Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks

Alfonso Renart, Pengcheng Song and Xiao-Jing Wang

Volen Center for Complex Systems

Brandeis University

Waltham, MA 02454

Neuron, in press

Running Title: Homeostatic regulation of spatial working memory.

Correspondence: Xiao-Jing Wang at above address. Phone: 781-736-3147, fax: 781-736-2915, email: xjwang@brandeis.edu

Abstract

The concept of bell-shaped persistent neural activity represents a cornerstone of the theory for the internal representation of analog quantities, such as spatial location or head direction. Previous models, however, relied on the unrealistic assumption of network homogeneity. We investigate this issue in a network model where fine-tuning of parameters is destroyed by heterogeneities in cellular and synaptic properties. Heterogeneities result in the loss of stored spatial information in a few seconds. Accurate encoding is recovered when a homeostatic mechanism scales the excitatory synapses to each cell to compensate for the heterogeneity in cellular excitability and synaptic inputs. Moreover, the more realistic model produces a wide diversity of tuning curves, as commonly observed in recordings from prefrontal neurons. We conclude that recurrent attractor networks in conjunction with appropriate homeostatic mechanisms provide a robust, biologically plausible theoretical framework for understanding the neural circuit basis of spatial working memory.

Introduction

Neural correlates of the internal representation and memory of spatial information (such as an angle or a location) have been observed in many brain regions. Examples include mnemonic persistent activity in the prefrontal and parietal cortices associated with working memory (Gnadt and Andersen 1988; Funahashi et al. 1989; Chafee and Goldman-Rakic 1998; Rainer et al. 1998), head-direction selective activity in cortical and sub-cortical structures (Rank Jr. 1985; Taube 1995; Sharp et al. 2001) and place cell activity in the hippocampus (O'Keefe and Dostrovsky 1971; Wilson and McNaughton 1993; McNaughton et al. 1996). Neurons in these areas encode and store spatial information, even when explicit sensory stimuli are absent. For instance, if a monkey must remember the spatial location of a transient stimulus for a few seconds in order to perform a delayed behavioral response, single neurons in the dorsolateral prefrontal cortex display location-specific activity that persists during a delay period after the stimulus has been withdrawn (Goldman-Rakic 1992). The spatial selectivity of a cell's memory activity can be quantified in terms of a broad bell-shaped tuning curve with the maximum at the cell's preferred location. Given this type of encoding, if neurons of the network were labeled and arranged in space according to their preferred locations, the activity pattern during the delay period would be localized, with the memory of the stimulus location being encoded in the peak location of the network activity profile. Such bell-shaped firing patterns ('bump' states) provide a population representation of the remembered spatial cue (Figure 1B), which can be used to guide the behavioral response. Working memory function arises from the ability of the network to sustain localized persistent firing patterns by virtue of its internal reverberatory dynamics during the delay period when the specific external input is no longer present (Wang 2001). The bump states can, in this sense, be interpreted as dynamical attractors of the working memory network.

A theoretical framework has been developed for understanding the properties of neural activity recorded during the maintenance of spatial information. This framework assumes that the localized neural firing patterns are self-sustained by synaptic interactions in a local recurrent network (Ben-Yishai et al. 1995; Camperi and Wang 1998; Compte et al. 2000; Zhang 1996; Samsonovich and McNaughton 1997). A crucial feature of these models is the assumption that the local network is homogeneous, with the properties of all the cells in the network being identical. As

a result, any localized activity pattern displaced spatially leads to another activity pattern of the same shape but peaked at a different location. For large networks, this ensures that the continuous nature of spatial information is preserved by a continuous family of bump attractors, each encoding a different stimulus location. This paradigm has been successful in describing the phenomenology of spatial working memory experiments (Camperi and Wang 1998; Compte et al. 2000), place cell activity (Tsodyks and Sejnowski 1997; Samsonovich and McNaughton 1997) and head-direction related activity (Skaggs et al. 1997; Zhang 1996; Redish et al. 1996; Xie et al. 2002), and it has also been postulated to play an important role in the optimal read-out of noisy population codes (Deneve et al. 1999). Furthermore, behavioral results in monkeys (White et al. 1994) and humans (Ploner et al. 1998) show that the accuracy of saccades to remembered target locations decreases as the memory period increases in a way consistent with the predictions of a model with a continuum family of bump attractors (Compte et al. 2000).

The homogeneity assumption is, however, biologically unrealistic. Indeed, in a given neural population, the membrane properties vary considerably from cell to cell (see, e.g., Mason and Larkman (1990)) and the number and efficacy of synaptic connections are also variable (Braitenberg and Schütz 1991). The effect of heterogeneities in these models has not been systematically characterized, but previous studies suggest that even with a small amount of heterogeneity most of the continuum of bump attractors disappear; only a few (2 or 3) discrete spatially-localized attractor states survive (Zhang 1996; Tsodyks and Sejnowski 1997; Stringer et al. 2002). In the context of a spatial working memory task, if a typical cue-specific localized pattern of activity is not stable and instead drifts systematically toward one of the few stable positions, the memory of the initial cue would be lost, and the network would lose its functional relevance. It is therefore unclear whether this paradigm provides a satisfactory model for the internal representation of spatial information in the brain, when realistic levels of neural heterogeneities are taken into account.

In this study we use a biologically plausible network model to characterize in detail the effect of heterogeneity on the stability of persistent bump states and show that, in the absence of 'fine-tuning' in the cellular properties of different cells, the network does not support spatial working memory function. This fine-tuning problem is a general feature of systems encoding internal representations of analog

features (Seung 1996; Wang 2001). A possible solution to the problem is that the network learns to tune itself through an activity-dependent mechanism. Several homeostatic mechanisms of this kind, thought to promote stability in neural circuits, have been uncovered (Marder 1998; Turrigiano 1999). We incorporated into our model an activity-dependent scaling of synaptic weights (Turrigiano et al. 1998; O'Brien et al. 1998; Desai et al. 2002) which up- or down- regulates excitatory inputs so that the long-term average firing rate is similar for each neuron. We found that synaptic scaling effectively homogenizes the network despite considerable heterogeneity in single cell and synaptic properties, including parameters regulating the scaling process itself. As a result, robust working memory function is recovered under biologically plausible conditions.

Results

We investigated the active maintenance of spatial information in the presence of heterogeneities using a recurrent cortical network model of spatial working memory. The model (see Methods) consists of two reciprocally connected cell populations, pyramidal cells (labeled by their preferred cues θ from 0 to 360 degrees on a circle) and interneurons (see schematic Figure 1A). In most simulations, the excitatory recurrent connections were all-to-all, with a magnitude that decayed as a function of the difference between the cells preferred cues. In some simulations (see below) the pyramidal cells were connected probabilistically and with random synaptic efficacies. In these simulations, the probability of connection decreased as the difference between the cells preferred cues increased. Without heterogeneity, the network model reproduces the salient electrophysiological observations from the behaving monkey in an oculomotor delayed-response (ODR) task (Compte et al. 2000). A transient spatial input triggers a localized activity pattern in the network (Figure 1B). This 'bump state' outlasts the stimulus, resulting in a persistent activity state sustained by excitatory synaptic reverberations (Wang 2001). The ideal memory function of the network is realized when all cells are identical. In this case, when different stimuli are used in different trials, in any given trial the peak location of the persistent firing pattern (as measured with the population vector method (Georgopoulos et al. 1982)) remains close to the stimulus location throughout the memory period, so that the stimulus location can be accurately read out several seconds after the stimulus

offset (Figure 1C).

[Figure 1 about here.]

Effect of Heterogeneity

This network behavior is extremely sensitive to heterogeneities. When the leak potential V_L (membrane resting potential) varies from cell to cell according to a Gaussian distribution (mean $< V_L >= -70$ mV and standard deviation SD(V_L) $= 1$ mV) the bump state elicited by a stimulus at any location systematically drifts toward one of two privileged positions in the network (Figure 2A top left). In this condition, after a delay period of five seconds, an estimate of the position of the cue from the peak location of the bump would lead to the incorrect conclusion that 19 out of the 20 stimuli were shown at one of the two privileged locations. We found that the drift speed $v(\theta)$ of a bump state is unique at each spatial location θ. This means that knowing only the peak location θ of the bump state at any given time is enough to know the speed $v(\theta)$ at which the bump is moving through that location. No extra information about, for instance, the location of the stimulus at the beginning of the trial, or the time since the trial started, is needed. Interestingly, the drift speed is determined by the distribution of local average excitability $U(\theta)$ (shown in Figure 2A top right) across the network. $U(\theta)$ measures the average excitability of a local region of the network around θ with a size of the order of the width of the localized activity profile (see Methods). Specifically, we observed that $v(\theta) = k\ dU(\theta)/d\theta$, where k is a constant (see Methods). We used this finding to build a phenomenological model of the dynamics of the bump state in the presence of heterogeneity, described by the dynamical equation

$$d\theta/dt = k\ dU(\theta)/d\theta + \eta(t) \quad , \tag{1}$$

where $\eta(t)$ is a white noise (see Methods). This simple model accurately describes the time evolution of the bump state's peak location $\theta(t)$ obtained in the simulations of the original spiking network model (Figure 2A bottom).

This formulation allows us to draw an analogy between the drift of a bump state's peak location, and the motion of a particle with speed $v(\theta)$ at each location θ. The function $-U(\theta)$ can, accordingly, be interpreted as an 'energy function'. The drift speed is determined by the gradient $dU(\theta)/d\theta$, and the privileged locations with high

local excitability are given by the maxima of $U(\theta)$, at which $v(\theta) = 0$. Intuitively, due to the random distribution of single cell parameters across the network, some clusters of cells become slightly more excitable than others. The regions where this average local excitability is high constitute the privileged locations in the network, and the difference in excitability between nearby locations determines how fast the bump moves through them. This explains how the presence of heterogeneity results in the systematic drift of the bump states.

