Model of Global Spontaneous Activity and Local Structured Activity During Delay Periods in the Cerebral Cortex

Daniel J. Amit[1,2] and Nicolas Brunel[2,3]

[1]Racah Institute of Physics, Hebrew University, Jerusalem, Israel and [2]INFN, Sezione di Roma, Istituto di Fisica, Università di Roma, La Sapienza, Ple Aldo Moro, Roma, Italy
[3]Present address: LPS, Ecole Normale Supérieure, 24, rue Lhomond, 75231 Paris Cedex 05, France

We investigate self-sustaining stable states (attractors) in networks of integrate-and-fire neurons. First, we study the stability of spontaneous activity in an unstructured network. It is shown that the stochastic background activity, of 1–5 spikes/s, is unstable if all neurons are excitatory. On the other hand, spontaneous activity becomes self-stabilizing in presence of local inhibition, given reasonable values of the parameters of the network. Second, in a network sustaining physiological spontaneous rates, we study the effect of learning in a local module, expressed in synaptic modifications in specific populations of synapses. We find that if the average synaptic potentiation (LTP) is too low, no stimulus specific activity manifests itself in the delay period. Instead, following the presentation and removal of any stimulus there is, in the local module, a delay activity in which all neurons selective (responding visually) to any of the stimuli presented for learning have rates which gradually increase with the amplitude of synaptic potentiation. When the average LTP increases beyond a critical value, specific local attractors (stable states) appear abruptly against the background of the global uniform spontaneous attractor. In this case the local module has two available types of collective delay activity: if the stimulus is unfamiliar, the activity is spontaneous; if it is similar to a learned stimulus, delay activity is selective. These new attractors reflect the synaptic structure developed during learning. In each of them a small population of neurons have elevated rates, which depend on the strength of LTP. The remaining neurons of the module have their activity at spontaneous rates. The predictions made in this paper could be checked by single unit recordings in delayed reponse experiments.

It has recently been pointed out (Bernander *et al.*, 1991; Rapp *et al.*, 1992; Amit and Tsodyks, 1992; Shadlen and Newsome, 1994) that the apparently low background spontaneous activity in cortex, ~1–5 spikes/s, is high and prominent when looked at from the vantage point of the post-synaptic neuron. The reason is that the connectivity in neo-cortex is very high [of the order of 20 000 synapses per neuron (e.g. Braitenberg and Schüz, 1991), so a neuron receives an enormous number of afferent spikes per second before it engages in any stimulus-related computational activity. Bernander *et al.* (1991) and Rapp *et al.* (1992) employed advanced analytical and numerical methods to evaluate in detail, within cable theory (Rall, 1977), the effect of this high background input on neural physiological parameters. Amit and Tsodyks (1992) emphasized the effects of this high background on the performance of the network as an attractor neural network, sustaining stimulus selective elevated rates following the removal of the stimulus. Observing that the additional afferents due to a reverberation provoked by a stimulus are small in comparison with the background, it was concluded that when operating as attractor networks the neurons can be described as linear elements on the input side. This has been an underpinning of essentially all neural network modeling (e.g. Amit, 1989).

The common conclusion to the first three studies is that neurons operating in the cortex have their physiological parameters significantly renormalized by the flux of spontaneous activity. A common assumption is that at this low activity rate, the spikes arriving on different synapses are uncorrelated. As a consequence networks in a living cortex are assemblies of effective elements, rather remote from the isolated neurons in a slice. The parameters of the renormalized neuron can, in principle, be computed from those of the bare one (Bernander *et al.*, 1991; Rapp *et al.*, 1992), or can be determined phenomenologically.

It is a curious fact that in constructing explanatory schemes for computational phenomena in cortex spontaneous activity has never been dealt with. This is true both for phenomenological explanations and for detailed models. What underscores the ommission is the fact that it is always the case that alongside any computational dynamics of a cortical region there is, depending on the stimulus, the option that the region will remain at spontaneous activity levels, if and when the stimulus does not relate to any previously learned or coded features. Recently we have presented a detailed quantitative account of delay activity correlations in working memory (Amit *et al.*, 1994), again avoiding the issue of spontaneous activity. At this point the need for a remedy has become pressing.

The prominent presence and the essential effect of the background neural activity, renews the question of the origin of this activity in the neural dynamics and of the mechanisms that keep it stable. Why would the cortex be stable at a rate of ~1–5 s^{-1}? One could invoke various external agents, but an explanation in terms of the system's own dynamical rules and parameters would be more satisfying.

Moreover, both anatomy and neurophysiology suggest a modular or columnar organization of the cortex on the scale of 1 mm^2 (Braitenberg and Schütz, 1991). This creates a local environment in which stimulus selective enhanced spike rates have been observed during the delay period following the removal of the stimulus, in delay response tasks, in several areas of the association cortex (e.g. in IT cortex: Fuster and Jervey, 1981; Miyashita and Chang, 1988; in prefrontal cortex: Funahashi *et al.*, 1989; Wilson *et al.*, 1993; in posterior parietal cortex: Koch and Fuster, 1989; for a review see Fuster, 1995). We refer to such persistent activity as 'delay activity'. Another question is therefore the origin of this self-sustained delay activity, which preserves an active memory of a visual stimulus after it is removed.

In what follows we shall argue that an explanation to both questions is available. What makes this program non-trivial is the fact that it must deal with several aspects concurrently.

First, since spontaneous activity is non-specific and global, it must be an attractor which is global on the cortical level. In other words, spontaneous rates at the observed levels have to reproduce the same rates when considered as afferents in

plausible cortical conditions, and these rates have to be stable, i.e. local fluctuations in rates, due either to variations in local spontaneous activity or to local selective delay activity, have to be dynamically suppressed. As we will see, this stability requirement imposes severe constraints on the various neural and synaptic parameters of the model neural assembly. One must then confirm that synaptic modifications due to learning do not destabilize this spontaneous activity state, consistent with the global parameter constraints.

