

Background-activity-dependent properties of a network model for working memory that incorporates cellular bistability

Christopher P. Fall¹, Timothy J. Lewis^{1,2}, John Rinzel^{1,2}

¹ New York University Center for Neural Science, New York, NY 10003, USA

² Courant Institute of Mathematical Science, New York University, New York, NY 10012, USA

Received: 25 June 2004 / Accepted: 14 December 2004 / Published online: 1 April 2005

Abstract. In models of working memory, transient stimuli are encoded by feature-selective persistent neural activity. Network models of working memory are also implicitly bistable. In the absence of a brief stimulus, only spontaneous, low-level, and presumably nonpatterned neural activity is seen. In many working-memory models, local recurrent excitation combined with long-range inhibition (Mexican hat coupling) can result in a network-induced, spatially localized persistent activity or “bump state” that coexists with a stable uniform state. There is now renewed interest in the concept that individual neurons might have some intrinsic ability to sustain persistent activity without recurrent network interactions. A recent visuospatial working-memory model (Camperi and Wang 1998) incorporates both intrinsic bistability of individual neurons within a firing rate network model and a single population of neurons on a ring with lateral inhibitory coupling. We have explored this model in more detail and have characterized the response properties with changes in background synaptic input I_o and stimulus width. We find that only a small range of I_o yields a working-memory-like coexistence of bump and uniform solutions that are both stable. There is a rather larger range where only the bump solution is stable that might correspond instead to a feature-selective long-term memory. Such a network therefore requires careful tuning to exhibit working-memory-like function. Interestingly, where bumps and uniform stable states coexist, we find a continuous family of stable bumps representing stimulus width. Thus, in the range of parameters corresponding to working memory, the model is capable of capturing a two-parameter family of stimulus features including both orientation and width.

1 Introduction

There is much interest in how the brain can maintain the persistent neural activity that encodes recent stimuli and

is thought to be the basis of working memory (Fuster 1988; Goldman-Rakic 1995). During behavioral tasks, this persistent elevated neuronal firing can last for tens of seconds after the stimulus is no longer present. Such persistent activity appears to maintain a representation of the stimulus until the response task is completed. Several theoretical mechanisms for the maintenance of persistent activity have been described, including local recurrent synaptic feedback and intrinsic cellular bistability (Durstewitz et al. 2000; Wang 2001). Of these, recurrent connectivity at the local circuit level has received the most attention, from Hopfield models to biophysically elaborate, cell- and conductance-based networks.

Wilson and Cowan suggested that meaningful insight into the behavior of neural ensembles might be gained by a mean field approach describing the space and time coarse-grained activity of local populations of excitatory and inhibitory neurons (Wilson and Cowan 1972). They showed that such firing rate models, in a spatially distributed system with local recurrent excitation and wider-range inhibition, could result in stable, spatially localized, “bump-like” patterns of activity (Wilson and Cowan 1973). Reducing from a two to a single-population formulation by incorporating a “Mexican hat” connectivity kernel, Amari proved the existence and stability of standing “bump-like” patterns of activity in such a rate model (Amari 1977). In particular, he demonstrated the crucial dependence of stable patterns on the imposed background firing activity of the network. Sompolinsky and coworkers also found bump solutions for a scalar model with a combined excitation-inhibition kernel (Ben-Yishai et al. 1995), and they used this property to code for visual orientation tuning on a ring geometry.

Amari identified a parameter regime of kernel properties (the amount of excitation and inhibition) and uniform background activity where a stable bump state coexisted with a stable uniform state. This kind of network bistability is thought to be necessary in a model for working memory. That is, a brief transient input can induce a resting network to evolve into a bumplike pattern. Such a bump is then susceptible to being returned to the rest state by a transient memory-erasing perturbation. Bump states

Correspondence to: C. P. Fall
(e-mail: fall@cns.nyu.edu,
Tel.: +1-212-9987775)

in models formulated on infinite spatial domains or rings have translation invariance, and therefore a bump's location is viewed as an encoded stimulus feature.

There is now renewed interest in the concept that individual cortical neurons might have some inherent ability to sustain persistent activity in response to a transient stimulus *without* network recurrence. The remarkable finding that individual cortical neurons can sustain graded persistent firing in the entorhinal cortex (Egorov et al. 2002) suggests that neurons elsewhere in the central nervous system may be capable of similar behavior. Indeed some spinal cord neurons show properties that have been modeled as intrinsic bistability (Booth and Rinzel 1995). At the intersection of Amari-type network bistability and intrinsic cellular bistability is a working-memory model by Camperi and Wang (1998). The Camperi–Wang (C–W) formulation comprises an integrodifferential network of conditionally bistable cells on a ring with a Mexican-hat-like synaptic weight kernel. Here the intrinsic dynamics for individual cells is bistable rather than the graded multistability seen in the experiments of Egorov et al. (2002). The central insight from their work is that intrinsic bistability in single neurons provides robustness against noise and distracters as compared with the Amari- or Ben-Yishai-type model. Here we reconsider the C–W model and explore the response properties as a function of the background input rate, I_o , and stimulus width, p .

Our computational results suggest that bistable working-memory properties of the C–W model are found in only a relatively restricted range of I_o , further restricting the already limited range of parameters supporting bistability previously found by Camperi and Wang (1998). Background network activity is therefore a crucial parameter for this model. We found bumps and uniform states in C–W coincide only in a small range at low I_o values. Over a higher and much larger I_o range the C–W model is not bistable, but instead its bump solution is a global attractor that encodes a stimulus-specific location. Here the uniform resting state is not stable, and instead the model only holds the memory of the most recent spatially patterned transient input. Increasing I_o further we pass through another range of bistability and then into a regime where only a uniform elevated steady state exists. We show that the behavior within the regions of bistability is complicated and includes multistability and Turing-type low-amplitude bump patterns.

2 Methods

Camperi–Wang is a firing rate model designed to simulate the encoding of visuospatial cue orientation in a working memory task. The essential features of the C–W model are a circular domain, an intrinsic firing rate that is an S-shaped function of input, and an input function comprised of a thresholded convolution of other units' firing rates with a connectivity kernel that spans the entire circular domain. The S-shaped gain function gives the individual unit conditional bistability. This means that isolated units do not have two stable states at rest, but some steady input can bring individual units into the bistable range.

When coupled in the network, input consists of a constant imposed background I_o , a transient input pattern I_{cue} , and the recurrent input from other neurons in the system. As Camperi and Wang pointed out, the network (in a certain parameter regime) can sustain Amari-type bump activity without conditional bistability; however, the resistance to noise and distracting input patterns is lost. Maintaining a bump solution without conditional bistability requires enhanced background activity and/or tuning of the relative level of excitation to inhibition.