One could argue that with a sufficiently large number of synaptic inputs, heterogeneity could be averaged out in each local network region, so that the drift speed should decrease as the number of inputs increases. We studied the effect of increasing the number of inputs by increasing the size of the network (see Methods). We quantified the overall effect of heterogeneity by the average across the network of the absolute value of the drift speed $\bar{v} =< |v(\theta)| >_\theta$. In Figure 2B, we show that, for a fixed level of heterogeneity (the same used in Figure 2A), the mean drift speed $\bar{v}$ indeed decreases with the network size N as $\sim 1/\sqrt{N}$ (Figure 2B inset). This type of scaling with N is a general feature of heterogeneities in single-neuron parameters and can be obtained analytically (details of the calculation can be obtained from the authors upon request). In these simulations connectivity was all-to-all and homogeneous across cells. In Zhang (1996), heterogeneities in the synaptic efficacies in a fully connected network were considered and their effect on the mean drift of the bump states was shown to decrease as $\sim 1/N$. Intuitively, the extra factor $1/\sqrt{N}$ comes from averaging the N independent fluctuations associated with each afferent synaptic connection. We also considered sparse connectivity (see Methods), so that each neuron is randomly connected to C other neurons on average. The effect of this type of connectivity can be shown to scale as $1/\sqrt{CN}$. Thus, for large sparse networks one expects heterogeneity in synaptic properties to have a weaker influence than cellular heterogeneities on the stability of the bump states. Since simulating large sparse networks is computationally very costly, we have obtained most of our results in fully connected networks with no synaptic heterogeneities. The final check of our mechanism, however, has been performed in a sparse network with random synaptic efficacies (see Figure 6).

Although the effect of heterogeneity decreases with increasing network size, we found that increasing the level of heterogeneity (by increasing the variance of the heterogeneous parameter, or by taking into account heterogeneities in other single

cell parameters like capacitance C_m, firing threshold V_{th}, etc) at fixed network size results in a dramatic increase of the mean drift speed (Figure 2B). With realistic levels of heterogeneities, the mean drift speed is large even with $N \sim 5,000$ synaptic inputs per cell, a number typical for a cortical pyramidal neuron (Braitenberg and Schütz 1991). At high enough levels of heterogeneity the bump states stop being stable and the network shows a spatially disordered pattern of neural activity (not shown).

[Figure 2 about here.]

Figure 2C shows a typical example of a network of $4,096$ excitatory and $1,024$ inhibitory cells with a 2 mV standard deviation in V_L. In this network (corresponding to a particular realization (star in Figure 2B) of the distribution of heterogeneity marked by the downward triangle in Figure 2B) the peak position of the bump state 6 sec after stimulus offset is plotted across 100 trials for each of 8 equi-spaced locations (see Methods). The spatial distribution of final peak locations is strongly non-uniform. If the final peak location of the bump were used to guide a saccade toward the location of the previously shown stimulus in this case, saccadic responses for several stimuli would never be classified correctly, whereas others would always result in successful saccades. These features are clearly inconsistent with the experimental results from ODR tasks, in which behavioral performances of $\sim 90\%$ correct saccades are observed, regardless of the stimulus location (Funahashi et al. 1989; Chafee and Goldman-Rakic 1998). Given this level of performance, and because in the real biological network every single biophysical parameter is expected to be heterogeneous, we conclude that large numbers of connections *per se* do not solve the heterogeneity problem.

Homogenization by synaptic scaling

We showed that drifts of stimulus-elicited bump states are related to the differences in excitability between nearby network regions. This observation suggests that if some homeostatic mechanism could regulate, and approximately equalize the excitability of all neurons, then the network could be effectively homogenized, and systematic drifts of memory activity patterns would be reduced or eliminated. We investigated the hypothesis that a homeostatic mechanism might lead to an effective

homogenization of the network. Among the several types of known homeostatic processes (Marder 1998; Turrigiano 1999; Desai et al. 1999; Soto-Treviño et al. 2001), we focused in this study on the activity-dependent scaling of synaptic strengths, a mechanism discovered in experiments on neocortical neurons in culture slices and *in vivo*. By means of this mechanism, the distribution of excitatory synaptic contacts to a cell is up- or down-regulated multiplicatively in such a way as to control its overall level of activity over long periods of time (Turrigiano et al. 1998; O'Brien et al. 1998; van Rossum et al. 2000; Desai et al. 2002). We have implemented the dynamics of the scaling process through the following equation

$$\tau_g dg(\theta)/dt = -g(\theta)(r(\theta) - r_{tg}(\theta)) \quad , \tag{2}$$

where τ_g is the characteristic time of the process, $g(\theta)$ is a factor that multiplies all the excitatory synaptic conductances to a pyramidal cell with preferred cue θ, $r(\theta)$ is the instantaneous firing rate of the cell, and $r_{tg}(\theta)$ is the cell's 'target' firing rate, some level of long-term activity that the scaling process tries to impose on each neuron. In our simulations r_{tg} is drawn, for each cell, from a Gaussian distribution with a mean equal to the average excitatory activity when the network (without heterogeneities) is in a bump state, and a standard deviation equal to either zero (r_{tg} identical for all cells) or to some fixed fraction of the mean (see below).

The characteristic time of the scaling process found experimentally is long (hours to days) (Turrigiano et al. 1998). A long characteristic time of homeostatic regulation is necessary, so that different cells can have different levels of activity to encode information at shorter time scales. However, simulations and analysis of the behavior of large networks became infeasible, if the system contains widely different time scales. To bypass this difficulty, we have focused only on the estimation of the steady state value of the scaling factor for each pyramidal cell. In order to do this, we assume that the network is exposed to many spatial stimuli (one after another) over a long period of time. Since what matters for the steady state of the very slow scaling process is the integrated activity of each cell across different stimuli, we have replaced this temporal average by a spatial average carried out over several network simulations run in parallel. In each simulation a bump state is generated by a cue at a given location, and the different locations for the different simulations are evenly distributed (see Methods). By means of this procedure we have been able to use equation (2) with $r(\theta)$ equal to the average of the instantaneous firing rate of each cell across the different simulations run in parallel, and with a value of τ_g similar to

the other time constants of the problem.

The effect of scaling on the working memory function of the network can be assessed by studying the dynamics of the bump states when the scaling factor of each cell is set fixed and equal to its steady state value. After scaling, i.e. when the scaling process has reached its steady state, all excitatory synaptic conductances are multiplied by the steady state scaling factor g_{ss}, which results in excitatory synaptic currents of the form

$$I_{syn} = g_{ss} g_{syn}(V_m - V_{syn}) \quad , \tag{3}$$

where the parameter g_{syn} is the total excitatory conductance due to all the excitatory synaptic inputs to that neuron.

The systematic drift of the bump state after scaling is shown in Figure 3A for the same level of heterogeneity considered in Figure 2A. In all the simulations in Figure 3, the parameter r_{tg} was chosen to be identical for all cells and 40 cues were used to calculate the scaling factor of each neuron, with the location of the bump state at $(360/40)i, i = 1, 2, ..., 40$. This ensures that all neurons have effectively the same average long-term firing rate. After scaling, the systematic drift in the peak location of the bump state is almost completely suppressed despite heterogeneity. Figure 3B shows the reduction in drift speed when heterogeneities in different single cell parameters are considered. Although some parameters have a larger impact than others on the stability of the bump states, the scaling process is similarly effective in homogenizing all sources of heterogeneity that we have studied (Figure 3B). The homeostatic process adjusts the synaptic strength only by a few percents (Figure 3C). The steady state scaling factor of a cell depends, primarily, on the single cell properties of the cell. Indeed, the steady state scaling factor of each neuron can often be predicted with high accuracy from its cellular parameters (e.g. as a linear function of its capacitance or leak potential) (Figure 3C).

The effectiveness of the scaling process lies in its ability to homogenize the input-output relationships of different cells in response to synaptic inputs. In Figure 3D we have plotted the firing rate as a function of the excitatory synaptic input g_{syn} in equation (3) for 50 cells with different single cell parameters before ($g_{ss} = 1$) and after scaling. The range of values of g_{syn} used in the figure corresponds to that experienced by the cells in the network. Before scaling, the firing rate for a fixed value of g_{syn} varies widely from cell to cell due to heterogeneities in V_L, C_m, etc (Figure 3D left). The synaptic scaling process uses the long-term average firing rate

as a feedback signal to adjust g dynamically. After scaling, g reaches an optimal value, g_{ss}, so that the firing rate as function of the synaptic drive, g_{syn}, within the neuron's operational range, becomes virtually the same for all cells, independently of their intrinsic membrane parameters (Figure 3D right). Note that scaling does not affect the response of the cells to an injected current; hence the experimentally measured firing frequency as a function of the injected current intensity (i.e. the single cell's f-I curve) should still display a high degree of heterogeneity.

[Figure 3 about here.]

Because synaptic scaling does not exactly compensate for various kinds of heterogeneous single cell parameters, network homogenization is not perfect. Nonetheless, it is remarkably efficient, as can be seen by the fact that, after scaling, the distribution of the average local excitability $U(\theta)$ becomes essentially flat across the network (Figure 3A right). Although $U(\theta)$ is not exactly uniform, the differences in local excitability across the network become so small that the main determinant of the movement of the bump becomes the noisy input from external background activity, so that, similarly to the case when the network is homogeneous, the peak location of the bump state wanders slowly and quasi randomly, as in a diffusion process (Compte et al. 2000).

Heterogeneity in the long-term firing rate across the network

In the homeostatic process, we assumed a uniform distribution of many (40) cues, so that the long-term activity was roughly the same for all neurons. What happens when there is a biased distribution of spatial stimuli? We considered this question by using a decreasing number of cues for the calculation of the scaling factors in the same network studied in Figures 2A and 3A. Figure 4 shows the working memory behavior of the network after scaling using 20, 8 and 4 cues. Only when just four cues were used the performance of the network was compromised. In this case, during the homeostatic process, neurons with preferred locations close to the 4 used stimuli have higher firing rates, which result in smaller scaling factors. Therefore, after scaling, those cells have a lower excitability, and bump states tend to drift away from them. The opposite is true for neurons whose preferred locations are *in between* the cues used in the homeostatic process: with larger scaling factors they are more excitable and tend to attract the bump states (marked by dashed lines in

Figure 4C right). However, the scaling mechanism works well even with 8 cues. In general, as long as the number of cues is larger than the number of non-overlapping activity profiles that can fit into the network (approximately 4), homogenization by scaling is efficient. Thus, wide network activity profiles (equivalent to wide tuning curves for single cells) reduce the need for homogeneity in the distribution of spatial stimuli encoded by the network.

[Figure 4 about here.]