Second, on top of this great cortical attractor, local variations in synaptic connectivity due to learning must allow for selective (stimulus dependent) activity, either during the presence of a stimulus or following it. This implies a coexistence between spontaneous and selective activity states in the local module. Furthermore, in the local, selective attractor, neurons whose rates are not significantly enhanced (background neurons) usually maintain a spiking rate at levels similar to the global spontaneous activity.

In order for the selective attractors to be stable, noise due to spontaneous activity in the background neurons (who greatly outnumber those carrying the selective activity) should not wash out the signal in the selective population. This issue becomes even more stringent when one observes that the rates in a local selective attractor are usually close to the spontaneous rates (Miyashita and Chang, 1988; Wilson *et al.*, 1993; Nakamura and Kubota, 1995). Thus, one has to demonstrate that selective activity states with these properties can actually be stabilized with realistic parameters.

The Present Work — Assumptions and Summary of Results

In the present study we investigate the stable states (attractors) of a network of integrate-and-fire neurons as a framework for discussing the issues raised above. We make the following assumptions:

1. *Background activity is stochastic*. This is an expression of the empirical character of spontaneous activity, as well as a condition under which we can analyze its stability. We conceive the origin of the stochasticity in the unknown and sporadic inputs that affect remote parts of the cortex.
2. *In the absence of a stimulus, all neurons in the cortex see the same environment, apart from statistical fluctuations*. This expresses the position that the spontaneous activity is non-specific: whatever variation among neurons exists, it is due entirely to statistical fluctuations in the distribution of synapses. It also expresses the view that spikes emitted in this dynamical phase are due to fluctuations in the afferent input arriving at a neuron. It is exactly these fluctuations which we shall estimate below, in order to determine the stability of this underlying activity state.
3. *Spike emission times of different neurons are uncorrelated*. This implies the supposition that neural activity does not synchronize in the range of biological parameters. Experimental cross-correlations indicate some deviation from this hypothesis which are beyond the present account.
4. *The synaptic input to every cell is the linear sum of individual synaptic contributions*.
5. *The dynamics of both spontaneous and selective activity can be described by approximating the synaptic input of a typical neuron by a Gaussian stationary process*. This is justified if spikes arrive at a low rate on a large number of uncorrelated channels (hence a sum of many Poisson processes) per neural integration time (Amit and Tsodyks, 1991). In the case of excitatory as well as inhibitory inputs, each of the two inputs can be considered as an uncorrelated Gaussian process, due to the large number of incoming spikes of either type. The sum of the two processes can therefore be approximated by a Gaussian process, whose mean is the difference of the two means and whose variance is the sum of both variances. This is where the present approach goes beyond a naive 'mean-field' approach.

These assumptions imply that a reliable description is available in terms of spike rates and afferent currents, rather than actual spikes. In this context we study the attractors of the network and their stability against small fluctuations.

We find that when all neurons are excitatory it is impossible to stabilize realistic spontaneous rates. The stable states have all neurons either quasi-silent or firing at elevated rates. If inhibitory neurons are included, spontaneous rates of 1–5 spikes/s (s^{-1}) can be stabilized for reasonable values of the parameters of the network. If the integration time constants of the membrane potentials (of the spike emitter) for excitatory and inhibitory neurons are equal, stability requires that the average local excitatory and inhibitory synaptic inputs to every neuron of the network closely balance. If inhibition integrates its depolarization a couple of times faster, this condition can be relaxed and realistic spontaneous rates are stable even if the local excitatory inputs dominate.

We find that the spontaneous activity is globally stable, as is the state with all neurons silent in absence of external inputs. But a 'cortex' that is in an active state of spontaneous activity will not drift gradually into the silent state. A major coherent fluctuation on a very large scale will be required for such a transition.

In that same network, which can maintain spontaneous rates of the order of 1–5 s^{-1}, we observe the effects of learning in a local module, as expressed in the potentiation of the excitatory efficacies in some specific populations of synapses. It is found that when long-term potentiation (LTP) takes place one can observe two phases. At low potentiation levels (early learning) there is a non-selective enhancement of the rates of all cells that have been selective (visually responsive) to any of the stimuli. Then, when the average potentiated synaptic efficacy exceeds a critical value, the local network develops rather abruptly a variety of local attractors on the background of the stability of the uniform global one. The local attractors reflect the structure implied by the learning process. In such an attractor a small population of neurons, representative of a particular stimulus, remains active at elevated rates (of the order of 20 s^{-1}) following the removal of the stimulus. The actual rates depend on the mean potentiated synaptic value. The neurons of the local module which do not participate in the active representation of the particular stimulus (background neurons) remain at spontaneous rates. The state of selective delay activity in the module, following presentation of a stimulus, is an active memory of this stimulus. However, even in the presence of LTP and selective attractors with elevated rates, the local module has always the option of a stable uniform activity at rates close to the spontaneous ones. This state of activity is chosen by the module when the afferent stimulus is uncorrelated with what has been learned in the past (e.g. Amit and Brunel, 1995). These

phenomena are direct predictions for single cell recordings during learning in delayed match to sample experiments.

It should be emphasized that, besides the structural features of the synthesis of single neuron and network effects, this modeling effort provides another important bonus. It provides the first account of the coexistence of low attractor rates in a small subpopulation (~1% of the cells) with relatively high spontaneous rates (in ~99% of the cells), as is the case in the recordings during delayed match to sample experiments (Miyashita and Chang, 1988; Wilson *et al.*, 1993). The relatively high rates of spontaneous activity are present both in the wide cortical area and the neurons of the local module which do not participate in the foreground of a given retrieved memory.

The organization of the paper is as follows. We first describe the model network and its elements. Then we study the spontaneous stable states in a purely excitatory network. When inhibitory units are added it is shown how they stabilize a global attractor in which neurons have realistic spontaneous rates. Finally, we investigate the effect of learning on the stable states of the network.

