

Controlling activity fluctuations in large, sparsely connected random networks

A C Smith^{†§}, X B Wu[†] and W B Levy^{†‡||}

[†] Department of Neurological Surgery, PO Box 800420, University of Virginia, Charlottesville, VA 22908-0420, USA

[‡] Department of Psychology, University of Virginia, Charlottesville, VA 22908, USA

E-mail: wbl@virginia.edu

Received 25 August 1999

Abstract. Controlling activity in recurrent neural network models of brain regions is essential both to enable effective learning and to reproduce the low activities that exist in some cortical regions such as hippocampal region CA3. Previous studies of sparse, random, recurrent networks constructed with McCulloch–Pitts neurons used probabilistic arguments to set the parameters that control activity. Here, we extend this work by adding an additional, biologically appropriate, parameter to control the magnitude and stability of activity oscillations. The new constant can be considered to be the rest conductance in a shunting model or the threshold when subtractive inhibition is used. This new parameter is critical for large networks run at low activity levels. Importantly, extreme activity fluctuations that act to turn large networks totally on or totally off can now be avoided. We also show how the size of external input activity interacts with this parameter to affect network activity. Then the model based on fixed weights is extended to estimate activities in networks with distributed weights. Because the theory provides accurate control of activity fluctuations, the approach can be used to design a predictable amount of pseudorandomness into deterministic networks. Such nonminimal fluctuations improve learning in simulations trained on the transitive inference problem.

1. Introduction

The control of activity levels in recurrent neural network models has long been an important issue in computational neuroscience. Among recurrent networks, hippocampal area CA3 is particularly of interest, because its high degree of feedback connectivity is thought to be vital for associative learning in general (Marr 1971), and for the processing of polysensory temporal sequences in particular (Levy 1989). Because spatial navigation can be viewed as a form of sequence learning, measurement of the activity level during spatial exploration is of particular interest. In the rat hippocampus, the average firing rate of individual neurons during spatial behaviour remains remarkably constant at a low level, substantially below 1 Hz (Thompson and Best 1989, McNaughton 1996). Although small oscillations exist during awake behaviour in the form of the theta rhythm, large oscillations are not present. This contrasts with sleep where large-oscillation sharp wave bursts are observed corresponding to increased firing rates (requiring a more biophysically detailed model, see Levy *et al* (1998)). In addition to the

[§] Present address: Statistics Research Laboratory, Department of Anesthesia and Critical Care, 55 Fruit Street, Clinics 3, Massachusetts General Hospital, Boston, MA 02114, USA.

^{||} Author to whom correspondence should be sent.

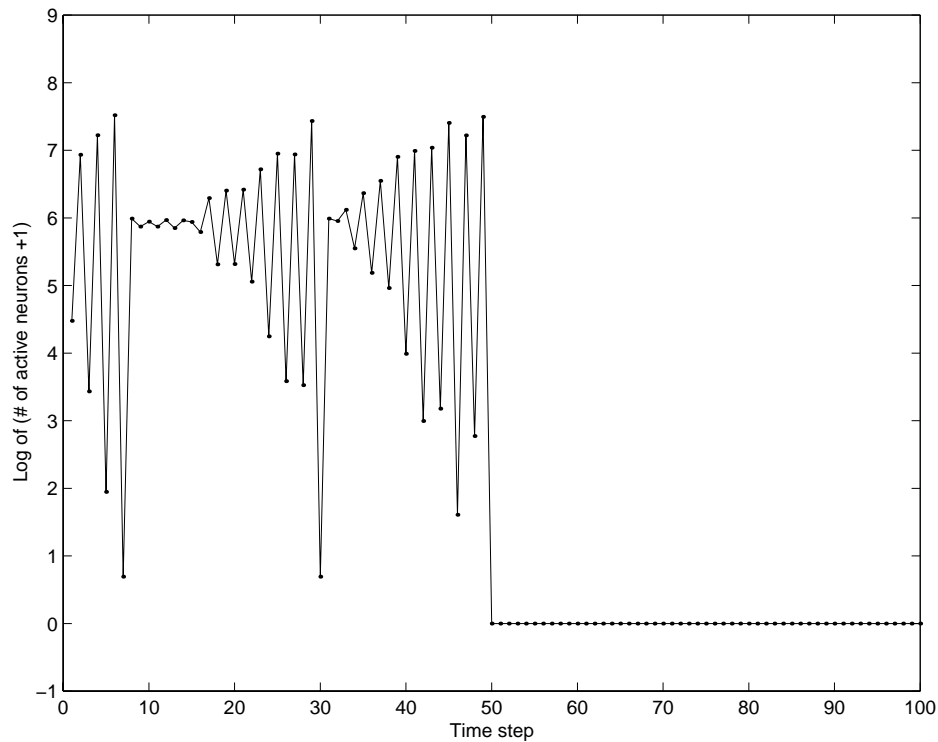


Figure 1. A complete loss of network activity occurs when parameter K_0 is set to zero in large networks. This figure shows fluctuations of activity over time for a random network with 4000 neurons and $K_0 = 0$. Parameter K_R is chosen according to equation (9) so that the mean activity is approximately 10%. Here log of (number of active neurons +1) is plotted against time. Thus, when the network dies (i.e. the number of active neurons reaches zero) the y-axis value is zero. Note that the activity approaches zero many times before a complete loss of network activity at time step 50.

experimental data, computer simulations of hippocampal networks also indicate that tight activity control is crucial for learning temporal sequences (Wu and Levy 1995, Wallenstein and Hasselmo 1997) and for maximizing their storage capacity (Levy and Wu 1996).

The problem we address in this paper is that of increasing the number of neurons in a simulation without introducing uncontrollably large oscillations. As motivation for our current work, we show in figure 1 characteristic activity oscillations in a 4000 neuron network. While the mean network activity over all time steps is held at approximately 0.1, activity oscillations between time steps become uncontrolled and result in a complete loss of network activation by time step 50. On the other hand, simulations at this same activity level using 512 neurons show no such oscillations.

Many models of recurrent networks have been described, ranging in complexity from biophysically realistic compartmental models (Hasselmo *et al* 1995, Lansner and Fransén 1992) to more abstract models (Hopfield 1982, Simmen *et al* 1996). Previous work which models network stability includes that of Wilson and Cowan (1972), where limit-cycle activity of coupled differential equations is considered, and Derrida *et al* (1987), where a Little–Hopfield network is used. More recently, Ginzburg and Sompolinsky (1994) have considered the situation for networks with a high degree of connectivity (unlike the sparse connectivity

of CA3) and Amit *et al* (1997) have analysed networks of integrate and fire neurons. Other recent models that deal with issues of activity and thresholding include those of Herrmann *et al* (1993), Buckingham and Willshaw (1993), Hendin *et al* (1997) and Chechik *et al* (1998). Since only minimal biological complexity is needed to reproduce many hippocampal functions, we investigate a model of hippocampal area CA3 that consists of recurrently connected McCulloch–Pitts neural-like units in which activity is restricted by global feedback inhibition. The connectivity of the model is sparse and asymmetric, consistent with known hippocampal anatomy (Miles and Wong 1986), in contrast to the architecture of the classical Hopfield model. CA3 had been found experimentally to be dominated by excitatory pyramidal cells with small feedback connectivity and input pathways from the entorhinal cortex and the dentate gyrus (see e.g. Tamamaki and Nojyo 1990, 1993, Li *et al* 1994, Amaral and Witter 1995). Our model, described in the next section, captures those features.

Previous simulations by our laboratory have demonstrated that this type of network is capable of a variety of dynamics including fixed-point, limit-cycle, and effectively aperiodic behaviours (Minai and Levy 1993b). Additionally, this research proposed an analytical form to predict the average network activity level based on the threshold, synaptic strengths, and fractional connectivity. However, the present study shows that this previous model is incapable of maintaining biologically plausible activity levels when the number of neurons is large, and at the same time, it presents a biologically minimal solution to this failure.

After describing the new model with its additional term, we show how a constant inhibitory term can be parametrized to control the mean network activity level and the average size of activity fluctuations. The inclusion of this constant becomes increasingly important for larger networks. In fact, we show (see the appendix) that large networks are guaranteed to be ill-behaved without such a term. Next, with an eye toward the practical use of the network in learning situations, the average network activity level is studied as a function of the number of active external input units. The analytical results are then further extended to the case in which strengths of the synapses are uniformly distributed, rather than all being equal. In each case, analytical predictions are compared with simulations. In the final section, we demonstrate how this theory can be applied to control simulations which learn. The main contribution of this paper is to provide an analytical expression that predicts a network's average activity and its fluctuations when this network is too large for explicit simulation.