Heterogeneity in the target firing rate

Our homeostatic mechanism is based on a dynamical regularization of the long-term neural firing rate to a target r_{tg}. However, since the biophysical parameter r_{tg} of the scaling process is also expected to vary from cell to cell, the scaling mechanism is useful only if the working memory behavior is robust when the parameter r_{tg} itself is heterogeneous. We tested the effect of heterogeneity in the target firing rate by assigning r_{tg} for each cell randomly from a Gaussian distribution with a standard deviation equal to 10% of the mean. Figure 5A shows that, although this increases slightly the mean drift speed $\bar{v}$, the increase remains constant as the level of heterogeneity in other single cell parameters is increased. This is in contrast to the case without homeostatic scaling, where the mean drift speed grows dramatically with the amount of cellular heterogeneities. Since larger cellular variability just results in larger cell-to-cell differences in scaling factor after scaling has taken place, the extent to which the input-output curves of different neurons are aligned after scaling is approximately independent of the level of heterogeneity in cellular properties. Thus, after scaling, one expects the performance of the network to be mainly related to the level of heterogeneity in the target rate of different neurons. To asses the effect of heterogeneity in r_{tg} in realistic conditions, we used a large network ($N_E = 4,096$), with a 10% heterogeneity level in the four cellular properties which have a significant effect on the firing rate (see caption of Figure 5). Figure 5B shows the mean drift speed as a function of the heterogeneity level in r_{tg}, up to 50%. The deterioration in performance as r_{tg} becomes more variable is extremely slow. As is shown in Figure 5C for a network with 50% heterogeneity in r_{tg} (corresponding to the square in Figure 5B), once the location of the bump is selected by a transient stimulus, no substantial systematic drift is observed. This shows that

the network is much more tolerant to heterogeneity in r_{tg} (which is a parameter of the scaling process itself and is not regulated homeostatically) than to heterogeneity in the other cellular parameters (which scaling makes largely irrelevant). To gain some intuition about this surprising result, we assessed in Figure 5D the variability in long-term firing rates as the level of heterogeneity in four cellular parameters is increased. In addition, we also show the variability in long-term firing rates that would result from scaling with comparable levels of heterogeneity in r_{tg}. This shows that heterogeneity in r_{tg} leads to smaller cell-to-cell differences in long-term firing rates than heterogeneity in any other parameter alone, and of course also smaller than all of them combined. Due to this remarkable property, a mechanism using long term-firing rate as a feedback signal seems ideal for the homogenization of the spatial working memory network.

[Figure 5 about here.]

Diversity of single neuron tuning curves

We performed a final set of simulations to check that the network could accurately encode spatial information after synaptic scaling, when all sources of heterogeneities were present at the same time. The results are shown in Figure 6. The network is composed of $N_E = 20,000$ randomly connected excitatory cells, each cell receives $C = 2,000$ synapses on average (The connection probability obeys a binomial distribution, with $p = C/N = 0.1$). Synaptic efficacies are also random (see Methods). A 20% level of heterogeneity in four cellular parameters and the target firing rate was used. In this case, the synaptic scaling factors have a standard deviation of 8% of the baseline (unity), and range from 0.6 to 1.3. Clearly, the location of the transient cue can be read out with little error after six seconds of delay. We explicitly checked that this behavioral performance could not simply be explained by the large network size. Without homeostatic scaling, no bump state is stable, and the network settles into a disorganized state of high activity. Therefore, despite a large amount of heterogeneity, the recurrent network as a whole acts as if it was effectively homogeneous, and is able to support a (quasi)continuum of bell-shaped persistent activity patterns.

At the same time, each individual neuron has a special combination of cellular and synaptic parameters that varies considerably from cell to cell. Therefore, single

neurons display a great variety of tuning curves. Ten examples are shown in (Figure 6B). For each neuron, simulation data are fitted by $r(\theta) = c_1 + c_2(\exp(c_3\cos(2\pi(\theta - \theta_{\text{pref}})/360))$. The amplitude is given by the peak firing rate minus the baseline, and the tuning width is defined at 60% of the peak amplitude. In Figure 6C, the half-width and peak amplitude of the tuning curve from single neurons are plotted against each other (color coded dots correspond to the examples shown in Figure 6B). Also shown are their histograms. A more than twofold variability in both tuning characteristics is evident, but the two do not appear to be strongly correlated. The main determinant of the height of the tuning curve of each cell is its target firing rate, whereas the correlation is weaker for the tuning curve width (Figure 6D).

In spatial working memory experiments, prefrontal neurons were found to display a wide range of tuning curves. Some are sharply tuned, others are broadly tuned (Funahashi et al. 1989; Chafee and Goldman-Rakic 1998). This diversity of spatial selectivity of mnemonic neural activity is reproduced by our model.

[Figure 6 about here.]

Activity dependent synaptic scaling is therefore efficient in its homogenization of the working memory network even in the presence of heterogeneities in the parameters which regulate the scaling process itself. Thus, it provides a robust solution to the heterogeneity problem in networks with a continuous family of localized activity patterns. We conclude that this theoretical framework represents a biologically plausible instantiation of spatial working memory at the circuit level.

Discussion

Using a biologically plausible cortical network model of spatial working memory, we studied the effect of heterogeneity on the active maintenance of spatial information in the form of persistent localized activity patterns. We found that the network's memory behavior is extremely sensitive to even a small amount of heterogeneity. This sensitivity results in the loss of stored information about a stimulus location in just a few seconds of the mnemonic delay period. The effects of heterogeneities are still unacceptably large even in networks with a realistic number of connections per cell ($N \sim 5,000$ in fully connected networks and $N \sim 20,000$, $C \sim 2,000$ in sparse networks). This is partly due to the weak dependence of the effect of

heterogeneities on the network size, according to a scaling law $\sim 1/\sqrt{N}$ (Figure 2B), a characteristic feature of disorder in single cell properties. By contrast, the heterogeneity effect decreases much faster for sparse connectivity ($\sim 1/\sqrt{CN}$) and for disorder in the synaptic coupling strength ($\sim 1/N$).

We found that when a homeostatic synaptic scaling mechanism was incorporated into the model, robust short-term storage of spatial information could be achieved in spite of biologically plausible levels of heterogeneity. Homeostatic plasticity scales the excitatory synaptic inputs to each cell to achieve a similar level of long-term activity for each neuron across the population, a process with a characteristic rate of 25%-50% change per 24 hours (Turrigiano et al. 1998). If the overall synaptic input to each cell is roughly the same during this period, the fine-grained (a few percent) adjustment of synaptic strengths by homeostatic scaling results in the alignment of the firing frequency versus synaptic input relationship of single cells with different biophysical properties (Figure 3D). This process renders the network effectively homogeneous.

The homeostatic mechanism is effective even in the presence of a large amount of heterogeneity in the scaling process itself (Figure 5). The only substantial effect of increasing variability in the target firing rate is a progressive destabilization of the uniform spontaneous state. For heterogeneities larger than $\sim 20\%$, it spontaneously evolves into a bump state (data not shown). This raises the experimental question of the actual variability of long-term (hours to days) firing rate across neurons in a cortical local network. On the other hand, it remains to be seen whether, in our model, the spontaneous state would be more robust with some adjustments of network parameters, or when other forms of homeostatic mechanisms (see below) are included. Finally, it is conceivable that bistability between a resting state and bump attractors involves additional mechanisms (such as a neural population which acts as a simple bistable switch, but with persistent activity which is not stimulus-selective). In any event, elucidating the circuit mechanisms for a continuum of localized persistent states represents a fundamental and difficult problem which is the focus of this paper.

Uniform distribution of long-term activity across the network

Effective or functional homogeneity is an essential requirement for spatial working memory function achieved through a quasi-continuum of attractor states. Our

results show that functional homogeneity can be dissociated from homogeneity in network parameters; it can be realized by activity-dependent regulatory mechanisms even in the presence of cellular heterogeneities. In order for such a mechanism to be effective, the distribution of long-term activity across the network has to be unbiased. This would be the case if the network is exposed to a uniform distribution of spatial cues over a long period of time. This is a reasonable assumption, since if a cortical circuit subserves spatial working memory for the animal, it should be constantly used by the animal in its daily life, and spatial cues that occur in real world are likely to be uniformly distributed.

One may ask whether the required functional homogeneity of long-term neural firing rates can be altered by over-training an animal with a biased non-uniform distribution of cues, for instance by using only eight stimuli as in a typical ODR experiment. The answer is not obvious since it depends on the relative importance of the eight cues used in the laboratory versus all the cues processed by the same working memory circuit in the animal's everyday life. Nevertheless, we tested the robustness of the homeostatic mechanism in our model, by varying the number of cues used to find the scaling factors (Figure 4). As expected, we found that scaling starts to fail when the number of cues becomes smaller than the number of non-overlapping activity profiles that fit in the network. Given the typical width of the neuron's tuning curves, of approximately 90 degrees, scaling ceases to be effective when only four cues or less are used. However, in the context of the task, the use of a very small (2-4) number of items is incompatible with the premise that the network subserves the encoding of an analog quantity (the spatial location) in a continuous manner. Finally, let us emphasize that our simulations with a small number of cues were done only to assess the robustness of the proposed homeostatic mechanism. To properly investigate the effect of a limited number of stimuli during training, the model should take into account the main driving force for learning during the animal's training, namely Hebbian long-term plasticity. Synaptic scaling in itself should not be viewed as a learning mechanism that creates the persistent firing patterns, and determines the animal's psychophysical performance. Instead, in addition to Hebbian plasticity, homeostatic regulation fine-tunes the synapses (by a few %) to homogenize the network.

Diversity of neuronal responses

In our network, heterogeneities in cellular parameters (specially in the target firing rate) lead to a sizable degree of diversity in stimulus selectivity characteristics, such as tuning curve height and width (Figure 6B-D). It would be interesting to do a quantitative analysis to estimate if this level of diversity is compatible with the available neurophysiological data, or if additional sources of diversity are needed. A main source of tuning diversity unrelated to random heterogeneities is likely to be the plurality of cell populations. For example, pyramidal neurons and different subtypes of interneurons are expected to exhibit different tuning curves (indeed, a recent study indicates that fast-spiking putative interneurons show broader tuning curves than regular-spiking putative pyramidal cells (Constantinidis and Goldman-Rakic 2002)). It is therefore important to emphasize that diversity of tuning properties should not be expected to arise exclusively from heterogeneity in cellular or synaptic properties. The same point applies to other forms of diversity of mnemonic activities, such as ramping-up or ramping-down time courses of delay period activity (see e.g. Chafee and Goldman-Rakic (1998)). These various neural firing patterns during the delay period are likely to have meaningful computational roles (like time integration, planning), and should be distinguished from heterogeneity at the cellular or circuit level.