Methods

The Network

We consider a local module that is a large network, containing some 10^5 neurons (a cortical column) embedded in a vast sea of neurons (the entire cortex). We have in mind a local module of 1 mm diameter in a living cortex, as is observed in inferior temporal or pre-frontal cortex (Miyashita and Chang, 1988; Wilson *et al.*, 1993). Two features distinguish the local module from the global network: the intensity of excitatory connectivity (high internally and low otherwise) and the range of the inhibitory neurons. Neurons receive three types of synaptic inputs: from recurrent (collateral) excitatory connections with other neurons in the same network; from inhibitory neurons in the module; and from excitatory neurons in other unspecified areas. The latter are considered excitatory, since only pyramidal cells have axons that are long enough. These afferents represent unrelated activity in the rest of the cortex, and carry the stochastic component induced by sporadic inputs from remote parts of the cortex (see assumptions above). They are modeled as Poisson processes with a spontaneous spike rate which is kept fixed. Estimates as to the proportion x of excitatory synaptic inputs on a given neuron from local collaterals vary between 50 and 80%. For numerical examples we use the value 50% (Braitenberg and Schütz, 1991). A schematic representation of the connectivity is sketched in Figure 1.

To highlight the difficulties and their potential resolutions we first consider the system in the absence of the inhibitory feedback. Then the context is enlarged and the network includes both excitatory and inhibitory units. The collateral connectivity in the local network has no geometric structure: a neuron has equal probability of synapsing on any other neuron in the local module. To remain within biologically acceptable territory, the number of inhibitory units in the module is taken to be 20% of the number of local excitatory neurons (Abeles, 1991).

The Integrate-and-fire Neuron

Dynamics of the Membrane Depolarization

The network is composed of leaky integrate-and-fire (IF) neurons (e.g. Ricciardi, 1977; Tuckwell, 1988), converted into an interacting system of synaptic currents and output spike rates (Amit and Tsodyks, 1991). The assumptions underlying this description of the neurons will be scrutinized below. Each neuron in the local network is characterized by its depolarization at the soma $V(t)$, which obeys to the integrator equation

$$\tau \dot{V}(t) = -V(t) + I(t) \tag{1}$$

where $\tau \sim 10$ ms is the integration time of the membrane depolarization at the spike emitting part of the soma and $I(t)$ is the synaptic current charging that part of the membrane. Note that the current is expressed in units of the potential. The underlying conductance is absorbed in the synaptic efficacy (equation 2).

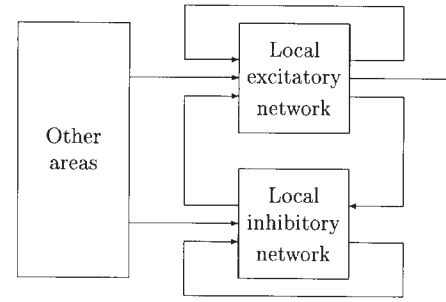

Figure 1. Connectivity scheme: excitatory and inhibitory neurons in the local network (module) receive collateral connections from neurons in the same module. Excitatory–excitatory local connections may be structured by learning. All other connections are structureless. Both excitatory and inhibitory neurons in the module receive excitatory synaptic inputs from elsewhere in the cortex.

When the depolarization reaches the threshold θ the neuron emits a spike, and the potential is reset to a value H after an absolute refractory period $\tau_0 \sim 2$ ms. In the situation we are considering, this absolute refractory period will not be relevant since the mean interspike interval will be very long. Yet it fixes the absolute scale of the rates, since it is the inverse of the maximal rate. The post-spike reset potential will be set to zero in numerical studies. We have checked that at low rates its actual value is immaterial.

Synaptic Inputs to the Cells

A neuron receives some $C \sim 20\,000$ synaptic contacts from other cells (Braitenberg and Schütz, 1991). A fraction x of these synapses come from the local module, while the remaining synapses come from outside. In the unstructured situation, the potential variability in synaptic efficacies on a neuron is schematized in the following way: the synaptic efficacies J_i (i = 1,..,C) on the dendrites of a given neuron are assumed to be randomly distributed, according to a distribution $P(J)$, with mean J and standard deviation (SD) ΔJ. This distribution can account for randomness in the synaptic transmission, different electrotonic distances of the synaptic sites on the dendrites, etc.

Dynamics of the Afferent Currents

The effective, afferent current $I(t)$, due to the temporal variation of the synaptic conductances provoked by afferent spikes and charging the spike-emitting integrator in equation (1), obeys the equation:

$$\tau' \dot{I}(t) = -I(t) + \sum_{i=1}^{C} J_i \tau \sum_k \delta\left(t_i^k - t\right) \tag{2}$$

i goes over the synaptic sites on the dendrites, the sum over k is over all spikes arriving at a given site, and t_i^k is the time of arrival of spike number k at synapse i. τ' is the time constant of the conductance changes at the synaptic sites (e.g. Frolov and Medvedev, 1986; Amit and Tsodyks, 1991). τ multiplying J_i gives J_i the dimensions of I_i.

If $\tau' << \tau$, i.e. the time constant of the currents is much shorter than that of the potentials, one can replace I in equation (1) by the source term on right-hand side of equation (2), i.e.

$$\tau \dot{V} = -V + \sum_{i=1}^{C} J_i \tau \sum_k \delta\left(t_i^k - t\right) \tag{3}$$

In this case, on a time scale of the order of τ, the depolarization $V(t)$ is a sum of unit contributions each with amplitude J_i.

Dynamics of a Single EPSP

Let us assume that $V = 0$ at $t = 0$, and a spike arrives on a synapse i with

efficacy J_i. In the absence of other spikes, the potential, for $t > 0$, will be given by

$$V(t) = J_i \exp(-t/\tau)$$

This is the unitary EPSP in the model, when the time constant of the currents is very short. The mean of J_i, J, is also the average EPSP amplitude. Estimates of J range from 0.05 to 0.5 mV (McNaughton *et al.*, 1981; Komatsu *et al.*, 1988; Thompson *et al.*, 1988; Sayer *et al.*, 1990; Mason *et al.*, 1991). We use an average of 0.2 mV, and with a threshold of 20 mV, some 100 simultaneous synaptic events will bring the post-synaptic neuron to threshold. The variability in amplitude of unitary EPSPs from different synapses is taken of the same order as their average amplitude, thus $\Delta \sim 1$ (Mason *et al.*, 1991).