2. Network model

The fixed fan-in network considered here consists of n binary (0/1) neurons, where each of these neurons is randomly connected to n_c other neurons. Given that the number of active neurons at time t is $m(t)$ and that $z_j(t)$ is 0 or 1 depending on whether neuron j fires or not, the net excitation representing the voltage of the cell, $y_i(t)$, is given by

$$y_i(t) = \frac{\sum_{j=1}^n w_{ij} c_{ij} z_j(t-1)}{\sum_{j=1}^n w_{ij} c_{ij} z_j(t-1) + K_R m(t-1) + K_0 + K_I \sum_{j=0}^n x_j(t)} \quad (1)$$

where w_{ij} represents the weight value between neurons i and j , and c_{ij} is a 0/1 variable indicating connection from neuron j to i . The term $\sum_{j=1}^n w_{ij} c_{ij} z_j(t-1)$ represents the excitatory synaptic conductance for the i th neuron. Parameters K_R and K_I are constants used to control feedback and feedforward inhibitions respectively. A binary external input to each neuron i at time t is given by $x_i(t)$. The neuron fires (i.e. $z_i(t) = 1$) if either $x_i(t) = 1$ or if $y_i(t) \geq \theta$ where θ is a fixed threshold, taken here to be 0.5 in the simulations. This equation is similar to the equations considered in Minai and Levy (1993a, b), except here the constant K_0 has been added to the denominator. Although we use global shunting inhibitory feedback,

which receives extensive use in other network models (e.g. Grossberg 1988, Furman 1965, Carandini and Heeger 1994), the results are easily applied to subtractive inhibition (Minai and Levy 1993a).

The external input model described here is a simplification of what is likely to be the case. Instead of collapsing the entorhinal cortex (EC) and dentate gyrus (DG) inputs to a single binary input, a more sophisticated model would assume a moderate level of EC-based excitation spread over the whole network with a small set of CA3 neurons getting a DG input. We have assumed the combination of the DG and EC inputs would lead to cell firing. Furthermore, we hypothesize that only at the highest levels of inhibition would this externally triggered cell firing be inhibited. Thus, the total external input contributes a constant amount to the activity and inhibition at each time step. Activity levels of the model are only perturbed by fluctuations due to recurrent excitation and recurrent inhibition.

Given the existence of a non-zero threshold and assuming all weights are equal to a constant w , equation (1) can be rearranged in the subtractive inhibition format (Minai and Levy 1993a) such that

$$z_i(t) = \begin{cases} 1 & \text{if } \sum_{j=1}^n c_{ij}z_j(t-1) - \alpha m(t-1) - \beta - \gamma m_e \geq 0 \\ 0 & \text{if otherwise} \end{cases} \quad (2)$$

where $m_e = \sum_{j=1}^n x_j$ with constant net input excitation and α , β and γ are given by $K_R\theta/(w(1-\theta))$, $K_0\theta/(w(1-\theta))$ and $K_I\theta/(w(1-\theta))$, respectively.

Following similar arguments to those in Minai and Levy (1993a, b), the firing condition for non-input neuron i at time t can be written

$$\sum_{j=1}^n c_{ij}z_j(t-1) \geq \alpha m(t-1) + \beta + \gamma m_e. \quad (3)$$

Note that the effect of changing β in equation (3) is the same as changing the constant product γm_e . Denoting $M_1 = \alpha m(t-1) + \beta + \gamma m_e$, the average firing probability for a non-input neuron based on the previous time step's activity is given by

$$\begin{aligned} P(z_i(t) = 1 | x_i(t) = 0) &= P(M_1 \text{ or more active inputs to } i | x_i(t) = 0) \\ &\equiv \rho(m(t-1)) \\ &= \sum_{k=\lceil M_1 \rceil}^{n_c} \frac{\binom{m(t-1)}{k} \binom{n-m(t-1)}{n_c-k}}{\binom{n}{n_c}}. \end{aligned} \quad (4)$$

Because $m(t-1)$ and $(n-m(t-1))$ are reasonably large and $n_c/n < 0.1$, we can approximate the hypergeometric distribution by the binomial with probability, $p \equiv n_c/n$. Additionally, when $m(t-1)p > 5$ (Milton and Arnold 1990), we use the normal distribution. Hence, we write

$$\rho(m(t-1)) \approx \sum_{k=\lceil M_1 \rceil}^{n_c} \binom{m(t-1)}{k} p^k (1-p)^{m(t-1)-k} \quad (5)$$

$$\approx \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{1}{\sqrt{2}} \frac{M_1 - m(t-1)p}{\sqrt{m(t-1)p(1-p)}} \right) \right]. \quad (6)$$

Errors related to the above approximations are discussed in the final section of the results. Using the assumption that neurons fire independently (Minai and Levy 1993a, b), the activity of the network is binomially distributed. As a result, activity at time t can be related to activity at time $t-1$ by the following equation where the first and third terms are the mean and standard deviation of the binomial distribution:

$$m(t) = (n - m_e)\rho(m(t-1)) + m_e + O\left(\sqrt{(n - m_e)\rho(1 - \rho)}\right). \quad (7)$$

Re-writing in nondimensional form, we can calculate fractional activity, $r(t) = m(t)/n$, from

$$r(t) = \frac{(n - m_e)}{n} \rho(nr(t - 1)) + \frac{m_e}{n} + O\left(\frac{\sqrt{(n - m_e)\rho(1 - \rho)}}{n}\right). \quad (8)$$

With β and m_e both zero in our analysis, equation (7) reduces to the equation studied in Minai and Levy. The networks resulting from this simplification exhibit various effectively aperiodic, cyclic and fixed-point behaviours depending on the magnitude of α , a parameter which is controlled by varying K_R (Minai and Levy 1993a, b).

2.1. Choice of parameters p and w

In our simulations, we take network connectivity to be 0.1, independent of network size. We estimate the hippocampus' connectivity to be approximately 3–5% (based on 20 000 inputs per CA3 neuron of the 600 000 CA3 neurons counting bilaterally) and take a higher value (i.e. 10%) to reflect the higher degree of random connectivity within a thin slice of CA3 compared with the overall connectivity. The synaptic weight, w , is taken arbitrarily to be 0.4 and is also independent of network size. Parameters were selected so that a neuron will require 25 active neuron connections to meet the firing criterion, a value which is arguably consistent with physiological data (Andersen *et al* 1994).

3. Control of oscillations

Both K_R and K_0 can be used to control activity levels and the magnitude of oscillations. This is done for a given absolute activity, a , the fixed point described by the desired number of active neurons, by solving simultaneously for the unknowns K_0 and K_R in the equations

$$a = (n - m_e)\rho(a) + m_e \quad (9)$$

and

$$g = \frac{dm(t)}{dm(t - 1)} = \frac{d}{dm(t - 1)}((n - m_e)\rho(m(t - 1))) \Big|_{m(t-1)=a} \quad (10)$$

where g represents the gradient of the curve for $m(t)$ against $m(t - 1)$ at the fixed point a . Assuming the normal approximation (equation (6)), the derivative of the cumulative normal is just the normal so that equation (10) at $m(t - 1) = a$ becomes

$$g = (n - m_e) \frac{(p - \alpha)a + \beta + \gamma m_e}{2a\sqrt{2\pi ap(1 - p)}} \exp\left(\frac{-((\alpha - p)a + \beta + \gamma m_e)^2}{2p(1 - p)a}\right). \quad (11)$$

For stability at the fixed point, $|g|$ must be less than unity (see, e.g., Hale and Koçak 1991). Moreover, oscillations will tend to be minimal when g is close to zero. Solutions for equations (9) and (10) can be calculated in a few seconds using a program such as Mathematica (Wolfram 1996).