The equation describing the dynamics of firing activity $r = r(\theta_i, t)$ of each neuron is

$$\tau_0 \frac{dr}{dt} = -f(r) + g(I),$$

where $\theta_i = i \frac{2\pi}{N}$, $i = 1, \dots, N$, and τ_0 sets the time scale.

Throughout this report we show the dimensionless activity levels: Camperi and Wang multiplied this rate by 7 Hz in order to obtain reasonable spontaneous and active firing rates, but this does not affect the behavior of the model, and we ignore it here. In the C–W simulations, activity was turned off by an externally imposed transient global inhibition.

The key expression in the model is the cubic-shaped function of the firing rate, which, under appropriate parameters, allows for intrinsic bistability:

$$f(r) = c + r - ar^2 + br^3.$$

The input to the neurons consists of the convolution of the weights and firing rates plus the external input:

$$I(\theta_i, t) = I^{\text{ext}}(\theta_i, t) + \sum_{j=1}^n \frac{1}{N} W(\theta_i - \theta_j) r(\theta_j, t).$$

The external input consists of a uniform, steady background synaptic activity (I_o) plus a transient, nonuniform stimulus (I_{cue}):

$$I_{\text{ext}}(\theta, t) = I_o + I_{\text{cue}}(t) \left(\frac{1 + \cos(\theta)}{2} \right)^p.$$

The shape of the transient stimulus pattern is controlled by the positive exponent p ; a larger p results in a narrower stimulus pattern. As with the C–W simulations, I_{cue} is typically presented at a constant level for 0.5 s and is then zero. We note that the duration of presentation can have an effect on the final pattern of activity, but we have not explored this here.

In the model, the net input I , including both external and recurrent components, is thresholded such that $g(I)$ cannot contribute to making $dr/dt < 0$:

$$g(I) = I: \quad I > 0, \\ g(I) = 0: \quad I \leq 0.$$

The synaptic weight kernel is given by

$$W(\theta) = -W_I + W_E \left(\frac{1 + \cos(\theta)}{2} \right),$$

which is very similar to the stimulus shape relation.

Explorations of the model were conducted using custom code written in MATLAB (R13) and simulated on an Apple Macintosh G5 computer. This code is available

Symbol	Parameter	C–W condition	Range this manuscript
τ_0	Integration time constant	0.025 s	0.025 s
p	Stimulus shape exp	1	0.01–1000
I_o	Background activity	0.45	0–6
I_{cue}	Stimulus amplitude	1	0.1–2
a	Nonlinearity parameter	0.36	0.36
		(bistable)	
b	Nonlinearity parameter	0.038	0.038
c	Nonlinearity parameter	–0.2	–0.2
W_I	Inhibitory strength	2	2
W_E	Excitatory strength	2.6	2.6

upon request from the authors. A forward Euler integration scheme was used with a time step of 0.001s, and a smaller time step did not affect the results. Typically, 128 neurons were simulated in the data presented here, though more or fewer neurons were tested to assess any potential effects. As far as we were able to see, more neurons resulted in smoother curves but no change in qualitative results. Stability of the steady state bumps was assessed numerically by running simulations for long time periods (~ 50 s).

3 Results

3.1 Dynamics of bump generation

The most prominent property of the C–W model is the generation of a bump in response to a brief stimulus. A typical simulation with standard C–W parameters is shown in Fig. 1. The spatiotemporal evolution of the network activity toward the bump state has some notable features. Several of these features are associated with the time points indicated by letters in Fig. 1a, and these correspond to later panels. Our simulations begin with all cells at zero firing rate. The initial transient represents the network settling to an activity level determined by the background input strength, I_o . At time “d” the “standard” stimulus ($p = 1$, $I_{\text{cue}} = 1$, held for 0.5 s) initiates a bump pattern with exaggerated amplitude. At time “e” the stimulus is released and the network begins to relax to the steady state bump profile. The transient phase is sampled at time “f,” and the network has approached steady state by time “g.” Salient features of the steady state bump include a sharp transition in activity from neurons in the up state to neurons in the down state. At the edges of the bump and moving outward begins a graded shelf of slightly elevated neurons that transitions into a uniform shelf of neurons held at the lowest possible level of activity by strong lateral inhibition and the threshold on the input function $g(I)$.

The spatial profiles of the bump state (solid), the connection weight kernel (dotted), and the brief stimulus (dashed) are plotted in Fig. 1b. The connection strength is from the *central* neuron to adjacent neurons; negative values indicate net inhibition. This connection pattern is identical, but shifted, for each neuron in the field. The exponent p determines the width of the stimulus; large p is for a narrow stimulus.

To understand the bump profile, we first focus on the single-unit dynamics. The effect of constant input to an isolated neuron can be seen graphically (Fig. 1c) by plotting the rate of change of activity (dr/dt) with the processed input g treated as a parameter. The zero crossings correspond to steady state firing rates for a given g . Open circles in this plot indicate where $dr/dt = 0$ for $g = 0.5$. The upper and lower activity states are stable because the slope of the dr/dt curve is negative at these points, while the intermediate steady states are unstable because the slope there is positive. For $g = 0$ (no input to a neuron), there is a nonzero basal firing rate. Increasing g effectively shifts the dr/dt curve upward such that a neuron passes through a range of bistability and eventually to a single elevated firing state.

Within an intact network, the input g is not necessarily uniform. At each time t , each neuron has an r and g value and the network state is represented by a set of (g, r) pairs. The set may be viewed in the g – r plane, and hence the network’s evolution appears as a moving set of points. The snapshots in Fig. 1d–g of the g – r plane correspond to the labeled time points in Fig. 1a. The superimposed solid curve indicates the possible steady state values of a single cell as a function of g (from Fig. 1c). The curve’s S-shape reflects the intrinsic cellular conditional bistability, which does not change with time. At steady state, each neuron in the set lies somewhere along this S-shaped curve, but a neuron may or may not lie on the curve during a transient phase.

During the stimulus, neurons around $\theta = 0$ are forced to high activity and spread rightward along the upper branch, while some neurons are in transition between the upper and lower branch and therefore are not on the S-shaped curve (Fig. 1e). When the stimulus is released (Fig. 1f), neurons relax and those without sufficient excitatory input drop off the upper branch. Some neurons remain on the upper branch, thereby establishing a steady state bump solution. After a stable bump is established (Fig. 1g), the more active neurons support those with lower firing rates along the lower branch of the S-curve. These latter neurons comprise the graded shelf seen on the edge of the bump in Fig. 1a. The remainder of the neurons are suppressed to a uniform low firing rate due to the massive lateral inhibition in the connectivity kernel.