Independent Poisson Activation of the Input Channels

Suppose that each synapse, from the local module as well as from outside, is activated by an identical, independent Poisson process, and spikes arrive at each channel with uniform rate ν. (For simplicity we assume that ν is uniform. The argument can be extended to account for variability of ν due to variabilities in the connectivity. See also Discussion.) If ν is low, but in an interval τ the number of arriving spikes is high, due to the large number of input channels C, then the source of the depolarization has a Gaussian distribution, whose mean $\mu(\nu)$ and SD $\sigma(\nu)$ are rate-dependent. To calculate the output rate of the neuron we use two different neuronal models, the simplified IF neuron and the full IF neuron.

Simplified IF Neuron

To obtain a preliminary idea concerning the neuron's output rate, we use the following simple approximate treatment of the spike emission process. In every integration time interval τ we assume that there is no decay. In each of these intervals the depolarization V is equal to the sum of unitary inputs J_i arriving in that interval. This sum would be a Gaussian variable. At the end of the interval either a spike is emitted and the depolarization becomes zero or no spike is emitted and still the depolarization is reset to zero, this time as an expression of the exponential decay of the integrator. The probability for the emission of a spike in such an interval is taken to be equal to the probability that the depolarization goes above threshold. These are crude approximations, but they do provide reasonable guides. They are eliminated in the full IF neuron.

Within this framework, if the synaptic input to the neuron in the τ interval has mean μ and SD σ and the threshold is θ, the probability that a spike be emitted in an interval τ is:

$$\begin{aligned} \Pr(\nu) &= \int_{\theta}^{\infty} \frac{dI}{\sqrt{2\pi\sigma^2(\nu)}} \exp\left(-\frac{\left(I-\mu(\nu)\right)^2}{2\sigma(\nu)^2}\right) \\ &= \int_{\alpha(\nu)}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \end{aligned} \qquad (4)$$

where we have performed the change of variable $z = (I - \mu)/\sigma$. $\Pr(\nu)$ depends on the rate of the afferent spikes via $\alpha(\nu)$, expressing the 'distance' between the average input and the threshold in units of the SD of the distribution,

$$\alpha(\nu) = \frac{\theta - \mu(\nu)}{\sigma(\nu)} \qquad (5)$$

Note that all rates are expressed in spikes per integration time τ, or equivalently in units of τ^{-1}. The resulting output spike rate ν_{ext} is therefore given by

$$\nu_{out} = \frac{\Pr(\nu)}{\tau} \qquad (6)$$

in s^{-1}. If the resulting output spike rate is to be very low, we must have $\mu < \theta$. Otherwise the probability is > 0.5 per τ, and if $\tau \sim 10$ ms the rate will be > 50 s^{-1}.

Full IF Neuron

In a more realistic description of spike emission, the depolarization $V(t)$ is governed by equation (3), in which it is charged by a current of Gaussian distribution, uncorrelated at different times, and a spike is emitted every time the depolarization $V(t)$ reaches threshold. The output spike rate is given by the inverse of the mean time between two consecutive events in which the depolarization reaches threshold [the mean interspike interval is approximated by the solution of the mean first passage time (e.g. Ricciardi, 1977; Tuckwell, 1988) plus the absolute refractory period τ_0], namely

$$\nu_{out} = \left(\tau_0 + \tau \int_{\frac{H-\mu(\nu)}{\sigma(\nu)}}^{\frac{\theta-\mu(\nu)}{\sigma(\nu)}} du\, \varphi(u)\right)^{-1} \qquad (7)$$

with $\varphi(u) = \sqrt{(\pi)}\exp(u2)[1 + \mathrm{erf}(u)]$. $\mu(\nu)$ is the rate-dependent mean of the Gaussian distribution of afferents depolarizing the cell and $\sigma(\nu)$ is $\sqrt{2}$ times the SD of the cell depolarization $\sigma[V]$. In a stationary situation, $\sigma(\nu)$ is equal to the SD of the synaptic inputs $\sigma[I]$, since $\sigma[V] = \sigma[I]/\sqrt{2}$ (Tuckwell, 1988). The simplified model of the neuron is a good approximation of the full IF neuron when the spike rates are much lower than $1/\tau$ (100 s^{-1}).

Concerning the assumptions made above, the rates arriving at each synapse on a given neuron are low, yet the number of synapses is so high that the assumption of a high number of spikes arriving in an integration time of the depolarization is reasonable. The most tenuous assumption is probably that concerning the ratio of the two time constants. If the time constant of conductance changes is of the same order or larger than that of the depolarization (Bekkers and Stevens, 1989; Hestrin *et al.*, 1990; Zador *et al.*, 1990), the replacement leading to equation (3) is not valid and the conditions leading to the current-rate transduction function, equation (7), are not satisfied. We return to this issue in Discussion.

In this paper it is assumed that at low rates the process of spike emission of each neuron is a Poisson process. Effects of the absolute refractory period, or of the resetting of the membrane depolarization following the emission of a spike, are negligible if the interspike interval is long compared with the integration time. Hence we shall consider only solutions with rates much smaller than 1. In such situations the variability of the interspike intervals is naturally unity, avoiding the complications raised by Usher *et al.* (1994) and Softky and Koch (1993).

Results

Excitatory Unstructured Network of Simplified Neurons

We begin with a network of simplified excitatory neurons only. The output rate of a neuron is given by equations (4–6), in which the mean μ and SD σ of the afferent current in an interval τ are:

$$\mu = JC\nu\tau, \quad \sigma = J\sqrt{C\nu\tau\left(1+\Delta^2\right)} \qquad (8)$$

where C is the number of synapses per neuron and the excitatory synapses have a distribution of efficacies with mean J and SD $J\Delta$, as described in Methods. For example, in cortical conditions, taking τ = 10 ms, C = 20 000 and Δ = 0, i.e. uniform synapses, and ν = 2 s^{-1}, then

$$\mu = 400\,J \quad \text{and} \quad \sigma = 20\,J$$

Thus the mean number of spikes arriving at a neuron per integration time is typically a few hundred, with fluctuations of a few tens.