Cobweb plots (figure 2) show how equations (9) and (10) are employed to control activity level and gradient. In each graph (figures 2(a)–(f)), equation (7) is plotted (thick grey line) using K_0 and K_R values calculated to produce a fixed point activity of 10% ($a = 200$ neurons) and by specifying a particular gradient ($g = -1.25, -1, -0.5, 0.5, 1$ and 1.25 in (a)–(f) respectively). The network parameters used are shown in the figure legend. In each case, we choose a starting value for $m(t - 1)$ of 160 neurons which corresponds to a point that is approximately three standard deviations (i.e. $3\sqrt{np(1 - p)}$) below the desired activity fixed point. This starting point lies at the lower extremes of possible starting values given the network activity is assumed to be normally distributed. As expected, for gradients of

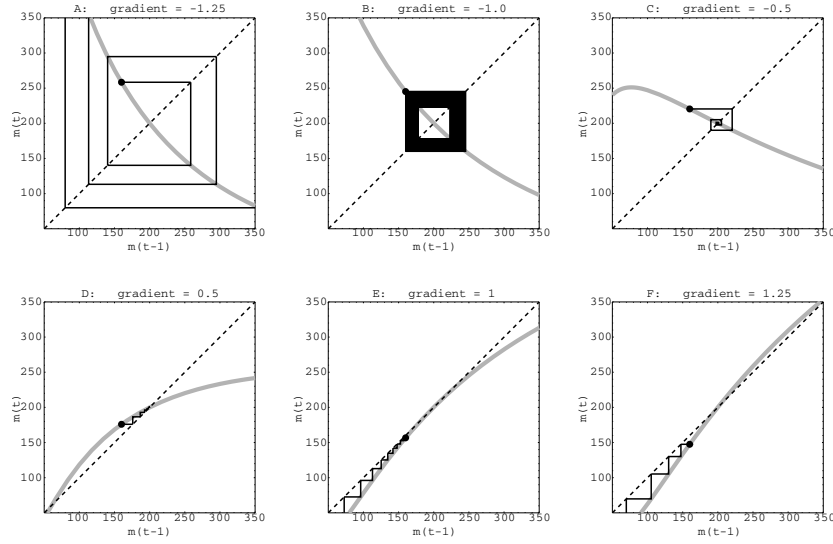


Figure 2. Graphs of $m(t)$ against $m(t-1)$ (thick grey curve) (equation (7)) using values of K_0 and K_R computed from equations (9) and (10). Here, activity, a , is assumed to be 10% (= 200 neurons) and different crossing gradients, g , shown above each graph, are used. Superimposed on each plot are cobweb convergence/divergence diagrams showing the typical convergence characteristics for equation (7) when the starting point (\bullet) is 160 neurons. The iteration procedure is plotted for 500 steps. Network parameters: $n = 2000$, $p = 0.1$, $w = 0.4$, $\theta = 0.5$, $m_e = 0$. Note that (E) and (F) converge to zero, (A) oscillates between high and low activity levels, (B) oscillates around the desired level but substantially far from it, while (C) and (D) converge to the desired level.

-1.25 (figure 2(a)) and 1.25 (figure 2(f)) and a starting point of 160, repeated iterations of equation (7) cause the activity values to diverge from the fixed point at $m(t-1) = 200$ neurons. In figure 2(a), the solution oscillates with increasing amplitude about the fixed point. In figure 2(f) the values of $m(t)$ tend to drop to zero. When $|g| = 1$ (figures 2(b) and (c)) the possibility exists for the solutions to converge or diverge. Figures 2(c) and (d) show, for $|g| < 1$, how the solution will converge to the fixed point. However, for all practical purposes, at each iteration, a new value for $m(t)$ is randomly generated from within an $O(\sqrt{np(1-p)})$ window around $m(t-1)$ (Amari 1972, Minai and Levy 1993a).

3.1. Comparison of theory with simulated results

Figure 3 illustrates the step-by-step activity fluctuations as a function of the pair (K_0, K_R) . Specifically, figure 3 shows predicted activity fluctuations over time for simulated networks with a mean activity of 10%. The values for K_0 and K_R (shown below the figure) are chosen so that g (shown above the figure) increases from -1.25 to 1.25 . The networks have identical numbers of neurons, connections, initial weights and threshold as the network in figure 2. For example, when $g = -0.75$, the solutions oscillate about the fixed point in a similar way as shown in figure 2(c). It is evident that the simulated network behaviour mimics the behaviour we would expect from the theoretical cobweb plots in figure 2. Note that a network modelled without K_0 would have stability properties somewhere between $g = -1.25$ and -1 since $K_0 = 0$ implies $-1.25 \leq g \leq 1.0$. Therefore networks with large numbers of neurons (2000 or more) will be highly unstable (see the discussion concerning our experience on this point and the appendix which includes an analysis of the implications of $K_0 = 0$). This instability

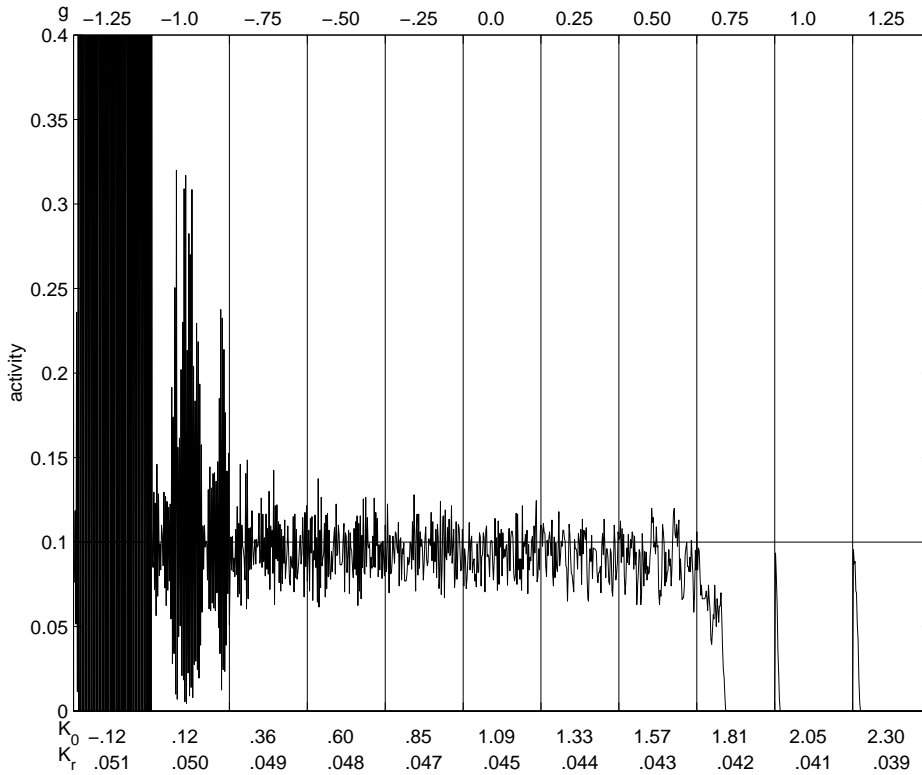


Figure 3. Graphs showing activity fluctuations for mean activity set at 10% (=200 neurons) and increasing crossing gradients. Each segment of the graph shows a typical plot of activity over time (100 time steps) for the indicated network parameters K_R and K_0 . The calculated gradient, g , for these parameters at 10% activity is noted above each segment of the graph. In each segment, the initializing activity is $m(t) = 160$ i.e. 0.08. The network consists of 2000 neurons with a connection probability of 10%. Initial weights are set at $w = 0.4$ and threshold is set at $\theta = 0.5$.

at $K_0 = 0$ implies a future role for $K_0 > 0$ when diverging activity must be avoided.

Figure 4 summarizes the results from simulations where α is fixed and g is varied between runs. Figure 4(a) shows the calculated relationship between K_R and K_0 for different constant activities at the gradients indicated in figure 4(b). (The points in figures 4(a) and (b) are just mapped sequentially between lines of identical activity.) These theoretical isoactivity curves in figure 4(a) are linear with slope $1/m(t - 1)$ when all other parameters are kept constant since equation (23) will be unchanged when M_1 (i.e. $\alpha m(t - 1) + \beta$ given $m_e = 0$) remains constant. At the same time, the gradient, g , in equation (10) is linearly related to α which explains the equal spacing of points on the isoactivity curves in figure 4(a) as the values of g were incremented by 0.1. This linearity is expected as we demonstrate using the hyperbolic tangent approximation in the appendix.