3.2 Bump shapes are not unique

For the standard C–W parameters, the steady bump profile depends on the stimulus shape. A very wide stimulus ($p = 0.01$, shown dashed in Fig. 2a) results in a bump that is essentially the same shape as the standard stimulus ($p = 1$, cf. Figs. 2a and 1b). Narrow stimuli ($p > 1$) result in narrow stable bumps, as shown in Fig. 2b for $p = 100$. The bump profiles that result from a variety of stimulus widths are shown in Fig. 2c. The range of widths in this family of bump patterns decreases from a maximum down to a minimum (that is larger than one cell wide). Stimuli that are too narrow ($p = 500$ shown) fail to produce a stable bump for a stimulus of amplitude 1. Interesting features that arise with the narrow bump patterns include a diminished

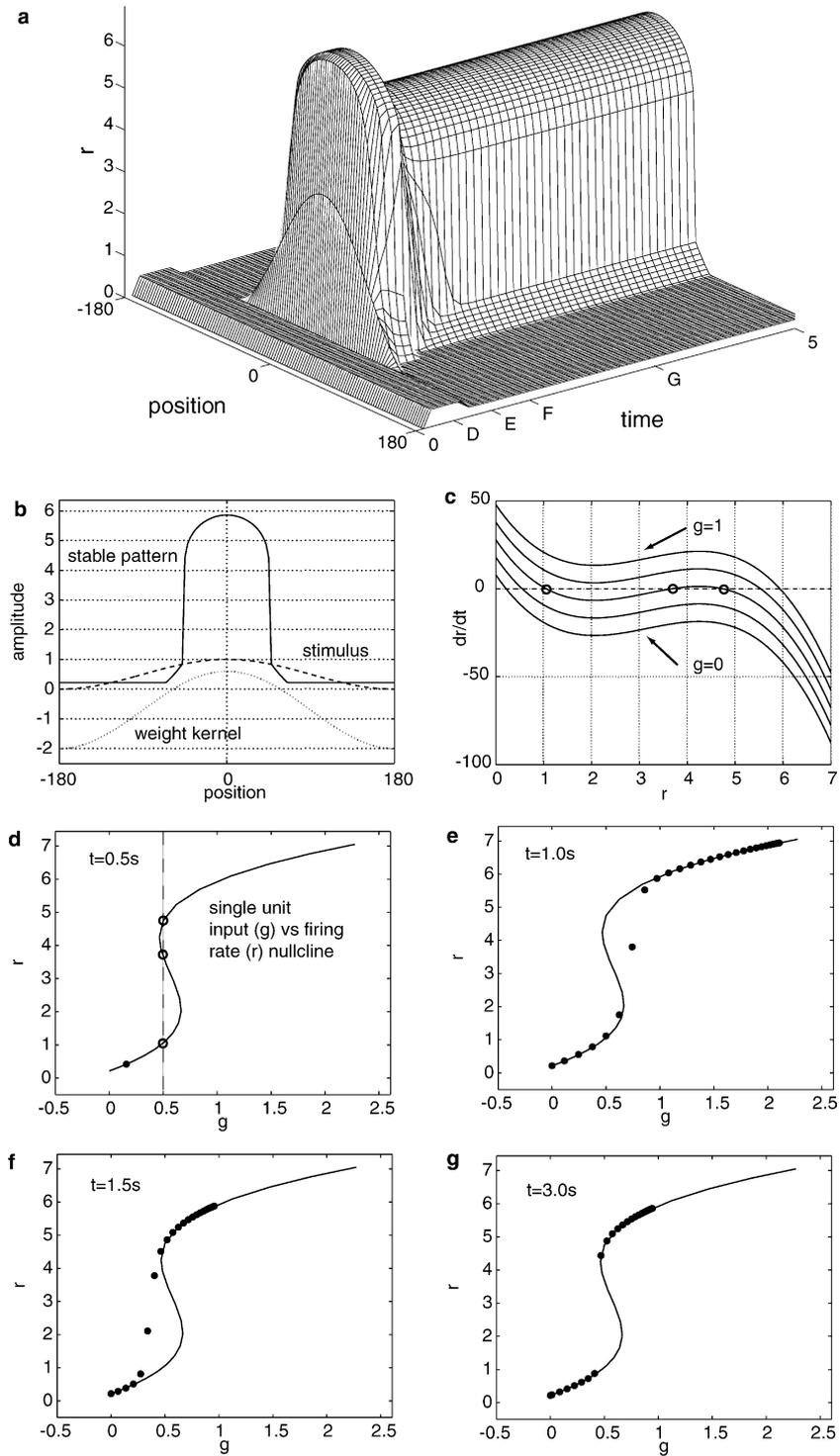


Fig. 1. C–W model behavior under standard parameters, with $I_o = 0.45$. **a** Firing rate as a function of space and time, showing bump pattern. The labels *D* through *G* along the time axis correspond to time points for panels **d–g**: 0.5, 1.0, 1.5, and 3 s. **b** Overlaid plots showing (*dots*) connection weight value as a function of distance from any given neuron, (*dash*) stimulus imposed for 0.5 s at $t = 0.5$ s, (*solid*) the firing rate as a function of space at $t = 5$ s. Note that the connectivity is strongly negative (mean -0.7) and, due to the threshold on $g(I)$, neurons outside of the bump are at $r \approx 0.2164$. **c** Intrinsic dynamics for a single isolated neuron with fixed input (g). Increasing g shifts the curve for dr/dt up such that individual neurons pass from having a single lower stable steady state through a range of bistability and eventually to a single elevated stable steady state. Intersection of

these curves with the *dashed* line indicating $dr/dt = 0$ corresponds to a steady state activity level. Shown are curves for $g = 0, 0.25, 0.5, 0.75, 1$. **d–g** Snapshots of network activity in the g - r plane. The *solid* curve indicates where input, g , and the firing function $f(r)$ are balanced for a single neuron. This curve comes from plotting the zero crossings in panel **c** as a function of g . The evolution of the network firing pattern is represented at various time slices, in panels **d–g**: *solid* dots indicate the distribution of all neurons on the r - g plane at various time points for standard conditions shown in **a**. **d** 0.5 s (just prior to stimulus). **e** 1.0 s (at end of stimulus). **f** 1.5 s. **g** 3.0 s. These time points are indicated on the time axis of panel **a**. At steady state, these *dots* must lie on the r - g nullcline

maximum amplitude, a smaller range of amplitudes for the high activity neurons, and an increased “shelf” of neurons in the low activity state (on the lower branch of the S-curve).

The state of the various neurons for wide and narrow stimuli can be seen better in the $g-r$ planes of Fig. 2d,e. The shape of the stable bump response depends strongly on stimulus width and only weakly on the amplitude (provided it is above threshold). If the network is stimulated with a square pulse instead of a bump-shaped cue profile, so that amplitude and width can be controlled separately, only the width has a significant effect (not shown). From the $g-r$ profiles we see that the transition from the lower to the upper branch does not necessarily occur at the “knee” of the S-curve. For a narrow bump more neurons spread out along the lower branch and move closer to the lower knee, presumably because there is less lateral inhibition. Many of these neurons on the lower branch in panel e are in the range of conditional bistability and might be available for transition to the upper branch to expand the bump. In contrast, the lower branch neurons in panel d are not in the range of conditional bistability and presumably are not available to expand the bump with a transition to the upper branch.