Experimental questions and tests

There are two main general assumptions in our implementation of the scaling process (described by equation (2)). The first one is that the goal of this process is to control the long-term activity of the cell, by keeping it close to some target value. Experiments with neuronal cultures show that the down- and up-regulation of synaptic strengths in response to pharmacologically-induced prolonged changes in activity tend to restore the stable activity level that the neurons had before the pharmacological agents were applied (Turrigiano et al. 1998). However, it remains to be determined quantitatively whether firing rates are indeed conserved. Our results showing that scaling is still effective in the presence of cell-to-cell fluctuations in target firing rate indicate that the conclusions of this study would still be valid in the absence of a precise conservation of long-term activity. The second assumption is that scaling takes place in each neuron on the basis of information regarding only

this neuron, in particular its firing rate. Evidence indicates that the signal used for the down-regulation of synaptic strength is the average level of post-synaptic depolarization (Leslie et al. 2001), which is a property of the single neuron. Additional experiments are needed to further test this important assumption, and to identify underlying molecular mechanisms (Rutherford et al. 1998; Zafra et al. 1991; Wetmore et al. 1994).

Although we have only considered one form of activity-dependent regulation, it is conceivable that the homeostatic process involves a regulatory network with multiple interconnected components, including those affecting intrinsic membrane properties (ion channels) (Marder 1998; Desai et al. 1999; Stemmler and Koch 1999), pre-synaptic mechanisms (Murthy et al. 2001; Burrone et al. 2002) and strength of inhibitory synapses (Soto-Treviño et al. 2001; Kilman et al. 2002). Another mechanism which should be explored is the contribution of bistability at the single cell level (Camperi and Wang 1998; Lisman et al. 1998; Koulakov et al. 2002; Stringer et al. 2002; Egorov et al. 2002) to the robustness of networks with a continuum of localized activity profiles. Future studies are needed to explore the differential/cooperative roles of these different processes in network homogenization and robust memory behavior.

Robustness in continuous attractor networks

Robustness represents a major open issue for many models of brain systems involved in the internal representation and short-term memory of analog quantities, such as stimulus location (Funahashi et al. 1989) or frequency (Romo et al. 1999), head direction (Sharp et al. 2001), eye position (Robinson 1989), place in the environment (McNaughton et al. 1996), and possibly numerical magnitude (Nieder and Miller 2003). An accurate way to retain the analog nature of the encoded information is by means of a continuum of neuronal firing patterns (continuous attractors) (Ben-Yishai et al. 1995; Camperi and Wang 1998; Compte et al. 2000; Zhang 1996; Samsonovich and McNaughton 1997; Seung et al. 2000). However, whether the brain does use a continuum of internal representations has not yet been demonstrated. In physiological experiments, discrete sets of stimuli (e.g., 8 spatial cues in working memory experiments) are commonly used. This is partly due to the limited time a neuron could be recorded stably while the complete stimulus set could be sampled. It would be highly desirable if the encoding scheme of analog stimuli

could be rigorously tested in future experiments. If the brain does use continuous internal representations, then regardless of the particular encoding scheme, a fundamental trade-off exists between the accuracy of memory maintenance and the fine-tuning of network parameters. Our results show that by using a feedback signal based on long-term neural activity, homeostatic control mechanisms are able to relax constraints on a large number of biophysical parameters, thereby allowing efficient maintenance of information when fine-tuning is destroyed by biologically plausible levels of heterogeneity. The present study suggests that the interplay between these adaptive processes and the dynamics of recurrent cortical circuits could underlie the robustness of delicate computational and memory operations in complex neural networks.

Methods

The Model

The model used has been described in detail in (Compte et al. 2000). Single neurons are described by the leaky integrate-and-fire model (Tuckwell 1988); synaptic currents are calibrated by the measurements of excitatory and inhibitory postsynaptic currents in cortical neurons. We have used the same values for all single cell and synaptic parameters as those listed in (Compte et al. 2000) except for the value of the four recurrent synaptic conductances. There are N_E and N_I excitatory and inhibitory cells in the network. In the fully connected simulations we set $N_I = N_E/4$, and the values of the four unitary maximal synaptic conductances were: NMDAR-mediated conductance to pyramidal cells $G_{EE} = 0.931/N_E$ μS and interneurons $G_{EI} = 0.75/N_E$ μS, and $GABA_A$R-mediated conductance to pyramidal cells $G_{IE} = 1.024/N_I$ μS and interneurons $G_{II} = 0.825/N_I$ μS. In some simulations the connections within the pyramidal cell population were random and sparse. In these simulations, the probability of connection $p(\theta_i - \theta_j)$ between two excitatory neurons i and j decreases with the difference between their preferred cues. We used the same functional form for $p(\theta_i - \theta_j)$ as the spatial connectivity in the fully-connected network (Compte et al. 2000), but with a prefactor chosen so that $\sum_j p(\theta_i - \theta_j) = C_E$. Since the connection is all-or-none, C_E is the average number of synaptic connections per cell. In all our sparse simulations $C_E = 0.1N_E$. If any two pyramidal cells happened to be connected, their synaptic connection had

maximal NMDAR-mediated maximal conductance of $G_{EE} = 0.931w$, where w is a random gaussian variable (independent for each synapse) with mean $1/C_E$ and standard deviation $0.1/C_E$. In these simulations, $N_I = C_E/4$.

A mean-field version of the model (described in detail in Brunel and Wang (2001)) was also used. The mean-field calculation we use expresses the instantaneous firing rate of the cell as a function of the statistical properties of its inputs, as determined by the mean and the fluctuations in its synaptic currents mediated by NMDA-, AMPA- and $GABA_A$ receptors.

Heterogeneity in a cellular parameter is introduced using a random distribution across the network. When the heterogeneity is in a voltage parameter such as the resting potential V_L or spike threshold V_{th}, what matters is the magnitude of heterogeneity (the standard deviation) compared to $(V_{th} - V_{reset}) = -50 - (-60) = 10$ mV. Hence the percentage of heterogeneity is expressed relative to 10 mV, e.g. 10 % means 1 mV.

The local average excitability $U(\theta)$

The function $U(\theta)$ represents the local average excitability around θ. To calculate it, we first quantify the excitability $E(\theta)$ of each cell. Intuitively, we measure the excitability of a cell as its long-term average firing rate over a long period of time when the distribution of spatial stimuli processed by the network is assumed to be unbiased. Specifically, for each neuron the average is performed for its firing rate over $N_c = 256$ bump states, each with the same synaptic input profile as in the homogeneous network, but with different peak locations interspaced by $360^o/N_c$. Since the bump is displaced one full circle, the complete set of excitatory synaptic inputs $\{I_i\}$, $i = 1, 2, ..., N_c$ is the same for every neuron. The excitatory synaptic inputs spanned by this set corresponds to the range of excitatory synaptic inputs used in Figure 5D. According to this definition, cell θ_1 is more excitable than cell θ_2 if, given the set of synaptic inputs $\{I_i\}$, the long-term firing rate of θ_1 is larger than that of θ_2. Since the distribution of single cell parameters is random across the network, $E(\theta)$ fluctuates abruptly from one cell to another. To calculate $U(\theta)$, we smooth out, or average locally, the function $E(\theta)$. We found empirically that a good agreement between the simple model Eq. (1) and direct network simulations was obtained by using the excitatory synaptic activity profile, $S(\theta)$, from the homogeneous network (normalized so that $\int_0^{360^o} S(\theta)d\theta = 1$) as a smoothing kernel. $S(\theta)$ is a bell shaped

function with width similar to the network activity profile, and measures the mean recurrent excitatory synaptic drive across the network when the network is in a bump state. Thus $U(\theta) \equiv \int_0^{360^o} S(\theta' - \theta)E(\theta')d\theta'$. The spatial scale of variation of $U(\theta)$, determined by the width of the localized activity profile, determines the number of privileged locations in the network. This number is, therefore, relatively insensitive to the number of cells in the network. In our simulations it is typically equal to two, although significantly narrower (wider) activity profiles would result in an increase (decrease) in the number of privileged locations.

Phenomenological model for the systematic drift speed

We found that the drift speed $v(\theta)$ of the bump's peak location at every point of the network θ in the presence of heterogeneity is very well approximated by a function proportional to the derivative of the average local excitability $U(\theta)$ at that point, $v(\theta) = k\ dU(\theta)/d\theta$. We calculated the value of the proportionality constant k by comparing the real drift speeds obtained from 50 simulations involving different network sizes and different types of heterogeneity, with those predicted by this phenomenological model, and then used it throughout our study. In addition to the systematic drift provoked by the heterogeneity, the bump's peak location fluctuates randomly in the simulations due to the random background activity impinging on every neuron. To emulate this random drift, we added a fluctuating component, $\eta(t)$, to the drift velocity in the phenomenological model, $v(\theta) = k\ dU(\theta)/d\theta + \eta(t)$. We model $\eta(t)$ as a white-noise with a standard deviation dependent on the network size. For each network size, N, we set this standard deviation so that, in a homogeneous network (when $U(\theta)$ is flat and $dU(\theta)/d\theta = 0$), the simple model reproduces the random walk process of the original spiking network of size N (Compte et al. 2000).

Calculation of the two-dimensional saccade distributions

Although our model is one-dimensional, we show the distribution of peak locations after 6 seconds of delay in Figure 2C in two dimensions, to ease the visual comparison with the behavioral data on the spatial distribution of memory guided saccades (e.g. Figure 1 in (Chafee and Goldman-Rakic 1998)). To do this, we fix the eccentricity on the two-dimensional plane. For each of the 8 stimuli, we calculate the mean angular

location of the bump state after 6 seconds across the 100 trials. Each particular trial is then located at a randomly chosen position on a circle with the center at the mean location, and a radius equal to the distance between the mean and the final peak location of the bump in that trial.

Steady state of the scaling process

To find the steady state scaling factor for each cell we used the mean field version of the model. For the particular realization of the network parameters under study, we created N_{rep} identical replicas of this network. In each replica, a very weak, spatially localized current, was persistently applied at some location to anchor the corresponding bump state. The location of the anchored bump state in the i^{th} replica was $i\ 360^o/N_{rep}$, where $i = 1, \ldots, N_{rep}$. Thus, if each network was made of N neurons, NN_{rep} equations were integrated simultaneously to simulate the time evolution of the firing rate of each neuron in each replica. N additional equations for synaptic scaling (eq. 2) were also simultaneously integrated. The scaling process was said to have reached steady state when all $N(N_{rep}+1)$ dynamical variables had converged numerically. Typically we used N_{rep} =8, 20, 40.