Results from simulations are summarized in figures 4(b) and (c). Figure 4(b) demonstrates that simulations can maintain a desired mean activity using the calculated values of K_0 and K_R . Note that the similarity between simulated and preassigned activities improves as activity levels decrease. However, as gradient g increases, the accuracy at 20% activity decreases because the mean of the fixed point is not necessarily the same as the fixed point of the mean. Figure 4(c) shows the coefficient of variation ($CV = \text{standard deviation}/\text{mean}$) of the activity

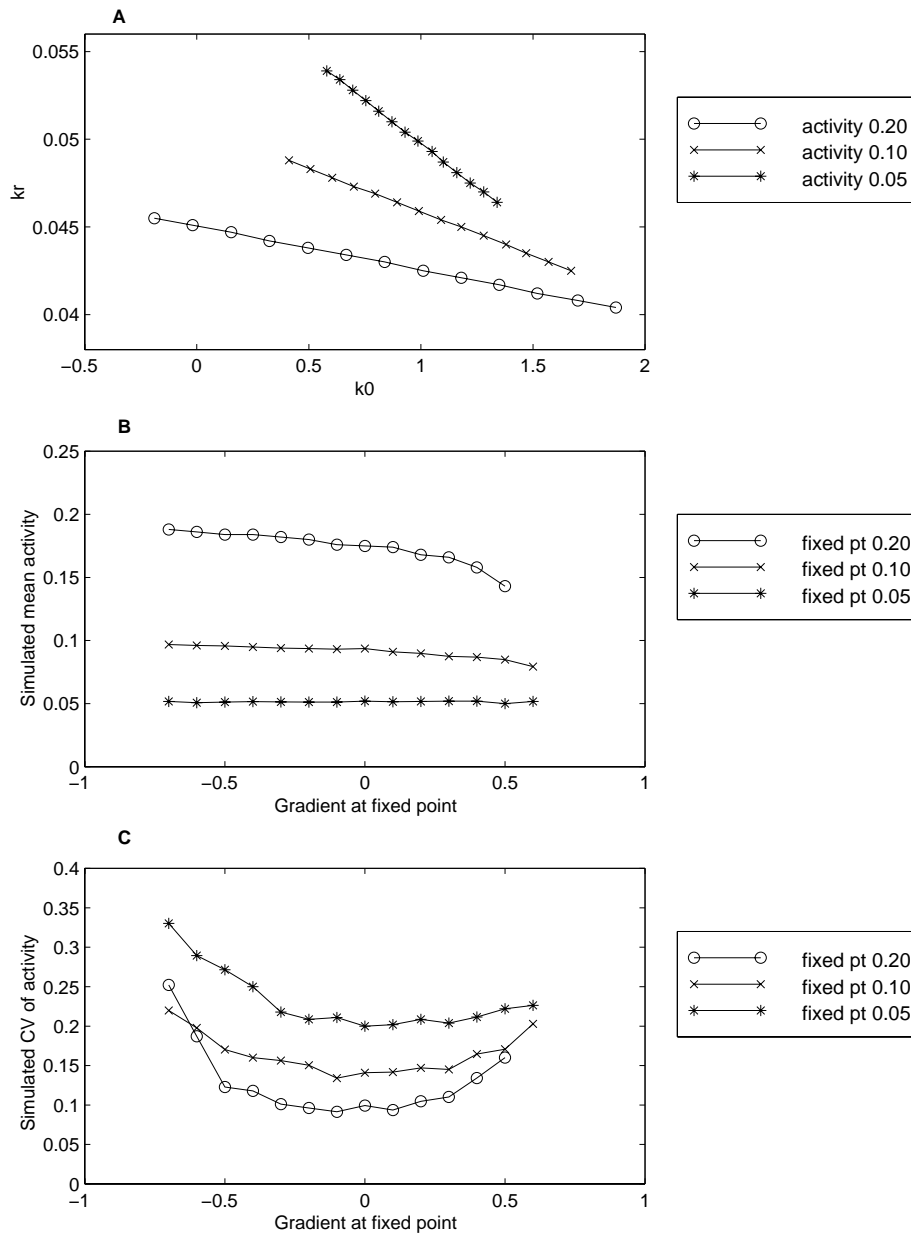


Figure 4. Summary of simulated network statistics using parameters from theory. (A) Values of K_R and K_0 calculated using equations (9) and (10) for three different activity levels ($a = 0.05, 0.1$ and 0.2) and different gradients (-0.7 to 0.6). The points in (A) map sequentially (relative to their respective symbols) to the points in (B). That is, (B) shows nearly constant activity for all (K_0, K_R) pairs except at the larger activities and gradients. As K_0 increases, the gradient at the fixed point increases. (B) Mean activity averaged across several randomly connected networks of 2000 neurons using values for K_R and K_0 chosen to fix network activity at 5, 10 and 20% for varying gradients at the fixed point. (C) Coefficient of variation (CV) of activity of these random networks using values for K_R and K_0 chosen to fix network activities at 5, 10 and 20% for varying gradients at the fixed point. Note that a zero gradient minimizes the variation and that CV is greater at lower fixed-point activities.

and demonstrates that fluctuations are indeed minimized when g is approximately zero for all three activities. Note that the CV is largest when activity is lowest. (In figures 4(b) and (c), we only show simulated results for gradients in the interval $[-0.7, 0.6]$ instead of $[-1, 1]$. Networks with parameters implying a g close to ± 1 were unstable because of the discrete rather than continuous nature of the actual return map of $m(t)$ against $m(t - 1)$.)

3.2. Scale-up to larger networks

Having shown that the oscillations are minimized when $g = 0$, we are particularly interested now in how an increase in network size affects the magnitude of the network oscillations when $g = 0$.

An interesting feature of equation (8) is that, for a given fractional activity, increasing n does not affect the shape of the curve of $r(t)$ against $r(t - 1)$. (This means that we would expect identical oscillatory behaviour about the fixed point for the same starting values of fractional activity.) However, as n is increased, K_R must be decreased and K_0 increased in order to achieve the same fractional activity and gradient.

At the same time, this increase in n tends to decrease the magnitude of oscillations by decreasing the size of the $O(1/\sqrt{n})$ region (equation (8)) within which the activity values tend to be restricted. Figure 5 shows simulated results for simulations run at different activities with the number of neurons increasing from 500 to 4000. All the networks are parametrized such that $g = 0$ at the fixed point. The changes in standard deviation of the activity with number of neurons for four different activities are shown in figure 5(a). As implied by equation (8) with increasing n , this standard deviation decreases. Also, this standard deviation decreases as activity decreases. To confirm the prediction of equation (8), that the size of the window around the field varies as $O(1/\sqrt{n})$, we plot in figure 5(b) the coefficient of variation of activity against $1/\sqrt{n}$ for four different activity levels where the same zero value for gradient at the fixed point is maintained in each case. As expected, an approximate linear relationship exists for all activities shown although for the low activity/small number of neurons case the dependence is not quite linear probably due to the use of the normal approximation to the binomial distribution.

An important point from the above observations is that as n increases without the constant K_0 in equation (1), the network activity oscillations will become unstable as the gradient of equation (7) diminishes below -1 . A rat hippocampal CA3 is typically estimated to have around 300 000 pyramidal cells and a fractional connectivity as low as 0.02 (Amaral *et al* 1990). In this case, without K_0 and assuming $w = 0.4$, such a simulation would not be stable for activities below $\approx 12\%$ (see the appendix for a more general analytical estimation of this value). Because this value is much higher than the $< 1\%$ activity in the rat hippocampus estimated by Thompson and Best (1989), a positive-valued K_0 is important for achieving biological relevance.

In summary, as n is increased, provided the appropriate K_0 and K_R values are obtained such that $g \approx 0$, there appears to be no limit on the size of network that can be activity-controlled using this method. Figure 5 indicates that the coefficient of variation of activity decreases as we increase n and thus networks at low activities become increasingly manageable when K_0 is present.

4. Variation of input size

Equation (7) can predict activity and its stability for different values of external activity, $\sum_{j=0}^n x_j(t)$. This is significant since external input into CA3 is a reality, and the relative