3.3 Background activity as a parameter

The strength of the uniform and steady background input I_o plays a strong role in determining the stable patterns observed in the C–W model. This is also the case in the Amari model (Amari 1977). For $I_o = 0.57$, which is above the C–W standard level of $I_o = 0.45$, we noticed that a stable pattern of minimal width (a single neuron on the upper branch state) was possible (Fig. 3a). This single-neuron bump is not possible at $I_o = 0.45$. At this elevated $I_o = 0.57$, while a very narrow bump is stable, the ability to hold a representation of an intermediate stimulus is lost, and such stimuli eventually evolve to a wide bump (not shown). The spatially uniform solution is still stable in this parameter regime.

At a somewhat higher I_o value, the uniform state loses stability (an analytic estimate of the critical value of I_o is derived in the appendix, $I_o = 0.606$). The ability to evoke a very narrow bump with a standard amplitude stimulus is also lost, but we did not determine precisely the critical I_o for this loss. With $I_o = 0.68$, the same stimulus that leads to a single-neuron elevated state for $I_o = 0.57$ now evokes a maximum width bump (Fig. 3a, dashes). Once the uniform stable state becomes unstable, we find evidence of a stable low firing rate bump state that exists entirely on the lower branch of the S-curve. That is, in a narrow I_o range beyond where destabilization of the uniform steady state occurs, a subthreshold stimulus breaks the symmetry of the uniform steady state and results in a stable lower branch bump as shown in Fig. 3b,c. This low-activity bump is reminiscent of a Turing-type pattern as seen in spontaneous pattern generation in reaction-diffusion models of morphogenesis (Murray 1989) and Wilson–Cowan models describing hallucinations and migraines (Ermentrout and Cowan 1979).

The effects of varying I_o over a large range are shown in more detail in Fig. 4. Each panel in Fig. 4 shows properties of the network’s stable long time response (after transients) versus I_o . The stable solution profile after a slight narrow ($p = 1000$, amplitude = 0.01) perturbation from a uniform initial condition is shown in Fig. 4a. Note that this is *not a spatiotemporal solution in response to a stimulus* as seen before, but rather the stable solution shape after a small perturbation as a function of the parameter I_o . As I_o increases from zero, we find several regimes: first there is a range of uniform low-activity solutions. Then we find a small I_o range of lower branch bump solutions (Turing-like patterns described above). Next we see a large I_o range over which the system evolves to a bump. Finally there is a spatially uniform high-activity state when inhibition is overcome by I_o . Both the amplitude and the width of the bump grow with I_o .

Some aspects of the network’s multistability are masked by using the same cue profile for all of the simulations in Fig. 4a. Therefore, we carried out additional simulations, adopting a continuation procedure. We varied I_o in small steps across a range, solving the model for each I_o and using the steady state solution at the previous value of I_o plus a small symmetry breaking perturbation (amplitude = 0.001, $p = 1$) as the initial condition for the next step. The results of forward (solid) and reverse (dots) continuation calculations, shown in Fig. 4b, reveal regions of bistability. The sigmoid portion of the curve corresponds to the spatially uniform solution; it can be determined analytically by evaluating I_o as a function of r . The range over which the uniform network is unstable is approximately $0.6 \leq I_o \leq 4.9$ (see appendix) and is shown dashed. The two asterisks in Fig. 4b bound the I_o range over which r is on the middle branch of $f(r)$. Because this state is unstable for an isolated cell, the uniform state here is unstable for the network. Outside of the asterisks on this uniform state curve, all cells are on the upper or all cells are on the lower stable branches of the S-curve. If the cells were not connected, such upper and lower steady states would be stable. With the cells connected in the network, the corresponding uniform states are stable to *uniform* perturbations. However, they are not necessarily stable to symmetry breaking perturbations (see appendix). The effects of network connectivity can destabilize the uniform state for I_o values outside of the range where an isolated cell would be unstable (between the asterisks). Because of the Mexican hat connectivity kernel, cells are pushed into more or less activity, which prevents them from coexisting stably at the same firing rate.

The network connectivity and intrinsic bistability impose significant hysteresis in the network response: when explored using continuation, the network is able to maintain a bump solution both below the lower destabilization point (around 0.6) and above the upper destabilization point (around 4.49) in I_o . In the range just above the lower destabilization point, small-amplitude stable bumps are seen (they are more obvious in panel c, which is an expansion of panel b); these are consistent with the Turing mechanism of destabilization (Ermentrout and Cowan 1979; Murray 1989; Bressloff and Coombes 1998). The I_o

