amplitude has suppressed the chaotic fluctuations. % variance accounted for by different PC's for chaotic spontaneous activity. Projections of the chaotic spontaneous activity onto PC vectors 1, 10 and 50 (in decreasing order of variance). For non-chaotic driven activity, projections of periodic driven activity are shown for PC's 1, 3, and 5. Projections onto components 2, 4, and 6 are identical but phase shifted by $\pi/2$. $N=1000$, $g=1.5$, $f=5$ Hz and $I/I_{1/2}=0.7$. **Bottom left panel:** Effect of input frequency on the orientation of the periodic orbit. Angle between the subspaces defined by the 2 leading PC's of non-chaotic driven activity at different frequencies and these two vectors for a 5 Hz input frequency. $N=1000$ and $I/I_{1/2}=0.7$ and $f=5$ Hz, $I/I_{1/2}=1.0$. **Bottom right panel:** Network selectivity to different spatial patterns of input. Signal and noise amplitudes in the input-evoked response aligned to the leading PC's of the spontaneous activity of the network. $N=1000$, $I/I_{1/2}=0.2$ and $f=2$ Hz. Chaos is completely suppressed only when input is aligned to the PC vectors with the 5 largest eigenvalues.

The top left panel of Fig. 5 illustrates the technique we developed to determine where the orbit of driven periodic activity lies in the full N-dimensional space of neuronal activities, relative to the trajectory of the chaotic spontaneous background. The analysis was done by diagonalizing the equal-time cross-correlation matrix of network firing rates given by, $D_{ij} = \langle (r_i(t) - \langle r_i \rangle)(r_j(t) - \langle r_j \rangle) \rangle$ where the average runs over t. The eigenvalues of this matrix expressed as a fraction of their sum, indicate the distribution of variances across different orthogonal directions in the activity trajectory. PCA revealed that despite the fact that network connectivity is a full rank matrix, the effective dimensionality of evoked states, even

while chaotic, is progressively smaller. In the example shown in the top right panels of Fig. 5, the leading 10% of PCs account for 90% of the total variance, whereas the non-chaotic driven state describes a 2D circular orbit.

We visualized subspace angles to reveal not only the difference in the shapes of trajectories corresponding to different network states, but also the difference in the orientation of the low dimensional subspace of these trajectories within the full space of neuronal activity. For two subspaces of dimension d_1 and d_2 defined by the orthogonal unit vectors V_1^a, for $a = 1, 2, \ldots d_1$ and V_2^b, for $b = 1, 2, \ldots d_2$, the cosines of the principal angles are equal to the singular values of the $d_1 \times d_2$ matrix $V_1^a \cdot V_2^b$. The angle between the two subspaces is the inverse cosine of the largest singular value of this matrix [19]. The lower left panel of Fig. 5 shows how a 5Hz reference frequency causes the orbit of driven activity to rotate as input frequency changes.

Finally, in the lower right panel of Fig. 5, we show the selectivity of the network in terms of signal and noise amplitudes in response to inputs aligned to the leading PCs of the spontaneous activity (i.e., to different spatial input patterns). Inputs aligned to the first 5 principal components of the spontaneous activity completely suppress the chaotic noise, resulting in periodic driven activity. For higher-order PCs, the network activity is chaotic. The point $a = 5$ corresponds to a phase transition analogous to that seen at $f = 7$ Hz in Fig. 4 (see also [20]). The noise shows more sensitivity to the spatial structure of the input than the signal, just as it did to its temporal structure in Fig. 4. Based on these analyses we can design spatially extended stimuli that evoke strong responses.

Our results show that in addition to revealing how the spatiotemporal structure of spontaneous activity affects input-evoked responses, these methods can be used (a), to infer input selectivity induced by network dynamics from experimentally accessible measures of spontaneous activity (e.g., from voltage- or calcium-sensitive optical imaging experiments) and (b), to design stimuli that evoke strong sensory responses. This is particularly true when selectivity is measured in terms of the ability to entrain the neural dynamics. Therefore, analysis of spontaneous activity can provide valuable information about the computational implications of neuronal circuitry.

[1] T. P. Vogels, **K. Rajan** & L.F. Abbott, *Annual Review of Neuroscience,* **28***: 357 (2005).*

[2] **K. Rajan** & L. F. Abbott, *Physical Reviews Letters,* **97***: 188104 (2006).*

[3] **K. Rajan**, *Ph.D. Dissertation, Columbia University in the City of New York (2009).*

[4] L. F. Abbott, **K. Rajan** & H. Sompolinksy, *The Dynamic Brain: An Exploration of Neuronal Variability and its Functional Significance, M. Ding and D. Glanzman eds., Oxford University Press (2009).*

[5] **K. Rajan**, L. F. Abbott & H. Sompolinksy, *Physical Reviews E.,* **82***: 01193 (2010).*

[6] **K. Rajan**, *Grace Hopper Celebration of Women in Computing, published by the Anita Borg Institute for Women & Technology and the Association for Computing Machinery (2010).*

[7] **K. Rajan**, L. F. Abbott & H. Sompolinsky, *Biomed Central Neuroscience, 11,* **O17***: 11 (2010).*

[8] **K. Rajan**, L. F. Abbott & H. Sompolinsky, *Advances in Neural Information Processing Systems (2010).*

[9] S. Song, P.J. Sjöström, M. Reigl, S.B. Nelson & D.B. Chklovskii, *PLoS Biology* **3***, 505, (2005).*

[10] V. L. Girko, *Theory of Probability and Its Applications (USSR)* **29***, 694 (1985).*

[11] H. J. Sommers, A. Crisanti, H. Sompolinsky & Y. Stein, *Physical Reviews Letters* **60***, 1895, (1988).*

[12] H. R. Wilson and J. D. Cowan,*Biophysics Journal* **12***, 1, (1972).*

[13] O. Shriki, D. Hansel & H. Sompolinsky, *Neural Computation* **15***, 1809, (2003).*

[14] H. Sompolinsky, A. Crisanti & H.J. Sommers, *Physical Reviews Letters* **61***, 259-262 (1988).*

[15] C. van Vreeswijk & H. Sompolinsky, Science **274**, 1724-1726 (1996).

[16] N. Brunel, Journal of Physiology, Paris **94**, 445-463 (2000).

[17] B. M. Broome, V. Jayaraman & G. Laurent, *Neuron* **51***, 467-482 (2006).*

[18] D.H. Hubel & T. N. Wiesel, *Journal of Physiology* **160***, 106-154 (1962).*

[19] I. C. F. Ipsen & C. D. Meyer, *American Mathematics Monthly 102, 904-911 (1995).*

[20] N. Bertchinger & T. Natschlger, *Neural Computation* **16***, 1413-1436 (1995).*