Self-reproducibility

As was mentioned above, when the activity is purely spontaneous it is considered uniform globally. In that case the

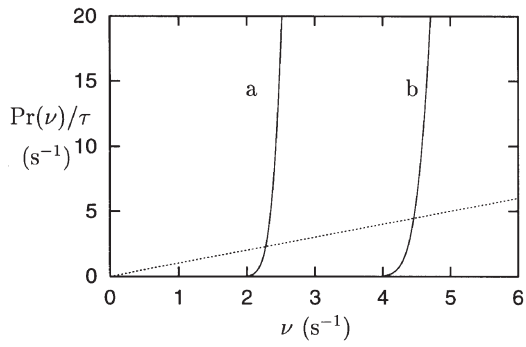

Figure 2. Numerical solution of equation (9) for two self-reproducing rates. Full curves: Pr(ν); straight line of slope unity (dotted): left-hand side of equation (9). The self-reproducing rates are the intersections for low non-zero rates. For $\theta/J = 500$ (curve a) $\nu = 2.3\ s^{-1}$, for $\theta/J = 940$ (curve b) $\nu = 4.5\ s^{-1}$. In both cases it is highly unstable (see text).

input to every neuron is equal on average, irrespectively of what proportion of the synapses are local and what are external. Since it is the rate itself which determines both μ and σ, the reproducibility of the spontaneous rate requires that the rate which determines these parameters be equal to the rate at which spikes are produced given those parameters, i.e. we must have

$$\nu = \frac{\Pr(\nu)}{\tau} = \frac{1}{\tau}\int_{\sigma(\nu)}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{9}$$

where α(ν) is given by equations (5) and (8). Given C and θ/J, one can solve equation (9) for the spontaneous spike rate which reproduces itself on average.

The full curves in Figure 2 are the function $\Pr(\nu)/\tau$ for $C = 20\,000$ and two values of θ/J. The solutions of equation (9) are the intersections of any of the curves with the straight line of slope unity in the figure. The intersections away from zero are at self-consistent average rates which are in the range of realistic spontaneous rates ($1–5\ s^{-1}$).

Local Stability

If a small fluctuation in the local mean rate displaces the rates in the system with no restoring force, this state is of little interest. Hence, we must look for solutions of equation (9) which are stable, i.e. which suppress fluctuations and restore the mean value of the spontaneous rates. For the rate to be stable against a fluctuation in the local mean rate, the change in the output rate must be smaller than the change in the input rate that caused it. The external synaptic input is assumed fixed, with each synapse coming from outside the local module activated at a rate ν_{ext}. The logic is that the dynamics of the external rates is outside our description, and the global constraint is that all spontaneous rates be equal. Local synapses are again activated at mean rate ν. They constitute a fraction x of the synapses on any neuron.

The μ and σ of the afferent input in an interval τ become:

$$\mu = C\left[x\nu\tau + (1-x)\nu_{ext}\tau\right]J \tag{10}$$

and

$$\sigma = J\sqrt{C\left(1+\Delta^2\right)\left[x\nu\tau + (1-x)\nu_{ext}\tau\right]} \tag{11}$$

The condition for self-reproducibility of rates is equation (9), where α(ν) is given by equation (5) with μ and σ given by equations (10) and (11). The condition for stability against fluctuations in the local rates is

$$\left|\frac{1}{\tau}\frac{d\Pr}{d\nu}(\nu^*)\right| < 1 \tag{12}$$

where ν* is the self-reproducing rate, the solution of equation (9).

We typically set $x = 0.5$ (Braitenberg and Schütz, 1991), i.e. afferents come in equal proportions from the local module and from the outside. Note that the results are not sensitive to the particular value of x. ν_{ext}, the external spike rate, is set at a fixed plausible value of spontaneous rate. Then the ratio θ/J is varied so that $\nu^* = \nu_{ext}$ is a solution of equation(9). This would be a globally consistent self-reproducing rate. The stability of a given self-reproducing rate is investigated by varying Pr(ν) with respect to ν, at fixed θ/J and ν_{ext}, to check the stability condition, equation (12).

It turns out that at the desired rates $1–5\ s^{-1}$ the condition (12) never holds. For equation (9) to have a solution at these rates one must have

$$\alpha(\nu) \approx 2 \tag{13}$$

i.e. the mean synaptic input at about two SDs below the threshold. For this value of α one can approximate dPr(ν)/dν (equation 9), by (see Appendix A for details)

$$\frac{1}{\tau}\frac{d\Pr}{d\nu}(\nu^*) \approx \alpha(\nu^*)\frac{\mu_I}{\sigma} \tag{14}$$

where $\mu_I = xC\nu\tau J$ is the average input coming from local collateral synapses. For the parameters chosen, μ_I/σ is of the order of 10, implying that the solutions at these rates are highly unstable. This instability can be read from Figure 2. Any perturbation of the rates will drive the network either to a state where all neurons are essentially quiescent, or to a state of high activity, in which the assumptions of the analysis do not hold. This is related to the fact that there are only excitatory elements in the network. The network lacks regulatory units that would control local perturbations in the rates. Note also that the ratios θ/J required to produce the low self-reproducing rates are very high. This again is due to the fact that we have only excitation. Including inhibition, the same self-reproducing rates are obtained with lower, more plausible, threshold to average EPSP ratios.

The simplified argument presented above serves to define the problem and to give an indication of the difficulties involved in finding a solution for the existence and the stability of self-reproducing spontaneous spike rates. As far as networks composed exclusively of excitatory neurons are concerned, a more careful analysis in a network composed of full IF neurons does not modify the conclusions. To have a stable attractor with low rates, it is necessary to introduce inhibitory neurons. This is done in the following.