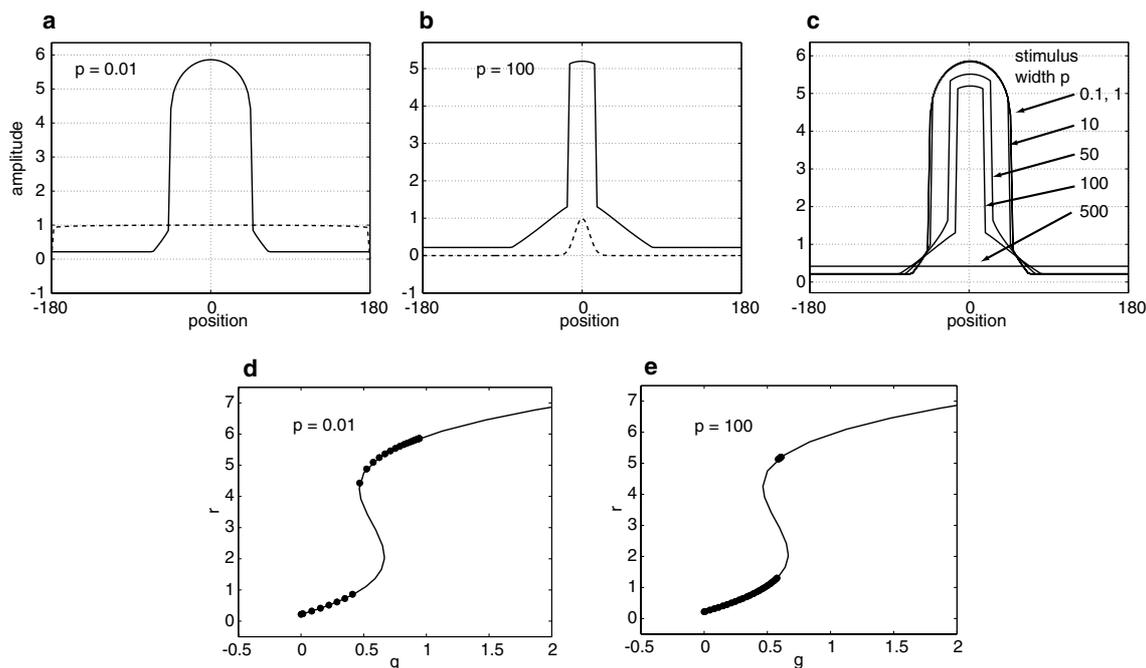


Fig. 2. Multiple bump states exist for the C–W model. Overlaid plots showing stimulus (*dashed*) imposed for 0.5 s starting at $t = 0.5$ s and the firing rate (*solid*) as a function of space at $t = 5$ s. The background input level is $I_o = 0.45$. Note that a *smaller* stimulus exponent results in a *wider* stimulus. **a** Stimulus exponent $p = 0.01$. Stimulus (*dashed*) is wider than standard conditions, but stable bump (*solid*) is the same as in the baseline case. The baseline stimulus exponent $p = 1$ already produces the maximum width stable bump, and broader stimuli do

not increase the bump width. **b** Stimulus exponent $p = 100$. Stimulus (*dash*) is narrower than standard conditions, and stable bump (*dash*) is narrower than baseline. **c** Summary plot showing stable bump pattern for a variety of stimulus exponents. Note that, for this input level ($I_o = 0.45$), there is a minimum sustainable input pattern. **d**, **e** Steady state distributions of neurons on the r – g plane for $p = 0.01$ and $p = 100$

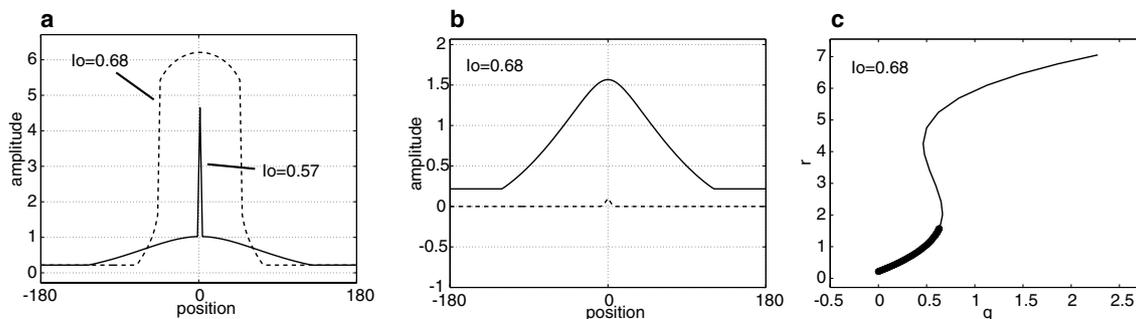


Fig. 3. Examples of the effect of tuning background firing rate, I_o , on network activity. **a** Narrow stimulus ($p = 10000$, amplitude = 1) with sufficient background activity ($I_o = 0.57$) leads to a single neuron sitting stably on the upper branch (*solid curve*). When the background activity level is increased to $I_o = 0.68$, the single-neuron upper branch state becomes unstable, and the elevated network activity induced by the single elevated neuron is enough to evolve a full bump (*dashed*).

At this background activity, the network is no longer conditionally bistable. **b** A subthreshold stimulus ($p = 1000$, amplitude = 0.1, *dashed*) with $I_o = 0.68$ as in **a** results in a stable lower branch bump. Note different vertical scale than panels **a** and **c**. **c** Neurons in **b** spread out below the knee of the lower branch solution. Shown are neurons (*circles*) in the g – r plane as in Fig. 1d–g. The *solid curve* is the g versus r steady state curve as before

range over which the spatially uniform solution in our typical 128-cell network is unstable (according to our numerical simulations of the responses to small perturbations) agrees well with our analytic results obtained by treating the network as a continuum (see appendix).

The remaining panels in Fig. 4 demonstrate more clearly the multiple solutions (i.e., more than two) available from the network in response to large stimuli. These panels show the results of presenting wide and narrow stimuli (of amplitude 1) to the network for a range of I_o . Panel d

shows the width of the bump solution for stimulus widths $p = 10000$ (solid) and $p = 0.001$ (dotted) over the range of I_o shown before. The width is expressed as the number of units on the upper branch (out of a total of 128 in these simulations). Panel e shows the width of bump solutions for *stimulus* widths similar to those shown in Fig. 2c over an interesting range of I_o . This panel shows that, particularly in the I_o range of 0.3–0.5, many stable bump width states are possible and these depend on the width of the stimulus. We note that there are similar multiple patterns