Fitting of single neuron tuning curves

The tuning curves in Figure 6 were calculated by fitting the 20 firing rates for each neuron from the simulations in Figure 6A with the four parameter curve $r(\theta) = c_1 + c_2(\exp(\mathrm{c}_3\cos(2\pi(\theta - \theta_{\mathrm{pref}})/360))$. We defined the height as $c_2(\exp(\mathrm{c}_3) - \exp(-\mathrm{c}_3))$ and the half width at approximately 60% of the peak firing rate, by solving σ from the equation $\exp(c_3\cos(\pi\sigma/360)) - \exp(-c_3) = \exp(-0.5)(\exp(\mathrm{c}_3) - \exp(-\mathrm{c}_3))$.

Acknowledgments

The authors thank A. Compte for providing the code for the model simulations and for help throughout the completion of this work. We also thank M.C.W. van Rossum, P. Miller, G. Turrigiano, E. Marder and L.F. Abbott for discussions. Support was provided by the National Institutes of Mental Health (MH62349, DA016455), the A.P. Sloan Foundation and the Swartz Foundation.

References

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

Braitenberg, V. and A. Schütz (1991). *Anatomy of the cortex*. Springer-Verlag, Berlin.

Brunel, N. and X.-J. Wang (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci. 11*, 63–85.

Burrone, J., M. O'Byrne, and V. N. Murthy (2002). Multiple forms of synaptic plasticity triggered by selective suppression of activity in individual neurons. *Nature 420*, 414–418.

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

Chafee, M. V. and P. S. Goldman-Rakic (1998). Neuronal activity in macaque prefrontal area 8a and posterior parietal area 7ip related to memory guided saccades. *J. Neurophysiol. 79*, 2919–2940.

Compte, A., N. Brunel, P. S. Goldman-Rakic, and X.-J. Wang (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex 10*, 910–923.

Constantinidis, C. and P. S. Goldman-Rakic (2002). Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J. Neurophysiol. 88*, 3487–3497.

Deneve, S., P. E. Latham, and A. Pouget (1999). Reading population codes: a neural implementation of ideal observers. *Nat. Neurosci. 2*, 740–745.

Desai, N. S., S. B. Nelson, and G. G. Turrigiano (2002). Critical periods for experience-dependent synaptic scaling in visual cortex. *Nat. Neurosci. 5*, 783–789.

Desai, N. S., L. C. Rutherford, and G. G. Turrigiano (1999). Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nat. Neurosci. 2*, 515–520.

Egorov, A. V., B. N. Hamam, E. Fransen, M. E. Hasselmo, and A. A. Alonso (2002). Graded persistent activity in entorhinal cortex neurons. *Nature 420*, 173–178.

Funahashi, S., C. J. Bruce, and P. S. Goldman-Rakic (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol. 61*, 331–349.

Georgopoulos, A. P., J. Kalaska, and R. Caminiti (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci. 2*, 1527–1537.

Gnadt, J. W. and R. A. Andersen (1988). Memory related motor planning activity in posterior parietal cortex of macaque. *Exp. Brain Res. 70*, 216–220.

Goldman-Rakic, P. S. (1992). Working memory and the mind. *Sci. Am. 267*, 110–117.

Kilman, V., M. C. W. van Rossum, and G. G. Turigiano (2002). Activity scales inhibitory synaptic strengths by regulatingthe number of postsynaptic gaba$_A$ receptors. *J. Neurosci. 15*, 1328–1337.

Koulakov, A. A., S. Raghavachari, A. Kepecs, and J. E. Lisman (2002). Model for a robust neural integrator. *Nat. Neurosci. 5*, 775–782.

Leslie, K. R., S. B. Nelson, and G. G. Turrigiano (2001). Postsynaptic depolarization scales quantal amplitude in cortical pyramidal neurons. *J. Neurosci. 21*, RC170:1–6.

Lisman, J. E., J.-M. Fellous, and X.-J. Wang (1998). A role for NMDA-receptor channels in working memory. *Nat. Neurosci. 1*, 273–275.

Marder, E. (1998). From biophysics to models of network function. *Annu. Rev. Neurosci. 21*, 25–45.

Mason, A. and A. Larkman (1990). Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex. II. electrophysiology. *J. Neurosci. 10*, 1415–1428.

McNaughton, B. L., C. A. Barnes, J. L. Gerrard, K. Gothard, M. W. Jung, J. J. Knierim, H. Kudrimoti, Y. Quin, W. E. Skaggs, M. Suster, and K. L. Weaver (1996). Deciphering the hippocampal polyglot : The hippocampus as a path integration system. *J. Exp. Biol. 199*, 173–186.

Murthy, V. N., T. Schikorski, C. F. Stevens, and Y. Zhu (2001). Inactivity produces increases in neurotransmitter release and synaptic size. *Neuron 32*, 673–682.

Nieder, A. and E. K. Miller (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron 37*, 149–157.

O'Brien, R. J., S. Kambol, M. D. Ehlers, K. R. Rosen, G. D. Kischback, and R. L. Huganir (1998). Activity dependent modulation of synaptic ampa receptor accumulation. *Neuron 21*, 1067–1078.

O'Keefe, J. and J. Dostrovsky (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely moving rat. *Exp. Brain Res. 34*, 171–175.

Ploner, C. J., B. Gaymard, S. Rivaud, Y. Agid, and C. Pierrot-Deseilligny (1998). Temporal limits of spatial working memory in humans. *Eur. J. Neurosci. 10*, 794–797.

Rainer, G., W. F. Assad, and E. K. Miller (1998). Memory fields of neurons in the primate prefrontal cortex. *Proc Natl Acad Sci (USA) 95*, 15008–15013.

Rank Jr., J. B. (1985). Head direction cells in the deep cell layer of dorsal presubiculum in freely moving rats. In G. Buzsáki and C. Vanderwolf (Eds.), *Electrical activity of the archicortex.*, pp. 217–220. Budapest: Akadémiai Kiadó.

Redish, A. D., A. N. Elga, and D. S. Touretzky (1996). A coupled attractor model of the rodent head direction system. *Network 7*, 671–685.

Robinson, D. A. (1989). Integrating with neurons. *Annu. Rev. Neurosci. 12*, 33–45.

Romo, R., C. D. Brody, A. Hernández, and L. Lemus (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature 399*, 470–474.

Rutherford, L. C., S. B. Nelson, and G. G. Turrigiano (1998). BDNF has opposite effects on the quantal amplitude of pyramidal neuron and interneuron excitatory synapses. *Neuron 21*, 521–530.

Samsonovich, A. and B. L. McNaughton (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci. 17*, 5900–5920.

Seung, H. S. (1996). How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. (USA) 93*, 13339–13344.

Seung, H. S., D. D. Lee, B. Y. Reis, and D. W. Tank (2000). Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron 26*, 259–271.

Sharp, P. E., H. T. Blair, and J. Cho (2001). The anatomical and computational basis of the rat head-direction cell signal. *Trends in Neurosci. 24*, 289–294.

Skaggs, W. E., J. J. Knierim, H. S. Kudrimoti, and B. L. McNaughton (1997). A model of the neural basis of the rat's sense of direction. In D. S. T. G. Tesauro and T. K. Leen (Eds.), *Advances in Neural Information Processing Systems 7*, pp. 173–182. MIT press, Cambridge, MA.

Soto-Treviño, C., K. A. Thoroughman, E. Marder, and L. F. Abbott (2001). Activity dependent modification of inhibitory synapses in models of rhythic neural networks. *Nat. Neurosci. 4*, 297–303.

Stemmler, M. and C. Koch (1999). How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate. *Nat. Neurosci. 2*, 521–527.

Stringer, S. M., T. P. Trappenberg, E. T. Rolls, and I. E. de Araujo (2002). Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. *Network 13*, 217–242.

Taube, J. S. (1995). Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *J. Neurosci. 15*, 70–86.

Tsodyks, M. and T. J. Sejnowski (1997). Associative memory and hippocampal

place cells. In *Advances in Neural Information Processing Systems 6*, pp. 81–86. MIT press, Cambridge, MA.

Tuckwell, H. C. (1988). *Introduction to theoretical neurobiology*. Cambridge: Cambridge University Press.

Turrigiano, G. G. (1999). Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends in Neurosci. 22*, 221–227.

Turrigiano, G. G., K. R. Leslie, N. S. Desai, L. C. Rutherford, and S. B. Nelson (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature 391*, 892–896.

van Rossum, M. C. W., G. Q. Bi, and G. G. Turrigiano (2000). Stable hebbian learning from spike timing-dependent plasticity. *J. Neurosci. 20*, 8812–8821.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosci. 24*, 455–463.

Wetmore, C., L. Olson, and A. J. Bean (1994). Regulation of brain-derived neurotrophic factor (BDNF) expression and release from hippocampal neurons is mediated by non-nmda type glutamate receptors. *J. Neurosci. 14*, 1688–1700.

White, J. M., D. L. Sparks, and T. R. Stanford (1994). Saccades to remembered target locations: and analysis of systematic and variable errors. *Vision Res. 34*, 79–92.

Wilson, M. A. and B. L. McNaughton (1993). Dynamics of the hippocampal ensemble code for space. *Science 261*, 1055–1058.

Xie, X., R. H. Hahnloser, and H. S. Seung (2002). Double-ring network model of the head direction system. *Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys 66*, 041902.

Zafra, F., E. Castren, H. Thoenen, and D. Lindholm (1991). Interplay between glutamate and gamma-aminobutyric acid transmitter systems in the physiological regulation of brain-derived neurotrophic factor and nerve growth factor synthesis in hippocampal neurons. *Proc. Natl. Acad. Sci. (USA) 88*, 10037–10041.

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

Figure captions

Figure 1 Spatial working memory in the homogeneous network. (**A**) Schematic description of the network architecture. Excitatory pyramidal cells are arranged according to their preferred cues (arrows). The excitatory-excitatory synaptic strength decreases as a function of the difference between the preferred cues of cells. Connections to or from inhibitory interneurons are uniform. (**B**) Spatio-temporal network activity in a bump state. Left. Rastergram for the excitatory population. A dot at position (t, θ) represents a spike fired at time t by a cell with preferred cue θ. The blue line represents the time evolution of the peak location calculated by the population vector method. Right. Localized activity profile during the memory period calculated by counting the number of spikes fired by each cell in the delay and dividing by the delay duration. (**C**) Time evolution of the peak location of a localized firing pattern ('bump state') in a homogeneous network of $1,024$ excitatory cells (see Methods) during the delay period. Each line corresponds to a simulation in which the stimulus was at a different location. The peak location of the bump provides a good estimate of the location of the remembered cue during the delay period.