The Role of Unstructured Local Inhibition

The main effects of the inhibitory input on the Gaussian process arriving at the soma of a neuron are to reduce the mean depolarizing current while increasing the SD of the net

depolarizing current. These effects will decrease the value of the ratio μ_I/σ in equation (14) and provide stability.

Both excitatory and inhibitory neurons are leaky integrate-and-fire neurons. Each type is characterized by a threshold θ_i, a post-spike hyperpolarization H_i, an integration time constant τ_i, with $i = E,I$ indicating whether the neuron is excitatory or inhibitory respectively. Each neuron receives a Gaussian white noise current of mean μ_i and SD σ_i, through C_i synaptic contacts, which are divided in C_{iE} excitatory synapses and C_{iI} inhibitory ones. The synapses in the network are now of four types, depending on all the possible types of pre- and post-synaptic neurons. For each synaptic type the efficacies are drawn randomly from the distribution $P_{ij}(J)$, with mean J_{ij} and SD $J_{ij}\Delta_{ij}$. i and j denote the post- and pre-synaptic type of neuron respectively. A fraction x_i of the excitatory connections come from outside the network and are activated at a rate ν_{ext}. The local excitatory synapses are activated at a uniform rate ν_E, while the inhibitory ones are activated at a uniform rate ν_I.

The number of parameters is large. To reduce this number we make the following choices, which we believe to be in a plausible region:

- both types of neurons receive an equal number of inputs C = 22 000, divided into C_{iE} = 20 000 excitatory and C_{iI} = 2000 inhibitory ones. The latter are assumed to originate all in the local module;
- both types of neurons receive an equal fraction of excitatory inputs from inside the local module. We choose $x = 0.5$;
- the relative SDs of the synaptic efficiacies are equal for all four synaptic types, i.e. $\Delta_{ij} = \Delta = 1$ for all i and j (Mason *et al.*, 1991);
- both excitatory and inhibitory post-spike hyperpolarizations are set to zero, i.e. $H_i = 0$ for all i.

We have checked that varying individually each of these parameters has a very mild effect on the dynamics of the network, provided the rates remain low.

The remaining parameters are

- the ratio of average strengthes of the inhibitory to the excitatory synapses, J_{EI}/J_{EE} and J_{II}/J_{IE};
- the ratio of the integration time constants τ_I/τ_E;
- the rates ν_E and ν_I.

The ratios θ_E/J_{EE} and θ_I/J_{IE} are determined by the requirement that preselected values for the rates ν_E and ν_I be self-reproducing given all the other parameters, and that the resulting $\nu_E = \nu_{ext}$. The input to each type of neuron is again Gaussian. The mean and variance of the input to an excitatory neuron are:

$$\mu_E = C_{EE}J_{EE}\left[x\nu_E\tau_E + (1-x)\nu_{ext}\tau_E\right] - C_{EI}J_{EI}\nu_I\tau_E \qquad (15)$$

and

$$\sigma_E^2 = \lambda\left(C_{EE}J_{EE}^2\left[x\nu_E\tau_E + (1-x)\nu_{ext}\tau_E\right] + C_{EI}J_{EI}^2\nu_I\tau_E\right) \qquad (16)$$

where $\lambda = 1 + \Delta^2$. Though the above expressions are largely self-explanatory, we discuss them in some detail in Appendix B. For inhibitory units we have

$$\mu_I = C_{IE}J_{IE}\tau_I\left[x\nu_E + (1-x)\nu_{ext}\right] - C_{II}J_{II}\tau_I\nu_I \qquad (17)$$

and

$$\sigma_I^2 = \lambda\tau_I\left(C_{IE}J_{IE}^2\left[x\nu_E + (1-x)\nu_{ext}\right] + C_{II}J_{II}^2\nu_I\right) \qquad (18)$$

We now perform the same type of analysis as in the case of excitation only, first in a network of simplified neurons in which excitatory and inhibitory neurons have identical characteristics, then in the more realistic network of leaky integrate-and-fire neurons.

Simplified Neurons

To simplify the qualitative argument, we set in equations (15–18) $\tau_I = \tau_E = \tau$ (i.e. equal inhibitory and excitatory time constants); $J_{EE} = J_{IE} = J$ (equal average excitatory efficacies to both types of neurons); $J_{EI} = J_{II} = g\,J$ (equal average inhibitory efficacies to both types of neurons); $\theta_I = \theta_E = \theta$ (equal thresholds); and $\Delta = 0$ (uniform synapses). In this case both types of neurons have equal rates ν and receive a synaptic input of the same mean and variance, i.e.

$$\begin{aligned}\mu &= \mu_{ext} + J(xC_{EE} - gC_{EI})\nu\tau \\ \sigma^2 &= \sigma_{ext}^2 + J^2(xC_{EE} + gC_{EI})\nu\tau\end{aligned} \qquad (19)$$

where μ_{ext} and σ_{ext}^2 are the mean and variance of the external input. Since with the present choice of parameters τ, ν, μ and σ are equal for both types of neurons, their self-reproducing rates ν^* are the solution of the same equation, i.e.

$$\nu^* = \frac{1}{\tau}\Pr(\nu^*) \qquad (20)$$

where $\nu^* = \nu_{ext}$, as above, and Pr is given as a function of ν by equation (4). The condition for the stability of a solution is again

$$\left|\frac{1}{\tau}\frac{d\Pr}{d\nu}(\nu^*)\right| < 1 \qquad (21)$$

Local Stability Analysis

We choose again ν_{ext} in the range 1–5 s^{-1}, and the self-reproducing rate $\nu^* = \nu_{ext}$. First, the self-reproducing rates are imposed by varying the ratio θ/J in equation (20). Then the variation of Pr with the local rate ν is computed at fixed ν_{ext} and θ/J. For a low self-reproducing rate, equation (21) reduces to the condition (see Appendix A for details)

$$\left|\frac{1}{\tau}\frac{d\Pr}{d\nu}(\nu^*)\right| \approx \left|\alpha\frac{\mu_I}{\sigma} + \alpha^2\frac{\sigma_I^2}{2\sigma^2}\right| < 1 \qquad (22)$$

where μ_I and σ_I^2 are, respectively, the mean and the variance of the synaptic input coming from the local collaterals, i.e.