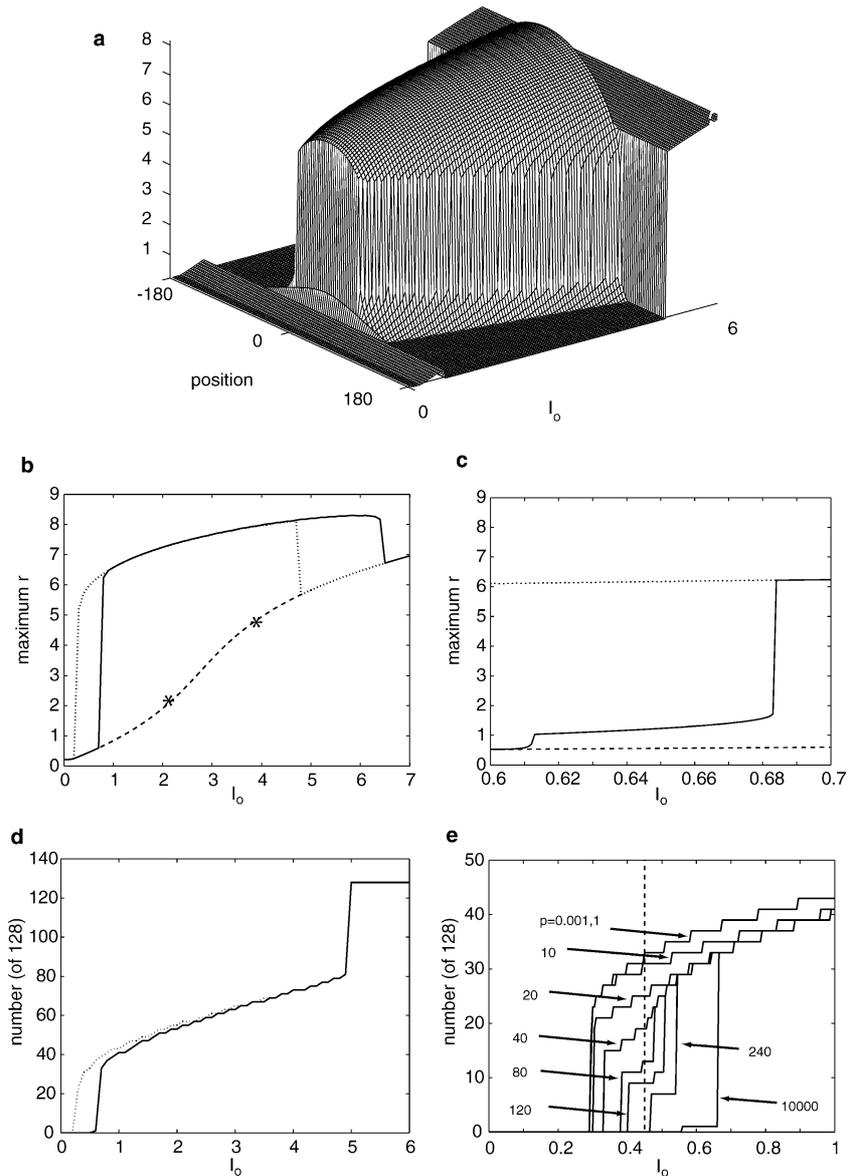


Fig. 4. Summary of the effects of background firing rate I_o on network activity. **a** Shape of stable bump resulting from a slight perturbation of uniform state ($p = 1000$, amplitude = 0.01) as a function of I_o . **b** The sigmoidlike curve represents the uniform steady firing rate of the network as a function of I_o and was calculated analytically. The uniform state is unstable where the curve is dashed. The asterisks represent the limits (in r) of the middle branch of the $f(r)$ curve, between which single units would be unstable if not in the network. The solid curve shows the maximal firing rate (across neurons) for each I_o shown in **a**. The solid curve is the maximum r for the forward continuation calculation (increasing I_o), and the dotted curve is the maximum r for the reverse continuation calculation (decreasing I_o); see text. Note that in this and following panels, the vertical transitions are not actually solutions; however, we have retained the lines

at high I_o corresponding to the upper range of bistability seen in panel **b**, but these were not explored in detail.

4 Discussion

Because of the recent growing interest in cellular bistability (Camperi and Wang 1998; Egorov et al. 2002;

for ease of viewing the range of the various regimes. **c** Expanded section of panel **b** showing lower branch bump. **d** Width of the bump (number of neurons in the high-activity state) as a function of I_o and two stimulus parameter widths, p . The amplitude, I_{cue} , is 1 for both cases. $p = 10000$ (solid) corresponds to a very narrow (single-unit) stimulus, while $p = 0.001$ (dashes) is a very wide stimulus. **e** An expanded view of bump width as a function of I_o and stimulus width. Stimulus width here is explored over the range of p for multistability, roughly corresponding to the plots in Fig. 2C. I_o is shown over a range that includes $I_o = 0.45$, as used in the original C-W simulations (vertical dashed line). As shown in the other panels, I_o greater than approximately 0.7 results in a bump for any stimulus. Note the range of I_o , corresponding to approximately 0.3–0.6, is where the network admits numerous solutions that depend on stimulus width

Loewenstein and Sompolinsky 2003) as related to persistent activity patterns, we have revisited the Camperi–Wang model in more detail. We found that the background synaptic activity I_o must be carefully tuned in order to exhibit bistability. This further constrains the already restricted parameter regime for working-memory-like behavior found by Camperi and Wang (1998). In an

adjacent and larger parameter region, the model supports bump states robustly, but the working-memory-like network bistability is lost. The globally attracting bump is still translation invariant, and thus a memory state for orientation, but the uniform rest state is no longer stable. In the working-memory regime we found an abundance of bump states that could encode a two-parameter continuum encoding stimulus width and location coexistent with a stable uniform state. In an adjacent, but still small, higher- I_o range, we found bistability between a large-amplitude bump state and a small-amplitude bump state. The small-amplitude bump results from a Turing-like destabilization of the uniform “rest state.” All of these states are translation invariant on the ring.

Intrinsic or cellular conditional bistability provides robustness against noise in bump location to this class of models for persistent activity (Camperi and Wang 1998). This robustness presumably occurs whether the model is tuned to the parameter regime where the bump is a global attractor or the bump coexists with a low-activity rest state. In either case, the intrinsic bistability underlies a sharp transition between neurons in the high-activity and low-activity states (upper and lower branches of the S-curve). This gap in activity levels serves as a buffer against the influence of noise or distractors (Camperi and Wang 1998). We would expect to find that graded multistability of the type explored by Egorov et al. (2002) is not robust against noise.

We have shown that, in the working-memory regime of the C–W model, there is more flexibility in representing stimulus features. One might expect increased encoding power in the presence of cellular bistability. Indeed, an ensemble of unconnected bistable neurons can maintain an arbitrary binary pattern. When such neurons are connected in a network, the connectivity serves to limit the available response patterns, but intrinsic bistability still imparts additional encoding capacity. The localized coupling favors those patterns with local similarity and only a few up/down transitions. We have seen that, in this enhanced encoding regime, narrow stimuli result in a stable narrow bump, and there is a continuous range of bump widths up to a maximum width determined by the connectivity kernel. In this regime, the C–W model can encode a two-parameter family of stimuli, with both mean and spread in orientation (in a visuospatial context).

These bump models offer insightful idealizations and allow one to explore parameter dependencies. To achieve the network bistability that would be required in a working memory task, care must be taken in the tuning of parameters for this model as well as in Amari-like models that do not incorporate intrinsic bistability. One might look to expand the parameter range for bistability by manipulating the intrinsic dynamics (the cubic function $f(r)$) or the connectivity kernel; however, Camperi and Wang (1998) found only a small region where changes in these parameters supported bistability. (We did explore somewhat the effect of varying the shape of the connectivity kernel in combination with varied I_o . Because of the interplay between intrinsic bistability and recurrence, the results are not easily generalizable and so we have deferred addressing

this pending a more detailed study. However, whenever the excitatory region of the connectivity kernel was sufficiently larger than the distance between cells and sufficiently smaller than the circumference of the ring, we saw no evidence suggesting that our conclusions needed revising.) The incorporation of idealized intrinsic bistability, leading to the S-shaped input-output function, is biophysically reasonable; for a similar approach in another context see Gruber et al. (2003). The simple summation of synaptic input combined with a threshold that prevents inhibition from contributing a negative influence to dr/dt or from forcing r below its basal level is less clearly justified.