Figure 2 Effect of heterogeneity on working memory function. (**A**) Top left. Time evolution of the bump's peak location in simulations in which stimuli at different locations were used in different trials. The leak potential for each of the $1,024$ excitatory cells was drawn randomly from a Gaussian distribution with SD=1 mV, i.e. $V_L = -70 \pm 1$ mV. The memory of the cue's location is lost in a few seconds in the presence of heterogeneity. Top right. Spatial distribution of local average excitability in this network. Regardless of the initial peak location of a bump (encoding the spatial stimulus), the bump drifts toward one of the two locations corresponding to the maxima of $U(\theta)$. Bottom. Trajectories obtained by numerical integration of the simple model (Eq. 1). This model reproduces the systematic drift observed in the original spiking network model (Top left). (**B**) Mean drift speed $\bar{v}$ versus number of excitatory cells N_E. Each point represents the mean $\pm$ SEM of 100 networks of the same size with a given level of heterogeneity. Squares: heterogeneity as in Figure 2A. Inset. Log-log plot of the data in squares along with a linear fit with slope -0.5. The mean drift speed decreases as $\sim 1/\sqrt{N}$. The circle, upward triangle and diamond show the increase in $\bar{v}$ each time a new source of heterogeneity (C_m=0.5 $\pm$ 0.05 nF ; V_{th}=-50 ± 1 mV ; g_L=2.5 $\pm$ 0.25 nS respectively) is added, and the downward triangle corresponds to networks with $V_L = -70 \pm 2$ mV . The mean drift speed increases fast with the level of heterogeneity at fixed N. Star. Value of $\bar{v}$ for the example shown in **C**, which belongs to the distribution represented by the downward triangle. (**C**)

Top. Spatial distribution of end locations calculated from equation (1) in a typical network with $V_L = -70 \pm 2$ mV and a realistic number ($N_E = 4,096$) of connections per cell. Each colored dot represents the peak position (see Methods) of the bump after a delay period of 6 sec, for a trial in which the stimulus was at the center of the square of the same color. 100 trials per stimulus are shown. If the dot is within the large circle around the square of the same color, the memory-guided saccade is considered as correct in that trial. The spatially biased and non-uniform saccade distribution is inconsistent with the behavioral data from ODR experiments. Bottom. Function $U(\theta)$ for this network, using the same scale as in **A**. Vertical lines represent the 8 locations of the bump states at the beginning of the delay, the colors match those of colored dots in the above scatter plot. Over time the dots move toward maxima of $U(\theta)$.

Figure 3 Network homogenization by activity-dependent synaptic scaling. (**A**) Left. Time evolution of the bump's peak location in simulations in which stimuli at different locations were used, for the same network in Figure 2A after scaling. The systematic drift speed is almost completely suppressed. Right. Spatial distribution of local excitability before (dashed) and after (solid) scaling. Scaling flattens out the excitability distribution, rendering the network effectively homogeneous. (**B**) Mean drift speed before and after scaling in networks with 256 excitatory cells with a 10% heterogeneity in leak potential (1 mV; cyan), membrane capacitance (red), threshold potential (1 mV; yellow), and leak potential plus membrane capacitance (magenta). Points with the same color refer to different sample realizations of heterogeneity distribution. Scaling is robust to different sources of heterogeneity. (**C**) Top. Scaling factor versus membrane capacitance for each cell in a network with C_m=0.5 $\pm$ 0.05 nF after scaling. Bottom. Same for a network with V_L=-70 ± 1 mV. The scaling factor is mostly determined by the values of the heterogeneous parameters for each cell. Note that the scaling factor is only a few percent of the baseline synaptic strength. (**D**) Firing rate as a function of the recurrent synaptic drive, g_{syn}, for 50 neurons chosen at random from a network with a 10% heterogeneity level in V_L, C_m, V_{th} and g_L. The excitatory synaptic current to each cell is $I_{syn} = g g_{syn}(V - V_{syn})$, where g is the scaling factor. Left. Before scaling ($g = 1$). Right. After scaling, g reaches its steady-state, g_{ss}, when the long-term firing rates of all cells become equalized in spite of their heterogeneous cellular properties. The shown range of the synaptic drive, g_{syn}, corresponds to the one that cells experience in the recurrent network.

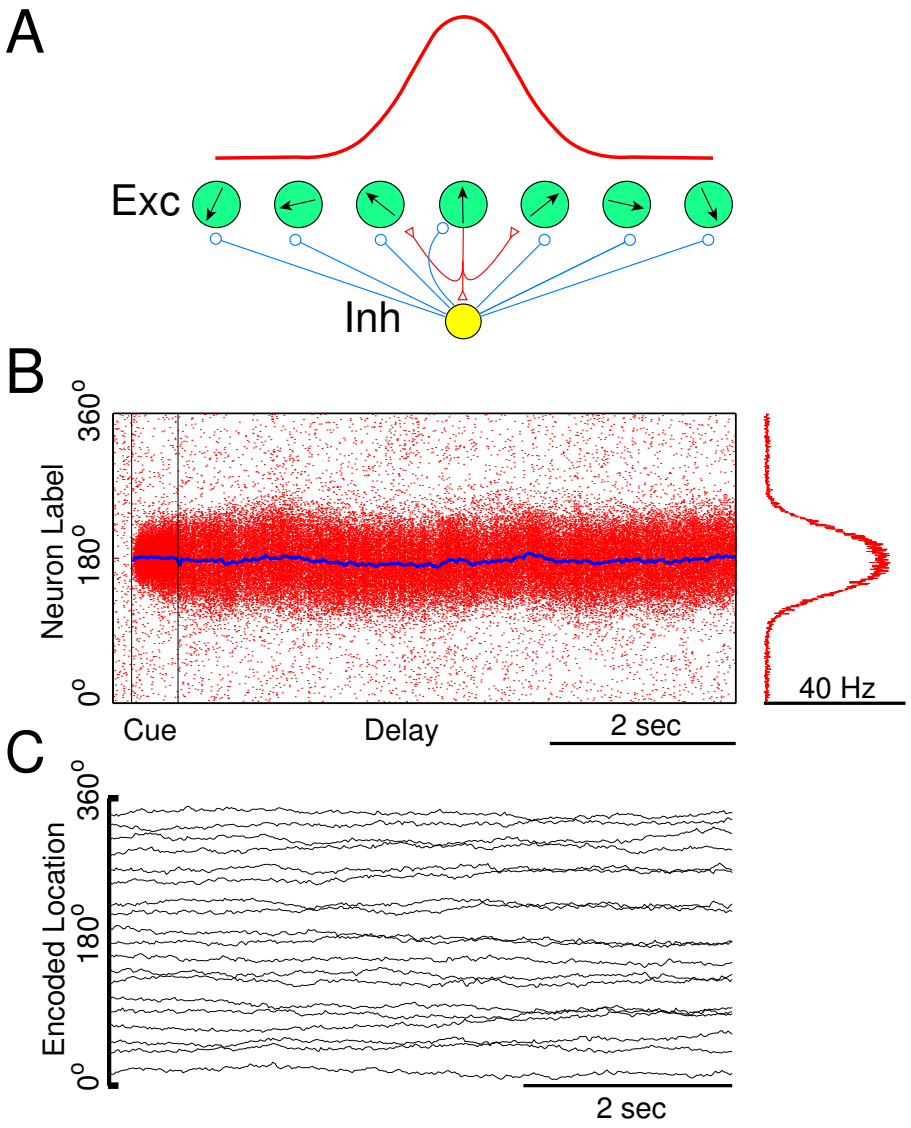
Figure 4 Robustness of homogenization by synaptic scaling with a decreasing number of discrete cues. The scaling factors were computed using 20 (**A**), 8 (**B**) and 4 (**C**) cues. The network is the same as in Figures 2A and 3A. Left. Temporal evolution of the bump peak location in simulations where 20 (evenly

spaced) cues were used in different trials. Right. Local average excitability $U(\theta)$ for this network, with the same scale as in Figures 2A and 3A. In the case with 4 cues (bottom), the dashed lines on the right panel represent the network subregions of higher local average excitability, which are those in between the four cues used during the homeostatic scaling process. Homogenization by scaling is efficient as long as the number of cues for scaling is larger than the number of non-overlapping network activity profiles which can be fit in the network (approximately 4).