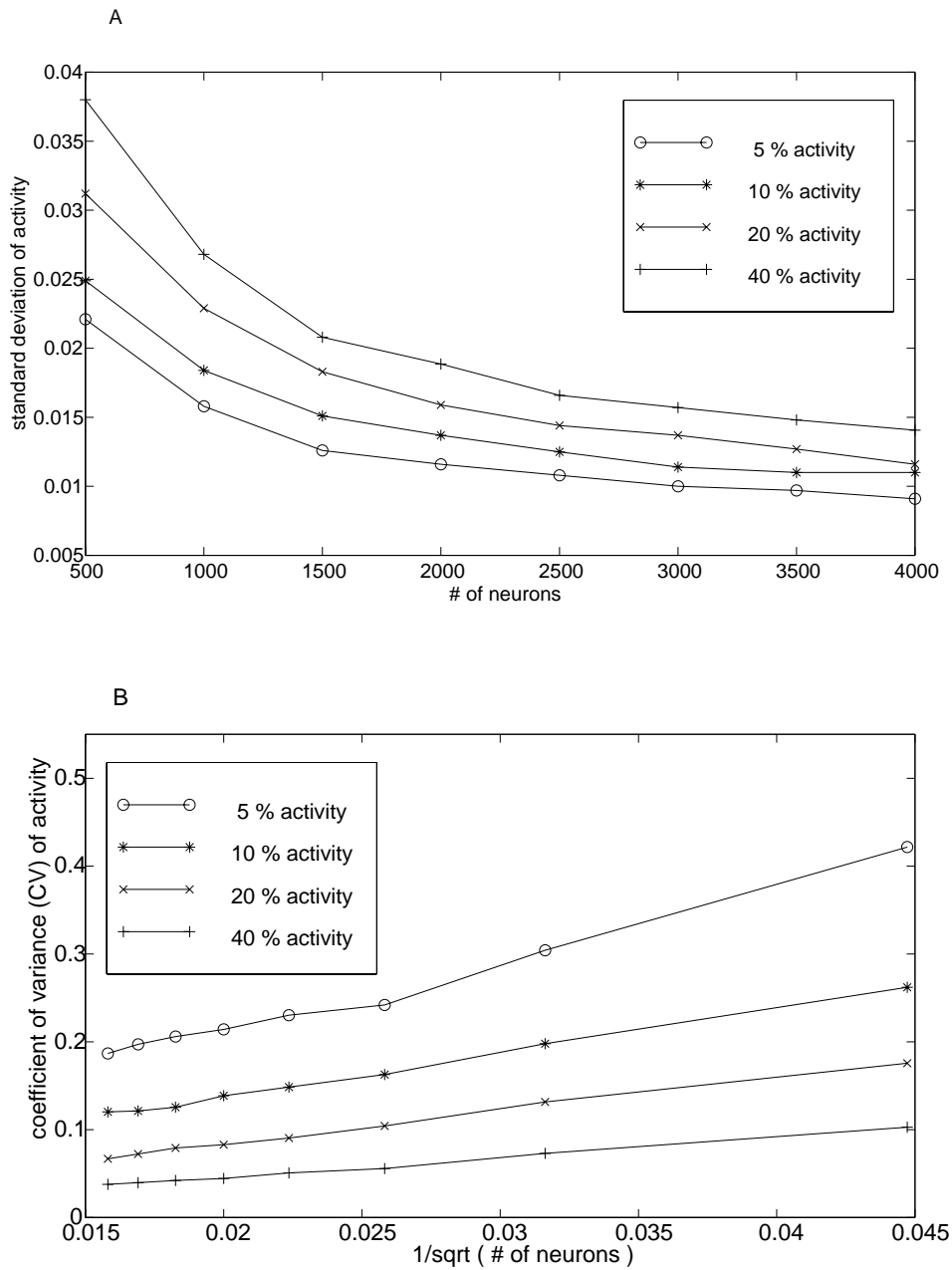


Figure 5. (A) The standard deviation of activity decreases as the number of neurons increases for each of the four indicated activity levels. Each point has parameters K_0 and K_R chosen such that $g = 0$. (B) Approximate linearization of the curves in (A). This is achieved by replotting the above data as coefficient of variation of the activity against $(\text{number of neurons})^{-1/2}$. Shown are results at four different activity levels. This linearizing transformation is suggested by equation (8).

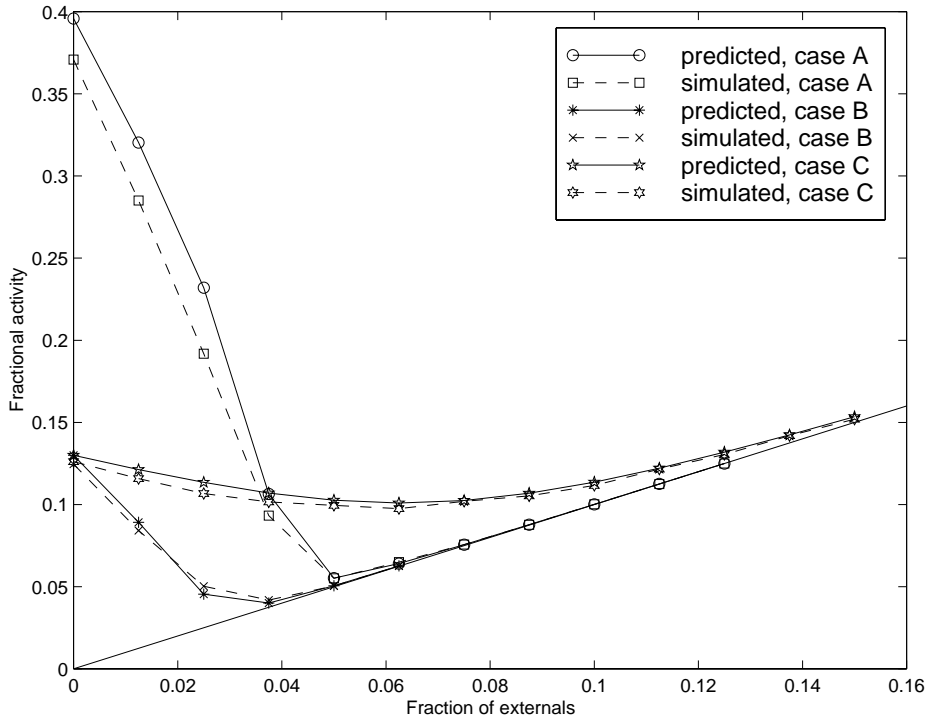


Figure 6. Comparison between theoretical and simulated prediction of average activity for networks with increasing ratio of external inputs to number of neurons. Three cases are shown corresponding to different pairs of values of K_I and K_R . Network parameters: $n = 2000$, $p = 0.1$, $\theta = 0.5$, $K_0 = 0.1$. Case A has $K_R = 0.041$ and $K_I = 0.03$. Case B has $K_R = 0.048$ and $K_I = 0.03$. Case C has $K_R = 0.048$ and $K_I = 0.01$. Note that as the fraction of externals increases the fractional activity tends to the line of slope 1 (shown as a solid curve on the graph).

size of this input is hypothesized to be an important variable (Levy 1989, 1996, Polyn *et al*, at press). Fortunately, changing this variable has little effect on predicting simulations (figure 6). For three different pairs of K_R and K_I , the predicted activities as a function of $\sum_{j=0}^n x_j(t)$ are very close to simulated results. In all three cases, the total fractional activity drops initially as the fraction of active externals increases. This drop is caused by the ‘feedforward’ inhibitory effect of the external neurons on the values of excitation, y_i . Eventually, as the external neurons account for more and more of the network activity, feedforward inhibition prevents other neurons from firing, and the fractional activity tends towards the fraction of active externals.

5. Implementation with a distribution of synaptic weights and estimation of model accuracy

Up to now the analyses assumed constant weights, but the results can be extended to predict activities in the case of uniformly distributed weights. This case is of interest because approximately uniform distributions of weights occur in networks filled to capacity by learning simple sequences (Amarasingham and Levy 1998).

Defining $M_2 = K_R \theta m(t-1)/(1-\theta) + K_0 \theta/(1-\theta) + K_I \theta m_e/(1-\theta)$, equation (4) can be replaced by

P (neuron firing)

$$= \sum_{k=1}^{n_c} P(k \text{ active input neurons}) \times P(k \text{ active neurons can fire the neuron}) \quad (12)$$

$$= \sum_{k=1}^{n_c} P(k \text{ active input neurons}) \times P(k \text{ samples from weights have sum } \geq M_2) \quad (13)$$

$$= \sum_{k=1}^{n_c} \frac{\binom{m(t-1)}{k} \binom{n-m(t-1)}{n_c-k}}{\binom{n}{n_c}} \times P(k \text{ samples from weights have sum } \geq M_2). \quad (14)$$

Using the normal approximation for the hypergeometric term and assuming independence among the uniformly distributed weights, we can invoke the central limit theorem for the second multiplier in equation (14) and then write

P (neuron firing)

$$\approx \int_{x=0}^{n_c} N_x(m(t-1)p, m(t-1)p(1-p)) \times \int_{y=M_2}^{\infty} N_y(x\mu, x\sigma^2) dy dx \quad (15)$$

where N_x and N_y are probability density functions for the normal distribution in x and y , respectively, each taking parameters mean and variance. Here, μ and σ^2 are the mean and variance of the uniform weight distribution.