In considering the C–W model, we have not addressed the issue of erasing the memory that is represented by the bump state. In the regime where a stable uniform rest state and bump coexist, the bump can be eliminated with transient global inhibition (Camperi and Wang 1998). We found that the rest state in this regime could be restored with sufficiently strong global or even shaped excitation. This is also the case for some other models of working memory (Laing and Chow 2001; Tegner et al. 2002). Many single-unit persistent recordings from cortex seem to demonstrate a burst of increased activity at the end of the delay period (Fuster 1988), which is consistent with such a mechanism for terminating working-memory patterns. In the higher- I_o regime where the bump state is a global attractor, a transient stimulus cannot permanently erase the bump because the rest state is unstable. In simulations, we have seen that a strong uniform stimulus can transiently leave the system in a uniform state. Since any fluctuation will drive the system back to the bump, this state cannot be maintained. Moreover such a massive and uniform excitatory transient is most likely not physiological.

These models are elegantly idealized, but we should keep in mind what they are predicting. It appears that the network bistability that we see in these models is merely a transition regime between a globally attracting down state and a globally attracting bump state. Although we did not explore this in detail, the C–W model also shows a regime of network multistability at yet higher I_o , where the transition is made from globally attracting bump behavior to globally attracting uniform high activity. These properties lead us to question whether the conceptual framework of network bistability (bump and rest states as coexistent attractors) shown in the Amari class of models represents the essence of working memory. What do we expect of working memory robustness as background network activity increases? It seems reasonable that working memory function might require some level of background activity. But should a working memory system lose the ability to erase an on-line memory or to exchange one for another as background levels increase? The transition from bistability to one bump state attractor as background activity is increased is a property of Amari class models. Perhaps the globally attracting bump state regime has some meaning for working memory if the dynamic effects of neuromodulators are included as other state variables. If so, we must explore on what time scale neuromodulators relating to attentional or reward mechanisms might be effective in restoring the system (or its parameters) to the

bistable regime or back to a uniform rest state. These considerations and questions apply to the whole class of such Amari-like models, whether or not intrinsic bistability is involved.

A prediction for *in vitro* experiments comes from the Amari framework of bump generation based on an upright Mexican-hat-like kernel: If such connectivity applies to a slice of generic cortex, then one might expect to see some evidence of working-memory behavior or persistent activity when parameters are tuned to appropriate regimes. (We note that, when invoking the Mexican hat framework, one usually envisions rapidly equilibrating synapses and long-range inhibition. However, various combinations of connectivity and relative time scales between interacting populations of excitatory and inhibitory cells can reduce to a Mexican-hat-type kernel for a single-population model.) Of course one cannot apply physiologically realistic stimuli to the slice in order to generate meaningful held or stored patterns. On the other hand, if modulatory agents are bath applied or excitability is otherwise increased (mimicking an increased I_o), the C–W model predicts spontaneous bumplike pattern formation. Amari-class models predict multiple bumps if the tissue size is large relative to synaptic footprint (Amari 1977). Once a bump has been produced, decreasing I_o should bring the system back to a bistable regime and the pattern could be erased by a globally applied (“inhibitory-like”) transient stimulus. The feasibility of such experiments will depend on many factors; of utmost necessity is adequate connectivity in the slice.

To date no evidence for spatially coherent bumplike activity *in vitro* has been reported, but only recently have techniques for capturing ensemble activity been refined (Grinvald et al. 1982). Interestingly, Turing destabilization and spontaneous pattern formation were first induced in distributed model chemical systems 50 years after Turing’s theory was published (Ouyang and Swinney 1991; Pearson 1993). Moreover, we cannot say beforehand whether such activity patterns would be spatially organized in a bump. It is more likely that bumps represent neurons adjacent in feature space that might not be adjacent in cortex. Thus, single-cell sensitivity in such ensemble measurements may be important (Peterlin et al. 2000). Some evidence for snippets of stored memories have been reported in slice/culture systems (Ikegaya et al. 2004), and spontaneous activity can also be used to understand cortical connectivity (Arieli et al. 1995).

Appendix

Stability of the uniform stable state

Here we determine the stability of the uniform steady state of the C–W model in its continuum limit:

$$\begin{aligned} \tau \frac{\partial r}{\partial t} &= -f(r) + g(I) \\ g(I) &= I, I \geq 0 \text{ and } g(I) = 0, \quad I < 0 \\ I &= I_o + \frac{1}{2\pi} \int_0^{2\pi} W(\theta - \tilde{\theta}) r(\tilde{\theta}, t) d\tilde{\theta} = I_o + W * r. \end{aligned}$$

Uniform steady states $r(\theta, t) = R$ are obtained by setting $dr/dt = 0$ in the above equations and solving $f(R) = g(I_o + W * R)$. The mean connectivity \hat{W} is given by $\hat{W} = \frac{1}{2\pi} \int_0^{2\pi} W(\theta) d\theta$, and thus, for uniform R , the convolution $W * r = \hat{W} R$. For the parameters and connectivity function used here (see methods), $\hat{W} = W_E/2 - W_I = -0.7$. Because of the threshold on the input function $g(I_o + \hat{W} R)$, there are two regimes of I_o to consider: $I_o \leq I_{\text{crit}}$ and $I_o > I_{\text{crit}}$, where $I_{\text{crit}} + \hat{W} R = 0$.

For $I_o \leq I_{\text{crit}}$, $I_o + \hat{W} R \leq 0$ and $g(I_o + \hat{W} R) = 0$, i.e., there is no input contributed by the network. In this regime of I_o , neurons in the network behave as they would in isolation, and the firing rate of the neurons is given by the real zero of $f(r)$, $r_o \approx 0.216486$ for standard C–W parameters. The uniform network continues to fire at r_o until we increase the background synaptic activity parameter I_o beyond $I_{\text{crit}} = -\hat{W} r_o \approx 0.15154$.

For $I_o > I_{\text{crit}}$, $I_o + \hat{W} R > 0$ and $g(I_o + \hat{W} R) = I_o + \hat{W} R$. In this regime, $I_o = f(R) - \hat{W} R$ gives the relationship between I_o and the firing rate at the uniform stable state R . The inverse relation gives R as a function of I_o , which is shown as a dashed line in Fig. 4b. Note that at standard C–W parameters there is a unique R for every I_o .