$$\mu_I = J(xC_{EE} - gC_{EI})\nu\tau; \quad \sigma_I^2 = J^2\left(xC_{EE} + g^2C_{EI}\right)\nu\tau$$
$$\sigma^2 = \sigma_I^2 + \sigma_{ext}^2$$

The first term on the right-hand side of equation (22) is the relative perturbation of the output rate due to the perturbation

of the mean of the synaptic input, while the second term corresponds to the relative perturbation due to the variation of the variance of the synaptic input. When $xC_{EE} >> gC_{EI}$, i.e. there is a predominance of excitatory over inhibitory inputs, the first of the two terms in the right-hand side of equation (22) dominates. Since there is a large number of excitatory incoming spikes, $\mu_I >> \sigma$ and condition (22) is violated. On the other hand, if excitatory and inhibitory inputs roughly balance, i.e. $xC_{EE} \sim gC_{EI}$ (with our parameters: C_{EE} = 10 000 and C_{EI} = 2000, implying $g \sim 5$, i.e. inhibitory synapses are ~5 times more efficient than excitatory ones), the SD of the synaptic input becomes of the same order or larger than the mean local input μ_I, which makes it possible for the stability condition to be enforced.

The critical value of g at which the self-consistent solution of equation (20) becomes stable is actually slightly larger than 5 in a wide range of rates. It varies from $g \sim 5.5$ for $\nu = 1\ s^{-1}$ to $g \sim 5.1$ for $5\ s^{-1}$. This means that the inhibitory inputs have to be slightly larger than the *local* excitatory ones for the self-consistent rates to be stable. The reason is that when $g = 5$, i.e. local excitation exactly balances inhibition, the variation in the SD of the input acts to destabilize the solution of equation (20). Note that the total synaptic input remains depolarizing even for $g > 5$, due to the external excitatory component.

To visualize the effect of inhibition on the stability of the solution of equation (20) we present in Figure 3 the graphical solution for two values of the rate ($\nu = 2.3$ and $4.5\ s^{-1}$), each for three values of the relative inhibitory synaptic efficacy (g = 3, 5, 5.5). To obtain the same self-reproducing rate for different values of g the ratio θ/J is adjusted. In Figure 3 there are two rates at each of which the straight line (dotted), the left-hand side of equation (20), intersects three curves, the right-hand side of (20): g = 3 (full lines), g = 5.0 (long-dashed lines) and g = 5.5 (short-dashed lines). As g increases the slope of $\Pr(\nu)/\tau$ at the intersection decreases. At g = 3 it is essentially as for g = 0 (Fig. 2). At g = 5 both slopes are still larger than unity, because of the variation of the variance of the local synaptic input upon a variation of the rate. At g = 5.5 both self-reproducing rates are stable.

For low g the low-rate stable self-reproducing solution to equation (20) is essentially at $\nu = 0$. It is kept away from zero by the fixed excitatory input from outside the module, which leave the neurons at extremely low rates. When g crosses its critical value the state in which all neurons have a spontaneous rate of several spikes per second, depending on the ratio θ/J, becomes stable and small variations around this rate are suppressed.

There is an intermediate regime in which the network has two stable self-reproducing rates, the quiescent state with ν essentially at 0 and a state with spontaneous rates of several spikes per second. The lower curve c in Figure 3 represents such a situation. It corresponds to $\nu^* = 2.3\ s^{-1}$ and g = 5.5. Large enough fluctuations may shift the network from one state to the other. When the inhibition increases further, the quiescent state becomes unstable. This is because when inhibition (i.e. g) increases, one has to decrease the ratio θ/J to preserve the same self-reproducing rate. Since the fixed excitatory rate from outside the module is equal to the internal self-reproducing rate, the external currents will increase relative to the threshold when g increases, destabilizing the quiescent state and leaving the state with spontaneous rates of several spikes per second the only stable state of the network. The quiescent state becomes unstable when the external inputs are such that the curve $\nu \rightarrow \Pr(\nu)/\tau$ has a unique solution ν^*. This is the case for the upper curve c of Figure 3. The self-reproducing rate $\nu^* = 2.3\ s^{-1}$ is stable

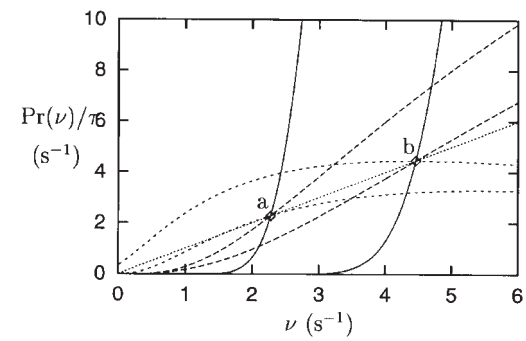

Figure 3. Graphical solution for two self-reproducing spontaneous spike rates $\nu^* = 2.3\ s^{-1}$ (*a*), $4.5\ s^{-1}$ (*b*), and three values of the relative inhibitory efficacy: $g = 3$ (full curves), $g = 5$ (long-dashed curves) and $g = 5.5$ (short-dashed curves). The (six) curves represent the function $\Pr(\nu)/\tau$, the right-hand side of equation (20). The dotted straight line is the left-hand side. Each ν^*, the intersection of $\Pr(\nu)/\tau$ with the straight line, is obtained for the three values of g by adjusting the ratio θ/J. The two solutions for ν^* are represented by diamonds in the figure. A self-reproducing rate is stable if the slope of $\Pr(\nu)/\tau$ at the intersection is lower than that of the straight line. Only for $g = 5.5$ (short-dashed curves) are the solutions stable. Note the similarity of the solutions corresponding to $g = 3$ (full curves) with Figure 2.

for $g > 5.45$, but the quiescent state is stable for $g < 5.7$. Thus both states are stable between these two values of g. On the other hand, the self-reproducing rate $\nu^* = 4.5\ s^{-1}$ becomes stable for $g > 5.18$, and coexists with the quiescent state only until $g = 5.35$. If one increases further g, the slope of the curve $\nu \rightarrow \Pr(\nu)/\tau$ at the self-reproducing rate becomes smaller than –1 and the self-reproducing rate becomes unstable. (We are in debt for this comment to an anonymous referee. However, this instability is an artifact of the discrete time formulation of our SIF neuron. The realistic, leaky IF neuron is stable for any inhibitory strength above the critical strength.)