Some results found using this approach are summarized in tables 1 and 2. First, we consider the magnitude of the errors between calculated and simulated networks when the assumption of constant weights is made (table 1). Here, four cases are considered where n increases and fractional activity simultaneously decreases such that the number of neurons firing at each time step, $m(t-1)$, is kept constant. That is, the fractional activity is decreasing in a way that is increasing proportional to the number of neurons. The idea is to assess the prediction of network activity as we move to the more biologically plausible low activities and the high number of neurons encountered in the mammalian hippocampus. In each case shown in table 1, the values of K_R and K_0 have been computed using equations (9) and (10) to give pre-specified activities and zero gradients. Columns 2 and 3 show the simulated network activities using these calculated values of K_R and K_0 . Column 2 shows the case when all initial weights are set to 0.4. Column 3 shows the case when initial weights are uniformly drawn from the interval [0.1, 0.7]. Corresponding relative errors 1 and 2 shown in columns 5 and 6 show how the assumption of constant weights produces errors which increase in magnitude as n increases. Errors found by employing equation (15) are smaller (see column 6), but also indicate that the theory overpredicts activity as n increases.

Note that as n increases, the value of M_2 increases so we are essentially integrating a smaller portion of the right-hand tail of the normal distribution. As this happens, the use of the normal approximation of the hypergeometric will become insufficiently accurate for our purposes, but more exact calculations are computationally feasible. Table 2 shows a comparison of simulated activity with activity calculated using the normal approximation (equation (15)) as before and also using the hypergeometric term itself (equation (14)). The values shown for K_r and K_0

Table 1. Comparison of simulated activities with constant weight w (column 3) and uniformly distributed (column 2) weights w_{ij} with activity predicted assuming $w = 0.4$ (column 4) and equation (7). The table shows K_R and K_0 values calculated such that $g = 0$ for decreasing fractional activity levels and increasing numbers of neurons such that $m(t - 1)$ is kept constant in each case. The constant weights are set to a value of 0.4 and the distributed weights have mean 0.4, distributed uniformly over the interval $[0.1, 0.7]$. Relative error between the predicted activity and the simulated activity with a weight distribution (relative error 1) is shown in column 5. Column 6 shows relative error 2 which is the equivalent comparison when constant weights are assumed. The simulated activity (columns 2 and 3) given for each n is the mean activity calculated from five different random networks. For each set of five networks, the relative error (defined here as $100 \times (\text{maximum activity} - \text{minimum activity})/\text{minimum activity}$) was found to be less than 5%. Network parameters: $\theta = 0.5$, $p = 0.1$.

Number of neurons	Simulated activity with distribution	Simulated activity w/o distribution	Predicted activity	Relative error (%) 1	Relative error (%) 2	K_R	K_0
500	0.196	0.175	0.2	2	14	0.04505	0.5050
1000	0.113	0.092	0.1	-12	9	0.04769	0.7689
2000	0.067	0.052	0.05	-25	-4	0.04987	0.9869
4000	0.038	0.028	0.025	-34	-11	0.05176	1.1760

Table 2. Comparison of activity calculated using the normal approximation to the hypergeometric in equation (15) with the solution estimated by using the hypergeometric directly. The simulated activity (column 2) given for each choice of K_R and K_0 is the mean activity calculated from five different random networks. For each set of five networks, the relative error (defined here as $100 \times (\text{maximum activity} - \text{minimum activity})/\text{minimum activity}$) was found to be less than 3%. As the networks grow in size, note the linear trend for the increasingly poor percentage error of the calculations based on the Gaussian approximation. On the other hand, the calculations based on the hypergeometric seem satisfactory and appropriate for even larger simulations.

n	Simulated activity	Calculated activity with Gaussian	Calculated activity with hypergeometric	K_R	K_0
500	0.1784	0.2005(12%)	0.1815(2%)	0.04553	0.5467
1000	0.0958	0.1018(6%)	0.0960(0%)	0.04830	0.8671
2000	0.0521	0.0515(-1%)	0.0525(1%)	0.05060	1.141
4000	0.0274	0.0250(-9%)	0.0276(1%)	0.05290	1.380
500	0.0539	0.0501(-7%)	0.0544(1%)	0.0614	0.0600
4000	0.0501	0.0500(0%)	0.0500(0%)	0.04767	1.586

for the first four cases were chosen by trial and error such that the activities calculated with equation (15) approximately maintained the network activity, $m(t - 1)$, at 100 neurons as in table 1. The last two cases show the situation where network fractional activity is constant at 5%. The fact that differences between the activity predicted using the hypergeometric term and the simulated network are very small indicates the accuracy of the analysis up to the use of the normal approximation. However, large networks require a more accurate estimate than that provided by the normal distribution. The disadvantage of just using the hypergeometric is that it is slow to compute, and it is not invertible. Replacing the hypergeometric tail probabilities by a polynomial approximation of a cumulative Poisson distribution seems like a sensible course when n is large and the activity region of interest is low ($<5\%$) and narrow (less than an order of magnitude).

6. Gradient as a variable that affects learning

The replacement of inhibitory parameters K_0 and K_R by the dependent variables activity a and gradient g seems useful for at least three reasons. First, activity is readily measurable in biological experiments. Second, gradient allows us to control the variation of activity in a systematic manner while keeping activity itself constant. Third, our previous results show that some variance in activity is good for network performance (Levy and Wu 1999). Thus, by introducing gradient, there is now a rational way to control activity variation from small to large.

Using networks with similar architecture to the random network described above but with weights which vary according to a postsynaptic Hebbian learning rule, simulations were trained on the transitive inference (TI) problem. The weights are modified according to the equation $\Delta w_{ij}(t) = \epsilon z_j(t)(z_i(t-1) - w_{ij}(t-1))$ where ϵ is a small positive constant. The simulations used a variety of K_R and K_0 values, but all paired parametrizations kept activity at approximately 10% while the gradient value was varied.

In the TI problem, five atomic stimuli (e.g. $A-E$) are used (Dusek and Eichenbaum 1997, Levy and Wu 1997, Wu and Levy 1998). When A and B are together, A is the right answer ($A > B$). When B and C are together, B is the right answer ($B > C$) and so on. Thus, subjects and networks are taught $A > B$, $B > C$, $C > D$, and $D > E$. The critical test of TI after learning these four pairs is the response to $B?D$.

As part of our hypothesis (Levy 1989, 1996) of the computation performed by the hippocampus, we have cast this problem as sequence prediction, whereby sequences in the form stimulus–decision–reinforcement are presented to the network using a staged learning paradigm (see Dusek and Eichenbaum 1997). Hence, we teach the network eight sequences:

$$\begin{array}{ll} (AB)(AB)(AB)aaa + + +, & (AB)(AB)(AB)bbb - - -, \\ (BC)(BC)(BC)bbb + + +, & (BC)(BC)(BC)ccc - - -, \\ (CD)(CD)(CD)ccc + + +, & (CD)(CD)(CD)ddd - - -, \\ (DE)(DE)(DE)ddd + + +, & (DE)(DE)(DE)eee - - -, \end{array}$$

where lower case letters correspond to decisions (e.g. ‘ aaa ’ means choose A) and $+/-$ symbols correspond to the desirable and undesirable outcomes. We turn on part of the goal ‘ $+$ ’ to induce the model to follow the path in state space that leads to the right answer. This method is consistent with the concept that most problem solving is performed with at least a partial idea of the correct answer (Levy *et al* 1995, Levy and Wu 1997). During testing, the network is initially randomized, given a stimulus pair for three time steps all the while turning on a fraction of the $+$ input code. Then the network is allowed to run, generating its own sequence of cell firings. We decode this sequence to discover the network’s prediction. Decoding is done using a cosine comparison test where we compare each state during testing to each state on the last trial of training. For example, when AB is presented to the network with a fraction of $+$ neurons turned on during testing, we see if the network decides that ‘ a ’ or ‘ b ’ is the correct decision portion of the sequence.

For each simulation, good performance is defined as in the behavioural experiments (Dusek and Eichenbaum 1997), i.e., 80% or more correct responses at the end of training for all four learned discriminations (i.e. A versus B , B versus C , etc) and the two inferred discriminations (i.e. B versus D and A versus E). Thus, a simulation must correctly predict responses to all six sequences on 80% or more trials to be counted as a good learner.