To find the stability of a uniform steady state $r(\theta, t) = R$, we linearize the C–W system around the uniform stable state (Ermentrout and Cowan 1979; Murray 1989; Zhang 1996; Bressloff and Coombes 1998):

$$r(\theta, t) = R + \eta(\theta, t),$$

$$\tau \frac{\partial}{\partial t} (R + \eta) = -f(R + \eta) + g(I_o + W * (R + \eta)),$$

where η is a small perturbation to the uniform steady state. Expanding the functions f and g in a Taylor series around $\eta = 0$ and using the facts that R is constant for the uniform steady state and $f(R) = g(I_o + \hat{W} R)$, we obtain

$$\begin{aligned} \tau \frac{\partial \eta}{\partial t} &= -f(R) - f'(R)\eta + g(I_o + \hat{W} R + W * \eta) + \dots \\ &= -f(R) - f'(R)\eta + g(I_o + \hat{W} R) \\ &\quad + g'(I_o + \hat{W} R) (W * \eta) + \dots \\ &= -f'(R)\eta + g'(I_o + \hat{W} R) (W * \eta) + \dots \end{aligned}$$

The stability calculation is different for the two ranges of I_o corresponding to the domain where $g(I) = 0$ and where $g(I) = I$. For $I_o \leq I_{\text{crit}}$, $I_o + R\hat{W} \leq 0$, $g(I) = 0$, $g'(I) = 0$, and the dynamics are determined only by $f(R)$:

$$\tau \frac{\partial \eta}{\partial t} = -f'(R)\eta.$$

That is, R is stable if $f'(R) > 0$, which is true for all values of R to the left of the left knee of $f(R)$ for the parameters used here.

For $g(I) = I$, $g'(I) = 1$ and

$$\tau \frac{\partial \eta}{\partial t} = -f'(R)\eta + (W * \eta), \quad I_o \geq I_{\text{crit}}.$$

Expanding η in a Fourier series $\eta(\theta, t) = \sum_k \eta_k(t) e^{ik\theta}$, we find $\tau \frac{\partial \eta_k}{\partial t} = -f'(R)\eta_k + \lambda_k \eta_k$, where λ_k are the eigenvalues

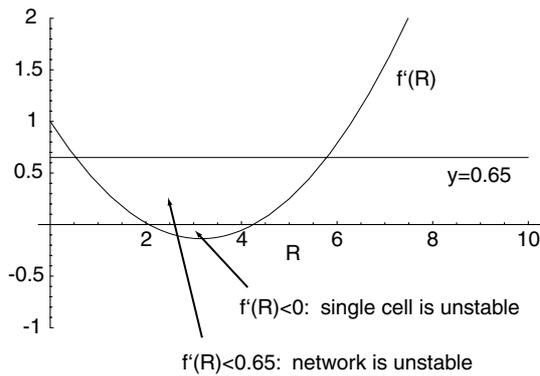


Fig. 5. Analytical determination of uniform stable state. Overlaid plots of $f'(R)$ and $y=0.65$. The uniform solution is stable when $f'(R) > \text{Re}[\lambda_1] = 0.65$

of the linear operator W^* corresponding to the eigenfunctions $e^{ik\theta}$. Thus, the uniform steady state is stable if

$$\text{Re} \left[\frac{-f'(R) + \lambda_k}{\tau} \right] < 0 \text{ or } \text{Re}[\lambda_k] < f'(R), \forall k.$$

The eigenvalues of the linear operator W^* are the Fourier coefficients of W , $\lambda_k = \frac{1}{2\pi} \int_0^{2\pi} W(|z|) e^{ikz} dz$. Because we have taken W to be a modulated and shifted cosine function, only the modes $k = -1, 0, 1$ have nonzero real parts. With the default parameters, the corresponding eigenvalues are $\text{Re}[\lambda_0] = -0.7$, and $\text{Re}[\lambda_{\pm 1}] = 0.65$.

As seen in Fig. 5, $f'(R)$ intersects $\text{Re}[\lambda_1] = 0.65$ at $R \approx (0.531, 5.785)$. Thus only the uniform steady states with values of R for which $f'(R) > 0.65$ are stable. These correspond to values of I_o in the interval $I_o \approx (0.606, 4.944)$. Note that *individual neurons* are unstable only for $f'(R) < 0$. These values accord well with our simulations.

References

Amari S (1977) Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol Cybern* 27(2):77–87

Arieli A, Shoham D, Hildesheim R, Grinvald A (1995) Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex. *J Neurophysiol* 73(5):2072–2093

Ben-Yishai R, Bar-Or RL, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. *Proc Natl Acad Sci USA* 92(9):3844–3848

Booth V, Rinzel J (1995) A minimal, compartmental model for a dendritic origin of bistability of motoneuron firing patterns. *J Comput Neurosci* 2(4):299–312

Bressloff PC, Coombes S (1998) Spike train dynamics underlying pattern formation in integrate-and-fire oscillator networks. *Phys Rev Lett* 81(11):2384–2387

Camperi M, Wang XJ (1998) A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *J Comput Neurosci* 5(4):383–405

Durstewitz D, Seamans JK, Sejnowski TJ (2000) Neurocomputational models of working memory. *Nat Neurosci* 3(Suppl):1184–1191

Egorov AV, Hamam BN, Fransén E, Hasselmo ME, Alonso AA (2002) Graded persistent activity in entorhinal cortex neurons. *Nature* 420(6912):173–178

Ermentrout GB, Cowan JD (1979) A mathematical theory of visual hallucination patterns. *Biol Cybern* 34(3):137–150

Fuster JM (1988) *The prefrontal cortex*. Raven Press, New York

Goldman-Rakic PS (1995) Cellular basis of working memory. *Neuron* 14:477–485

Grinvald A, Mankner A, Segal M (1982) Visualization of the spread of electrical activity in rat hippocampal slices by voltage-sensitive optical probes. *J Physiol Lond* 333:269–291

Gruber AJ, Solla SA, Surmeier DJ, Houk JC (2003) Modulation of striatal single units by expected reward: a spiny neuron model displaying dopamine-induced bistability. *J Neurophysiol* 90(2):1095–1114. Epub 2003 Mar 20

Ikegaya Y, Aaron G, Cossart R, Aronov D, Lampl I, Ferster D, Yuste R (2004) Synfire chains and cortical songs: temporal modules of cortical activity. *Science* 304(5670):559–564

Laing CR, Chow CC (2001) Stationary bumps in networks of spiking neurons. *Neural Comput* 13(7):1473–1494

Loewenstein Y, Sompolinsky H (2003) Temporal integration by calcium dynamics in a model neuron. *Nat Neurosci* 6(9):961–967

Murray JD (1989) *Mathematical biology*. Springer, Berlin Heidelberg New York

Ouyang Q, Swinney HL (1991) Transition from a uniform state to hexagonal and striped Turing patterns. *Nature* 352:610–612

Pearson JE (1993) Complex patterns in a simple system. *Science* 261(5118):189–192

Peterlin ZA, Kozloski J, Mao BQ, Tsiola A, Yuste R (2000) Optical probing of neuronal circuits with calcium indicators. *Proc Natl Acad Sci USA* 97(7):3619–3624

Tegner J, Compte A, Wang XJ (2002) The dynamical stability of reverberatory neural circuits. *Biol Cybern* 87(5–6):471–481

Wang XJ (2001) Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci* 24(8):455–463

Wilson HR, Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J* 12(1):1–24

Wilson HR, Cowan JD (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13(2):55–80

Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci* 16(6):2112–2126