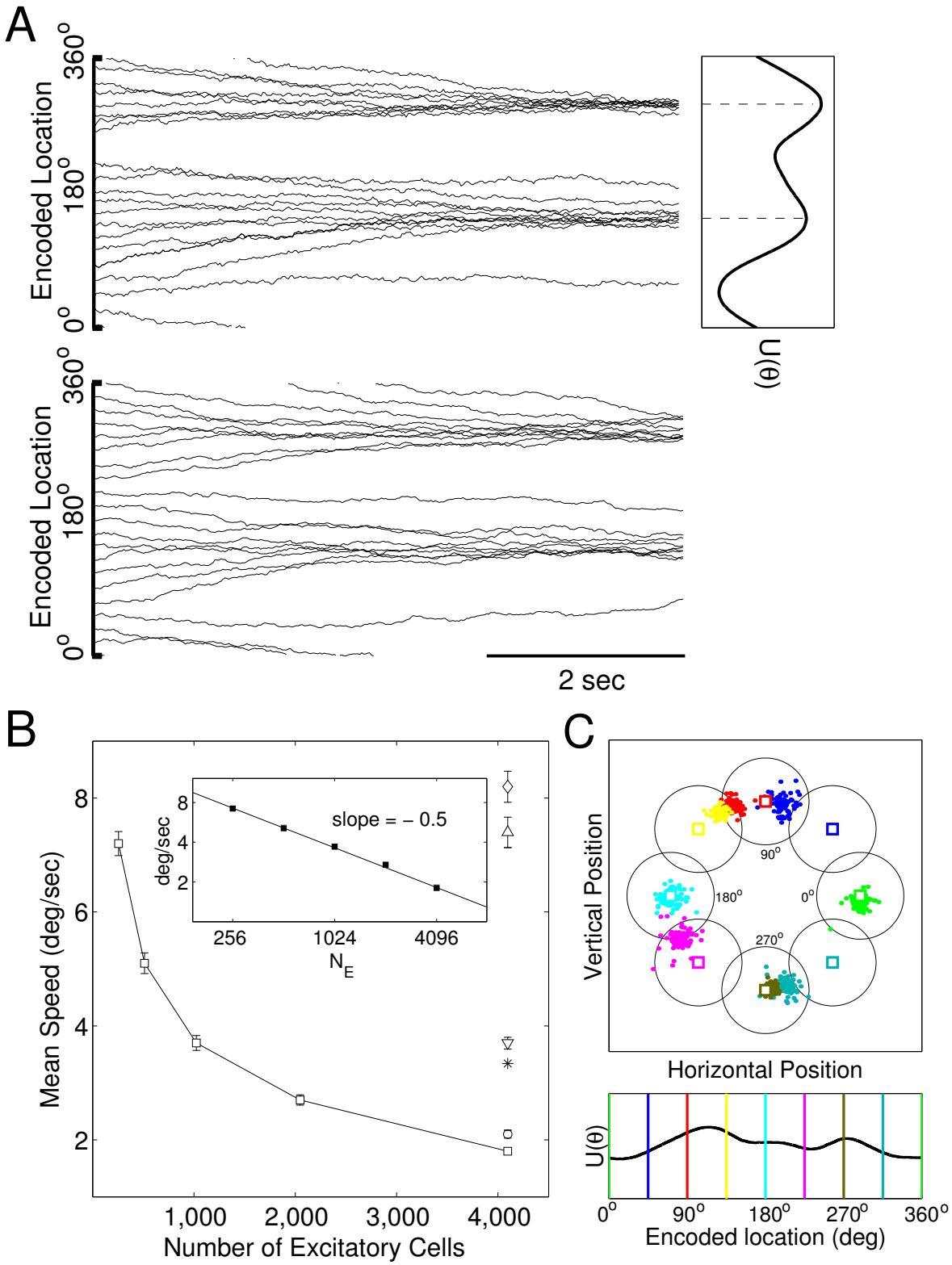
Figure 5 Robustness of the scaling mechanism in the presence of heterogeneity in the target firing rate of different cells. (**A**) Mean drift speed versus the level of heterogeneity before scaling (triangles), after scaling with SD(r_{tg})=0 (squares) and after scaling with a 10% heterogeneity in the target firing rate of different cells (r_{tg}=13.5 $\pm$ 1.35 Hz) (circles) for networks of 256 excitatory cells with heterogeneity in the leak potential and the membrane capacitance. (**B**) Mean drift speed versus the level of heterogeneity in r_{tg} for networks of $4,096$ excitatory cells and a 10% heterogeneity level in the leak potential and conductance, membrane capacitance and voltage threshold. For each heterogeneity level, four data points correspond to different random realizations of the distribution of heterogeneities. Note different scales in the drift speed in **A** and **B**. (**C**) Temporal evolution of the bump peak location for 20 trials that started at different cue locations, for a network with 50% heterogeneity in r_{tg} (marked with a black square in **B**). The memory of the transient input is maintained accurately despite the large amount of heterogeneity. (**D**) Relative variability in the distribution of neuronal excitability as a function of the level of heterogeneity in several cellular properties. *All* refers to the combined heterogeneity in the four cellular parameters. After scaling with heterogeneity in r_{tg} (bars on the right), the distribution of excitability across the network is much more narrow that the one resulting from heterogeneity in other cellular properties in the absence of scaling. In the simulations of this figure, we used 20 cues to calculate the scaling factors.

Figure 6 Robust encoding of spatial information by synaptic scaling and a wide diversity of tuning curves in a highly heterogeneous network. The network is composed of $N_E = 20,000$ randomly connected excitatory neurons, with an average of $C = 2,000$ synapses per cell, and with random synaptic efficacies (see Methods). A 20% heterogeneity in V_L, C_m, V_{th} and g_L was used, and the scaling factors were obtained using 8 cues and a 20% heterogeneity in r_{tg}. (**A**) Temporal evolution of the bump peak location in 20 trials with transient stimuli at different locations. The memory of the initial cue is well preserved during the 6 second delay period. (**B**) Examples of tuning curves from ten neurons. The firing rate of each cell for the 20 stimuli in **A** are shown as empty dots. Solid lines are obtained by fitting the data for each cell with the formula

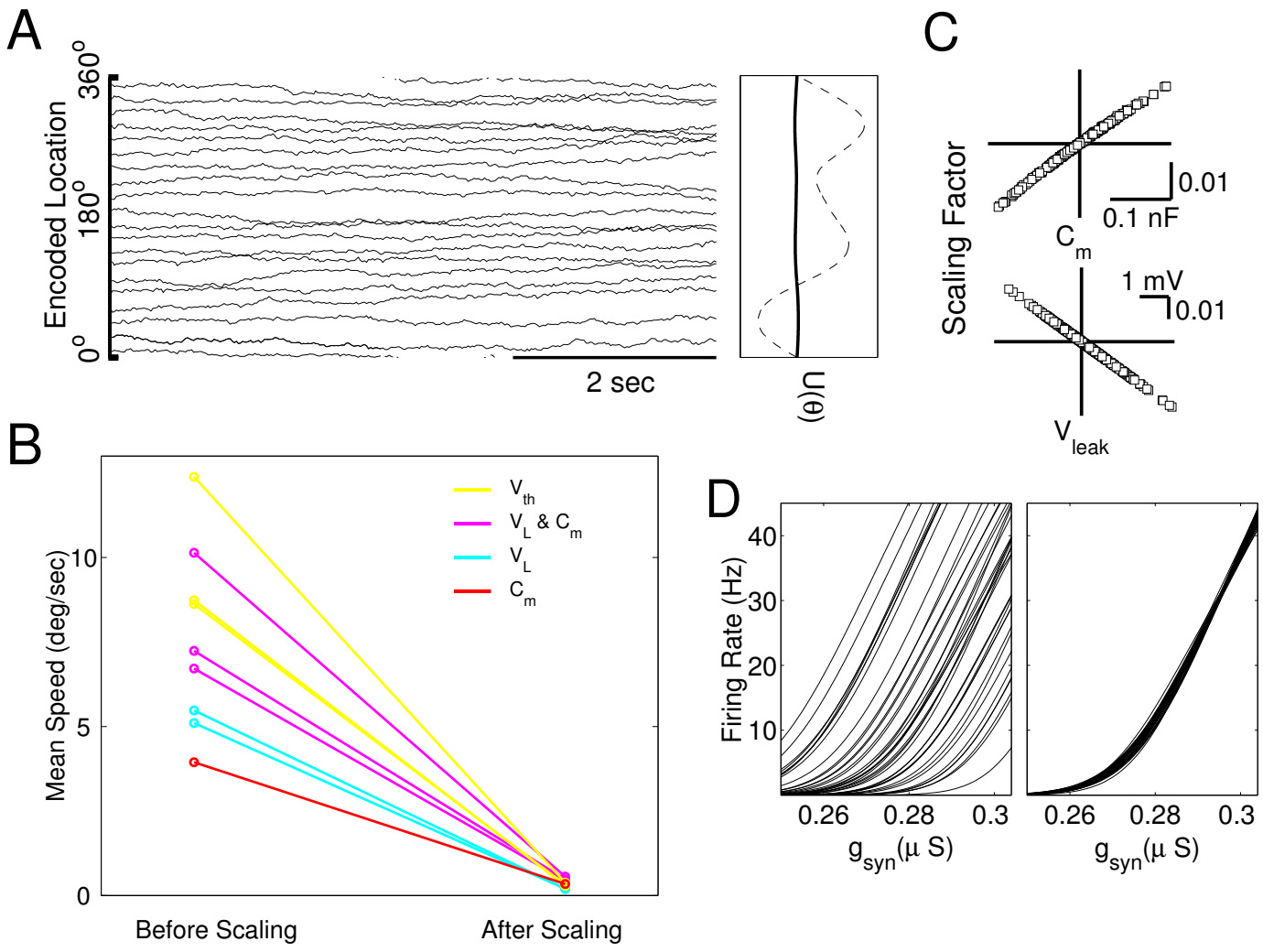
$r(\theta) = c_1 + c_2(\exp(c_3\cos(2\pi(\theta - \theta_{\text{pref}})/360))$, where θ_{pref} is the preferred cue of the cell, r is its firing rate and c_1, c_2 and c_3 are constants characterizing the shape of the tuning curve. (**C**) Tuning curve half-width versus tuning curve height (see Methods) for the $20,000$ neurons. The ten neurons in **B** are indicated by dots of the same color as the corresponding tuning curves. On the left and top are the corresponding histograms of half-width and height. (**D**) Tuning curve height (Top) is significantly, and half width (Bottom) more weakly correlated with the target firing rate of each cell.