Figure 7 shows the relationship between performance and the gradient setting. Apparently performance is optimized by a slightly negative gradient. For the same simulations, we also show measured variance in activity within each trial. Two points are worth noting. First,

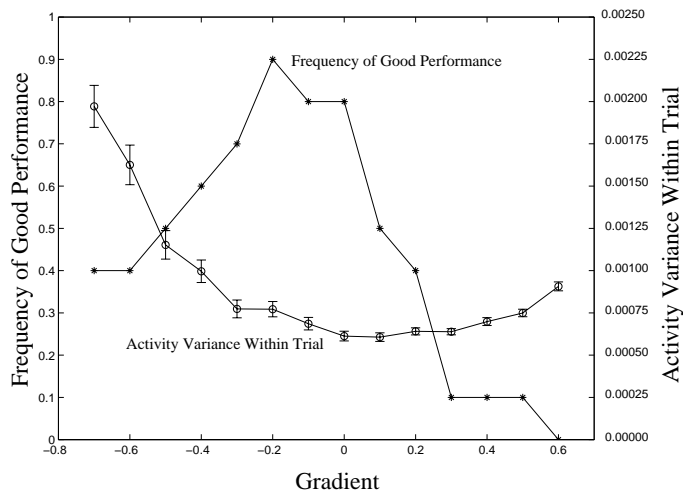


Figure 7. Frequency of good performance as a function of gradient (*) and simulated network activity variance as a function of gradient (o). The activity level was kept approximately constant at 10% by adjusting K_R and K_0 corresponding to different gradients while other network parameters were fixed ($n = 1024$, $\theta = 0.5$, $K_I = 0.018$). Note that the best performance is achieved at a slightly negative gradient instead of at zero gradient where activity fluctuations are minimized. Thus, some fluctuation in activity is good for network performance (also see Levy and Wu 1999). For each simulation, good performance is defined as 80% or more correct responses at the end of training for all comparisons AB , BC , CD , DE , BD , and AE . Ten networks were trained and tested. Thus, 70% good performance means 7 out of 10 simulations had at least 80% correct on all test pairs. Note that the level of failure on BD is similar to that of AB , i.e. poor performance is distributed evenly throughout the six comparisons.

the lowest variance of activity is not best for good learning and activity variance itself is not monotonically related to performance. This confirms and extends our previous findings that a moderate level of fluctuation during encoding helps learning (see Levy and Wu 1999). The second point hints that the larger frequency fluctuations of negative gradients are superior to the lower frequency oscillations of positive gradients.

7. Discussion

In this paper, we have presented a simple solution to the problem of the uncontrolled random oscillations (shown in figure 1) in biologically relevant models of the hippocampus used to explore the biological basis of cognition. As such a simple and biologically meaningful solution, the innovations here move the minimal model one step closer to the physiology since the previous model is unable to compute with the numbers of neurons that actually exist in hippocampal region CA3.

Because of computational limitations, mammalian brain circuitry must be simulated with many fewer neurons than actually exist in the brain region being studied. It is typically argued that the simulations will only improve as the number of neurons is increased. (For example, Tsodyks and Sejnowski (1995) increase the number of neurons while scaling the neurotransmitter release probabilities with the number of inputs to maintain average input.) In our earliest studies of the CA3 model, improvements were observed as we increased the number of neurons from 128 to 512. However, much larger networks are desirable to reproduce the exceedingly low activity levels that physiologists have recorded in the hippocampus including

region CA3 (Thompson and Best 1989). Unfortunately, we can only model these low activity levels if a large number of neurons is present. Thus, from this perspective of producing actual cell firing, it is critical that the models be scaled up in the sense of increasing the number of neurons while decreasing the active fraction over the period of neuronal integration. In contrast to our earlier scale-up successes, without the constant K_0 in the denominator of equation (1), scale-up from 1000 to 4000 neurons, while decreasing fractional activity, was unsuccessful. Figure 1 explains this failure.

The importance of scaling-up forced re-examination of the minimal biologically inspired neuronal equation used in previous work (Levy 1996). Specifically, we find (equation (23) in the appendix) that without a resting shunt large-scale simulations are certain to be unstable for activities below approximately 10%. From the results presented here concerning the role of a resting (i.e. constant) shunt on the stability of total activity in our CA3 model and its role in controlling the gradient of the return map, we conclude that this resting shunt seems to be the only necessary missing ingredient needed for continued scale-up of the model.

For a given network size, there are many combinations of inhibition parameters K_R and K_0 that can produce a given activity level. However, starting with a large enough K_R , as we decrease K_R and increase K_0 , the magnitude of the activity fluctuations passes through a minimum. As the network grows larger and synaptic weights and number of fan-in connections are kept constant, the values for K_R and K_0 must be altered to maintain the desired fractional activity and this minimum fluctuation level. Thus, it is necessary to adjust the constant inhibition K_0 as well as variable inhibition K_R to stabilize activity fluctuations in large networks.

The K_0 term is not only important for large networks but also important for relatively small networks (e.g. 1024 neurons used here) when learning hard problems like TI. Without K_0 term, only half of the networks tested here for the TI problem had performance consistent with experimental results. On the other hand, for a suitable K_0 , the frequency of consistent performance can be as high as 90% (see figure 7).

Although we have emphasized controlling activity fluctuations and although our TI results for $K_0 = 0$ show that overly large fluctuations are destructive to proper learning, we do not mean to imply that minimal activity fluctuations are always best. As is well known, adding randomness into neural networks can improve computation and pattern recognition (Kirkpatrick *et al* 1983, Hinton and Sejnowski 1983, Geman and Geman 1984). Moreover, in analogy to the helpfulness of fluctuations in pattern recognition, improved learning can result using a nonzero gradient (figure 7) or by injecting some noise during training (Wu and Levy 1999).

In summary, this paper shows the insufficiency of our previous minimal biological model if actual neuronal numbers and activities are to be employed. On the other hand, the simplicity and biological appropriateness of the constant K_0 —either as the resting conductance in the shunting model or as the threshold in the subtractive inhibitory model—encourage us to incorporate this parameter in our future studies. A similar but modified analysis should be possible for more elaborate models that include integrate and fire neurons.

Acknowledgments

This work was supported by NIH MH48161 to WBL and MH57358 to N Goddard (with subcontract no 163194-54321 to WBL), and by the Department of Neurosurgery. We would like to thank Matthew Harrison and Per Sederberg for their invaluable assistance with simulations.

Appendix. Exact solutions

Exact solutions can be found for the simultaneous solution of equations (9) and (10) when the hyperbolic tangent is used to approximate the binomial solution. Equations (9) and (10) become:

$$a = \left(\frac{n - m_e}{2} \right) \left[1 - \tanh \left(\sqrt{\frac{2}{\pi}} \frac{\alpha a + \beta + \gamma m_e - ap}{\sqrt{ap(1-p)}} \right) \right] + m_e \quad (16)$$

and

$$g = - \left(\frac{n - m_e}{2} \right) \frac{d}{dm} \tanh \left(\sqrt{\frac{2}{\pi}} \frac{\alpha m + \beta + \gamma m_e - mp}{\sqrt{mp(1-p)}} \right) \Big|_a. \quad (17)$$

Inverting equation (16) gives

$$\beta + \gamma m_e - ap + \alpha a = \sqrt{\frac{\pi}{2}} \sqrt{ap(1-p)} \tanh^{-1} \left(1 - \frac{2(a - m_e)}{(n - m_e)} \right) \equiv A \quad (18)$$

and making use of the fact that $d \tanh(x)/dx = 1 - \tanh^2 x$ and substituting equation (16) into (17) yields,

$$\beta + \gamma m_e + ap - \alpha a = \sqrt{\frac{\pi}{2}} \frac{\sqrt{ap(1-p)} g a (m_e - n)}{(a - m_e)(a - n)} \equiv B \quad (19)$$

where A and B are defined as the right-hand sides of the equations. Solving these equations simultaneously implies

$$K_R = \frac{w(1-\theta)}{\theta} \left(\frac{A - B}{2a} + p \right) \quad (20)$$

and

$$K_0 = \frac{w(1-\theta)}{\theta} \left(\frac{A + B}{2} - \gamma m_e \right). \quad (21)$$

Hence, for any choice of activity, a , and gradient, g , we can compute the approximate values for K_R and K_0 .

Note that when all parameters are fixed with the exception of K_R , K_0 and g , A is constant and B is linear in g . Adding equations (20) and (21) yields a linear equation for K_R in terms of K_0 :

$$K_R = -\frac{K_0}{a} + \frac{w(1-\theta)}{\theta a} (A + ap - \gamma m_e), \quad (22)$$

indicating that lines of isoactivity (cf figure 4(a)) should be linear with gradient of $-a^{-1}$.

Using the above equations, it is also possible to identify the minimum activity below which the network will be unstable if no resting shunt, K_0 , is included in the inhibitory term. Assuming no external inputs ($m_e = 0$) and putting $K_0 = 0$, equation (21) implies that $A = -B$. Using this fact, substituting equation (18) into (19) and then considering the condition of stability $g > -1$, we have

$$\tanh \left(\frac{1}{1 - \frac{a}{n}} \right) > 1 - 2\frac{a}{n}. \quad (23)$$

Solving this equation numerically indicates that fractional activity, a/n , must be larger than approximately 10% for stability. (This value is slightly lower than the 12% activity quoted in the section on scale-up which did not use the tanh approximation.) When a more practical criterion for stability of $g > -0.7$ is chosen, the critical activity below which the network becomes unstable is approximately 16%. Interestingly, equation (23) indicates that the criterion for stability when $K_0 = 0$ has no dependence on w , θ or p .

References

- Amaral D G, Ishizuka N and Claiborne B 1990 Neurons, numbers and hippocampal networks *Understanding the Brain through the Hippocampus: The Hippocampal Region as a Model for Studying Brain Structure and Function (Progress in Brain Research vol 83)* ed J Storm-Mathisen et al (Amsterdam: Elsevier) pp 1–11
- Amaral and Witter 1995 Hippocampal formation *The Rat Nervous System* 2nd edn, ed G Paxinos (San Diego, CA: Academic) pp 443–93
- Amari S-I 1972 Characteristics of randomly connected threshold-element networks and networks systems *Proc. IEEE* **59** 35–47
- Amarasingham A and Levy W B 1998 Predicting the distribution of synaptic strength and cell firing correlations in a self-organizing, sequence prediction model *Neural Comput.* **10** 25–57
- Amit D J and Brunel N 1997 Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex *Cerebral Cortex* **7** 237–52
- Andersen P, Trommald M and Jensen V 1994 Low synaptic convergence of CA3 collaterals on CA1 pyramidal cells suggests few release sites *Molecular and Cellular Mechanisms of Neurotransmitter Release* ed L Stjarne et al (New York: Raven) pp 341–51
- Buckingham J and Willshaw D 1993 On setting unit thresholds in an incompletely connected associative net *Network: Comput. Neural Syst.* **4** 441–59
- Carandini M and Heeger D J 1994 Summation and division by neurons in primate visual-cortex *Science* **264** 1333–6
- Derrida B, Gardner E and Zippelius A 1987 An exactly solvable asymmetric neural network model *Europhy. Lett.* **4** 167–73
- Dusek J A and Eichenbaum H 1997 The hippocampus and memory for orderly stimulus relations *Proc. Natl Acad. Sci. USA* **94** 7109–14
- Furman G G 1965 Comparison of models for subtractive and shunting lateral-inhibition in receptor-neuron fields *Kybernetik* **2** 257–74
- Geman S and Geman D 1984 Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–41
- Ginzburg I and Sompolinsky H 1994 Theory of correlations in stochastic neural networks *Phys. Rev. E* **50** 3171–91
- Grossberg S 1988 Nonlinear neural networks: principles, mechanisms, and architectures *Neural Netw.* **1** 17–61
- Hale J and Koçak H 1991 *Dynamics and Bifurcations* (New York: Springer)
- Hasselmo M E, Schnell E and Barkai E 1995 Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3 *J. Neurosci.* **15** 5249–62
- Hendin O, Horn D and Tsodyks M V 1997 The role of inhibition in an associative memory model of the olfactory bulb *J. Comput. Neurosci.* **4** 173–82
- Herrmann M, Ruppin E and Usher M 1993 A neural model of the dynamic activation of memory *Biol. Cybern.* **68** 455–63
- Hinton G E and Sejnowski T J 1983 Optimal perceptual interference *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (Washington, 1983)* (New York: IEEE) pp 448–53
- Holt G R and Koch C 1997 Shunting inhibition does not have a divisive effect on firing rates *Neural Comput.* **9** 1001–13
- Hopfield J J 1982 Neural networks and physical systems with emergent collective computational abilities *Proc. Natl Acad. Sci. USA* **79** 2554–8
- Kirkpatrick S, Gelatt C D Jr and Vecchi M P 1983 Optimization by simulated annealing *Science* **220** 671
- Lansner A and Fransén E 1992 Modelling Hebbian cell assemblies comprised of cortical neurons *Network: Comput. Neural Syst.* **3** 105–19
- Levy W B 1989 A computational approach to hippocampal function *Computational Models of Learning in Simple Neural Systems, The Psychology of Learning and Motivation* vol 23, ed R D Hawkins and G H Bower (San Diego, CA: Academic) pp 243–305
- 1996 A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks *Hippocampus* **6** 579–90
- Levy W B and Wu X B 1995 Controlling performance by controlling activity levels in a model of hippocampal region CA3. II: Memory capacity comes at the expense of context cell firing and compressed coding *INNS World Congress on Neural Networks* vol 1, pp 582–6
- 1996 The relationship of local context codes to sequence length memory capacity *Network: Comput. Neural Syst.* **7** 371–84
- 1997 A simple, biologically motivated neural network solves the transitive inference problem *Proc. IEEE Int. Conf. on Neural Networks (1997)* pp 368–71
- 1999 Some randomness benefits a model of hippocampal function disorder versus order in brain function—essays

- in theoretical neurobiology ed P Århem *et al* (Singapore: World Scientific) at press
- Levy W B, Sederburg P B and August D 1998 Sequence compression by a hippocampal model: a functional dissection *Computational Neuroscience: Trends in Research, 1998* ed J M Bower (New York: Plenum) pp 435–9
- Levy W B, Wu X B and Baxter R A 1995 Unification of hippocampal function via computational/coding considerations *Int. J. Neural Systems* **6** 71–80 (suppl.)
- Li X G, Somogyi P, Ylinen A and Buzsaki G 1994 The hippocampal CA3 network: an *in vivo* intracellular labelling study *J. Comput. Neurol.* **339** 181–208
- Marr D 1971 Simple memory: a theory for archicortex *Phil. Trans. R. Soc.* **262** 23–81
- McNaughton B L 1996 Cognitive cartography *Nature* **381** 368–9
- Miles R and Wong R K S 1986 Excitatory synaptic interactions between CA3 neurones in the guinea-pig hippocampus *J. Physiol.* **373** 397–418
- Milton S J and Arnold J C 1990 *Introduction to Probability and Statistics* (New York: McGraw-Hill)
- Minai A A and Levy W B 1993a Predicting complex behaviour in sparse asymmetric networks *Neural Information Processing Systems 5: Natural and Synthetic* pp 556–63
- 1993b The dynamics of sparse random networks *Biol. Cybern.* **70** 177–87
- 1994 Activity level in random networks *Neural Comput.* **6** 85–99
- Polyn S, Wu X B and Levy W B 1999 Entorhinal/dentate excitation of CA3: a critical variable in hippocampal models *Neurocomputing* at press
- Simmen M W, Treves A and Rolls E T 1996 Pattern retrieval in threshold linear associative nets *Network: Comput. Neural Syst.* **7** 109–22
- Tamamaki N and Nojyo Y 1990 Disposition of the slab-like modules formed by axon branches originating from single CA1 pyramidal neurons in the rat hippocampus *J. Comput. Neurol.* **291** 509–19
- 1993 Projection of the entorhinal layer II neurons in the rat as revealed by intracellular pressure-injection of neurobiotin *Hippocampus* **3** 471–80
- Thompson L T and Best E T 1989 Place cells and silent cells in the hippocampus of freely-behaving rats *J. Neurosci.* **9** 2382–90
- Tsodyks M V and Sejnowski T 1995 Rapid state switching in balanced cortical network models *Network: Comput. Neural Syst.* **6** 111–24
- Wallenstein G V and Hasselmo M E 1997 GABAergic modulation of hippocampal population activity: sequence learning, place field development, and the phase precession effect *J. Neurophys.* **78** 393–408
- Wilson H R and Cowan J D 1972 Excitatory and inhibitory interactions in localized populations of model neurons *Biophys. J.* **12** 1–24
- Wolfram S 1996 *Mathematica* (Cambridge: Cambridge University Press)
- Wu X B and Levy W B 1995 Controlling performance by controlling activity levels in a model of hippocampal region CA3. I: overcoming the effect of noise by adjusting network excitability parameters *INNS World Congress on Neural Networks* vol 1, pp 577–81
- 1998 A hippocampal-like neural network model solves the transitive inference problem *Computational Neuroscience: Trends in Research, 1998* ed J M Bower (New York: Plenum) pp 567–72
- 1999 Enhancing the performance of a hippocampal model by increasing variability early in learning *Neurocomputing* **26–27** 601–7