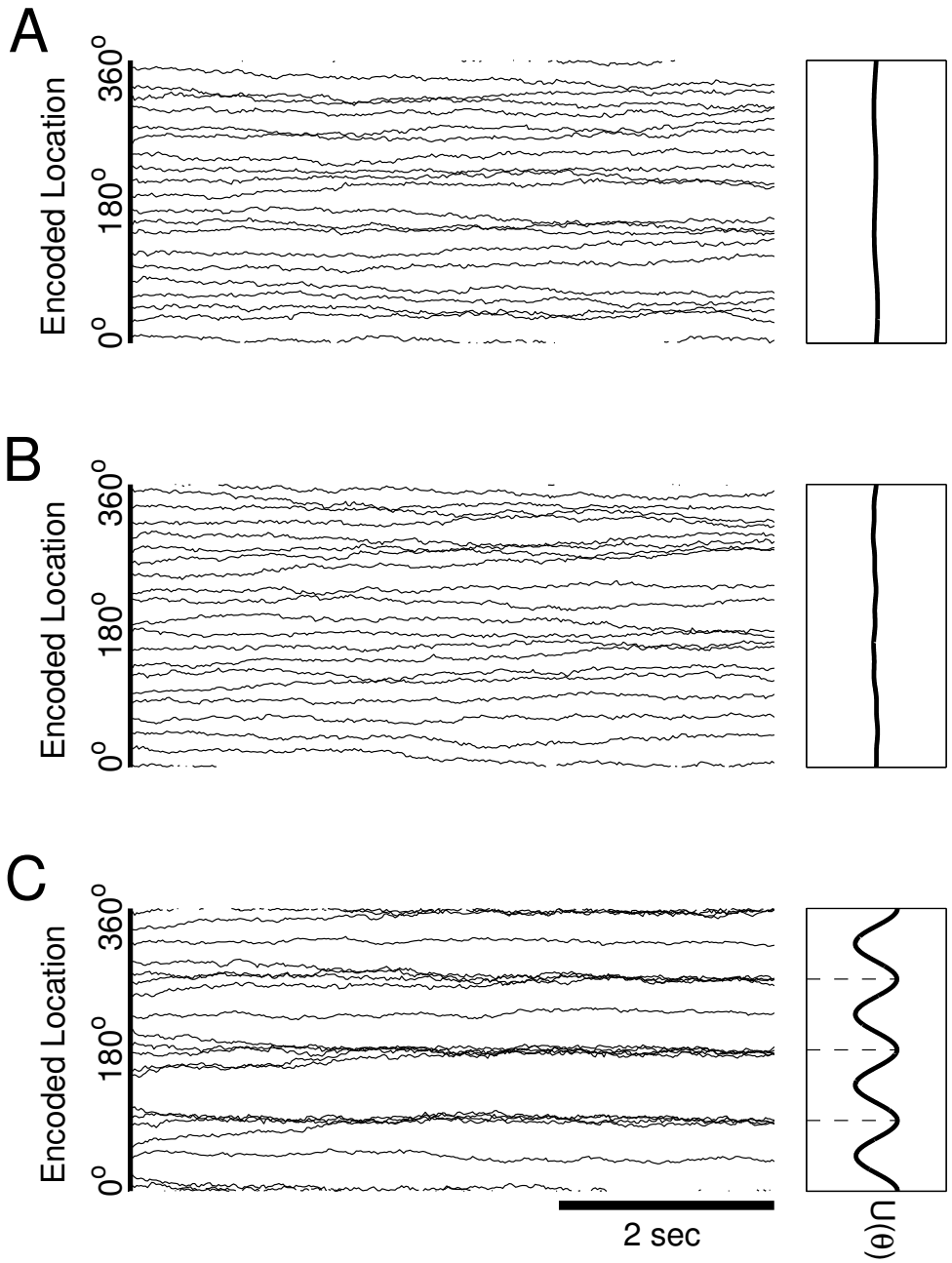
X.-J. Wang ; FIGURE 1



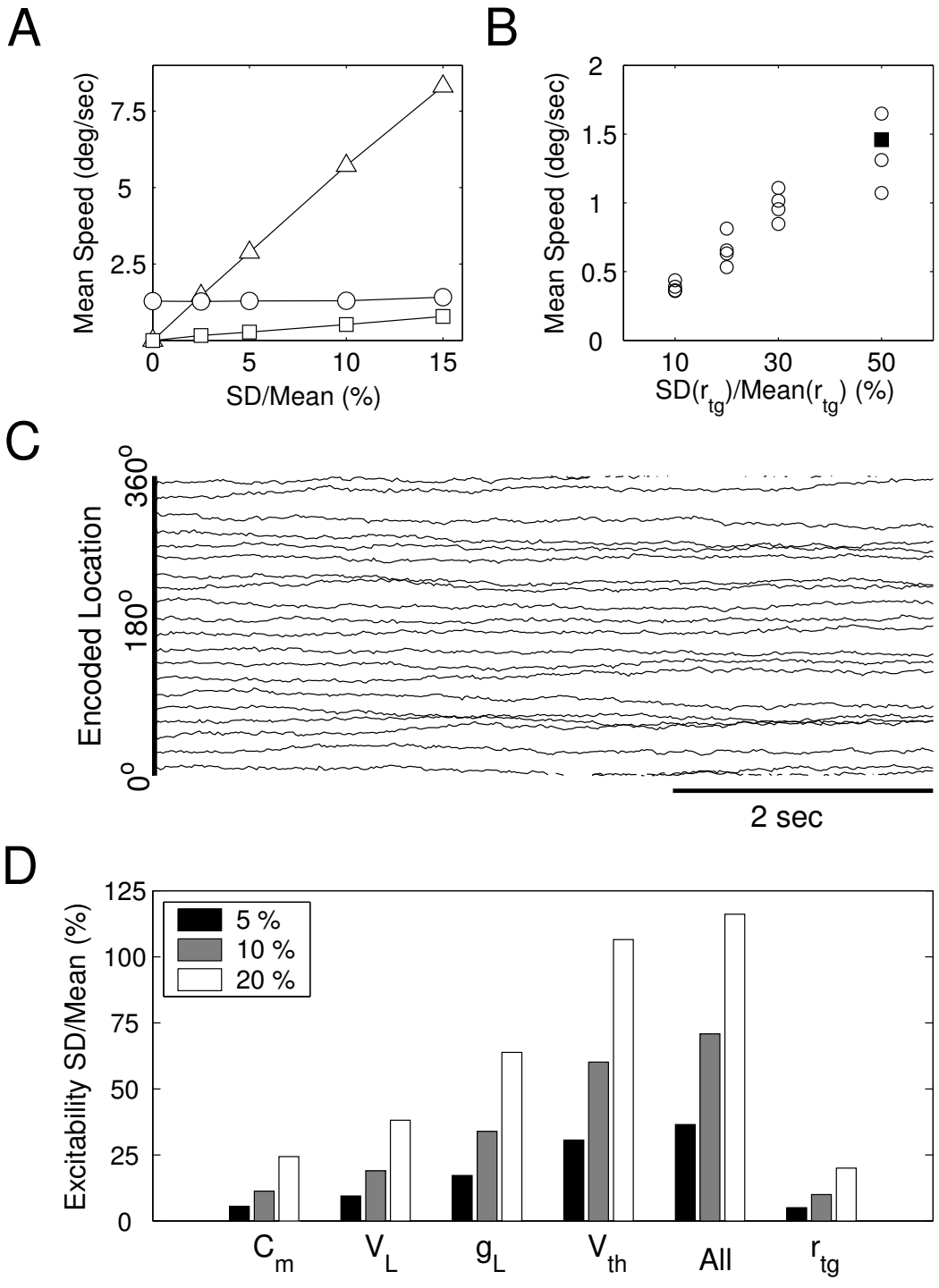
X.-J. Wang ; FIGURE 2



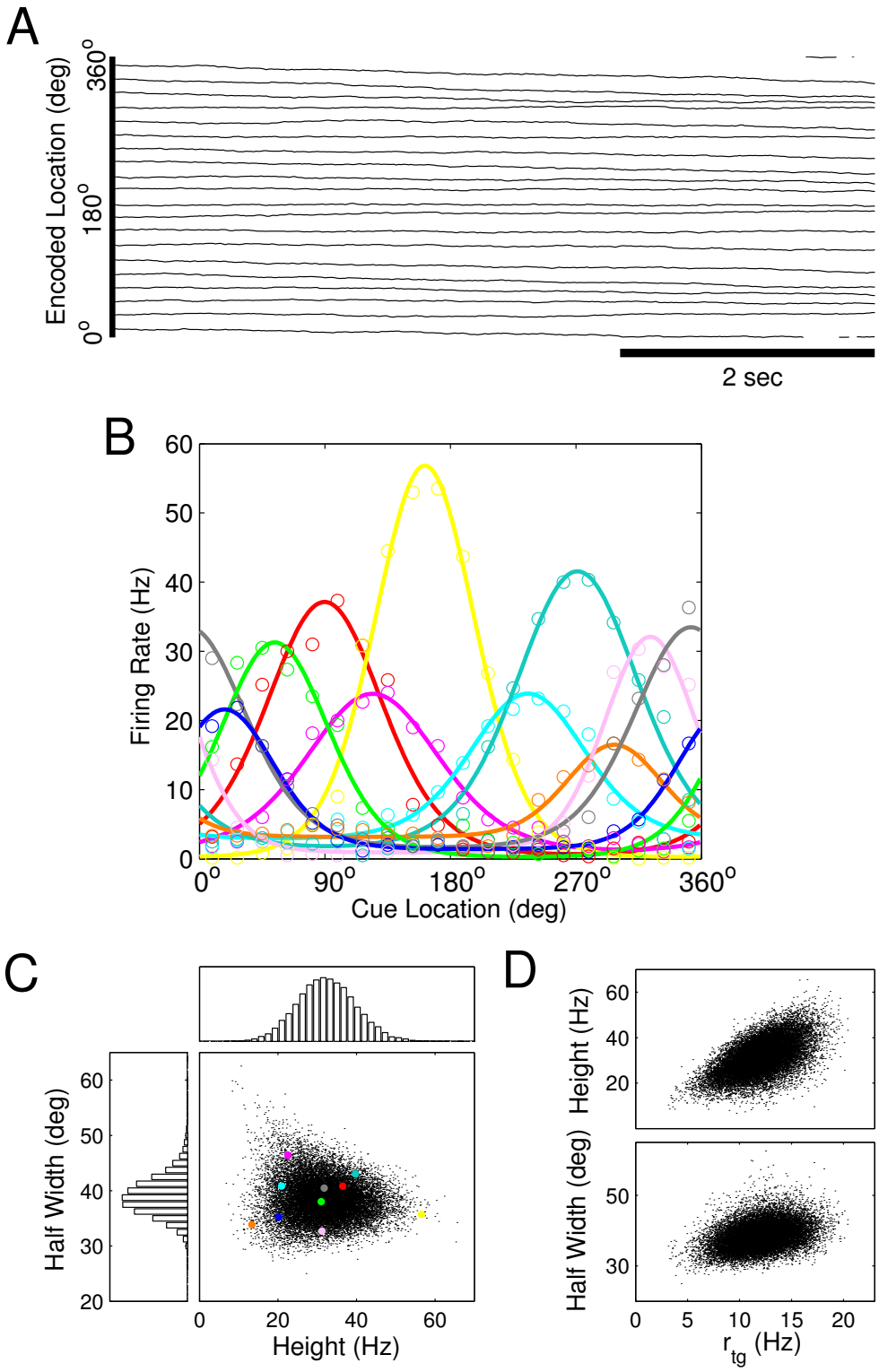
X.-J. Wang ; FIGURE 3



X.-J. Wang ; FIGURE 4



X.-J. Wang ; FIGURE 5



X.-J. Wang ; FIGURE 6