Integrate-and-fire Neurons: Stable Spontaneous Rates

In a network of integrate-and-fire neurons the self-reproducing rates for both populations are given by the solutions of the coupled equations

$$\nu_E = \phi_E(\mu_E, \sigma_E) \tag{23}$$

$$\nu_I = \phi_I(\mu_I, \sigma_I) \tag{24}$$

in which the means μ_I and SDs σ_I are functions of the rates of the excitatory and the inhibitory neurons, ν_E and ν_I, via equations (15–18). The right-hand sides are

$$\phi_i(\mu_i, \sigma_i) = \left(\tau_0 + \tau_i \int_{\frac{H_i - \mu_i}{\sigma_i}}^{\frac{\theta_i - \mu_i}{\sigma_i}} du\, \varphi(u) \right)^{-1}, \quad i = E, I \tag{25}$$

with φ given in equation (7) and τ_I the time constant of neuron of type I.

The stability analysis consists of considering the effect of a uniform displacement of the two rates: $(\nu_E, \nu_I) \rightarrow (\nu_E + \delta\nu_E, \nu_I + \delta\nu_I)$ in the local module and exposing the conditions for which both output rates are closer to the self-reproducing values than the input ones. Details are given in Appendix B. In the following we discuss a few simple examples in the space of parameters for which the solution of equations (23) and (24) is stable against small perturbations.

Equal Time Constants and Inhibitory Couplings. We start by taking equal integration time constants for excitatory and inhibitory neurons, $\tau_I = \tau_E$, and equal relative inhibitory efficacies for both types of neurons, $J_{EI}/J_{EE} = J_{II}/J_{IE} = g$. The equations for the self-reproducing rates are solved numerically. Then we carry the stability analysis of Appendix B and find that the spontaneous rates are stable if excitatory and inhibitory local inputs approximately balance. In fact, the mean local synaptic input should again be slightly dominated by the inhibition. Hence, the smaller number of inhibitory inputs (20%) must be compensated either by stronger inhibitory synapses or by relatively higher inhibitory rates. This is not a very stringent condition. In hippocampus inhibitory couplings are indeed more efficient than the excitatory ones (Traub and Miles, 1991), and there is evidence that inhibitory rates are actually higher (Simons, 1978; Connors and Gutnick, 1990).

Fast Inhibition. As inhibition grows faster, any perturbation in the rates in the excitatory network is more easily controlled, and the network should be more stable. We find that decreasing the integration time constant of the inhibition relaxes the condition on the relative inhibitory couplings and on the inhibitory spike rates. The effect of increasing the rates of the inhibitory neurons is very similar. We have studied the stability of spontaneous activity in a network with excitatory rate $\nu_E = 2\ s^{-1}$ and equal relative inhibitory efficacy on both excitatory and inhibitory neurons ($J_{EI}/J_{EE} = J_{II}/J_{IE} = g$), for three different values of $\tau_I = \tau_E$, $0.5\tau_E$ and $0.2\tau_E$ (10, 5 and 2 ms), varying g and ν_I. Decreasing the integration time constant of inhibitory cells has the effect of increasing the domain of stability of spontaneous activity. At parity of excitatory rates, the spontaneous attractor will be stable for lower values of the average inhibitory efficacy or inhibitory rate. For example, if $\nu_I = \nu_E = 2\ s^{-1}$, the critical inhibitory strength g is 3.53 for $\tau_I = 5$ ms and 2.24 for $\tau_I = 2$ ms. In particular, the spontaneous rates will be stabilized even with large dominance of excitatory inputs over the inhibitory ones. This is due to the fact that the inhibition adjusts its rates faster in case of a perturbation. But as an integration time constant becomes very small, transmission delays tend to destabilize the spontaneous activity and the network finds it easier to transit to the quiescent state. This effect is counterbalanced by an increase in g or ν_I. Hence, a rather strong inhibition seems to be required to obtain a robust spontaneous activity state.

Robustness to Variations of Threshold-to-EPSP Ratio. Figure 4 shows the region of stability of the spontaneous activity upon varying the threshold-to-EPSP ratios θ_E/J_{EE} and θ_I/J_{IE}, for relative inhibitory efficacies: $g = 2$, 3 and 5. For each value of g, the spontaneous activity state is stable below the corresponding curve. Taking 20 mV for both excitatory and inhibitory thresholds, $\theta_E/J_{EE} = 300$ corresponds to an average efficacy $J_{EE} = 0.067$ mV, while $\theta_I/J_{IE} = 50$ corresponds to $J_{IE} = 0.4$ mV. These values are well within the experimentally reported range (McNaughton *et al.*, 1981; Komatsu *et al.*, 1988; Thompson *et al.*, 1988; Sayer *et al.*, 1990; Mason *et al.*, 1991). Figure 4 indicates that for the corresponding values of g, the ratio θ_I/J_{IE} has to be smaller than θ_E/J_{EE} (or conversely the average excitatory synapses have to be stronger on inhibitory cells), which in turn means that inhibitory cells have a higher rate. A decrease in g requires a compensating decrease of J_{EE}/J_{IE}, which implies higher inhibitory spike rates. For example, in a network with $g = 3$ (inhibitory synapses on both excitatory and inhibitory

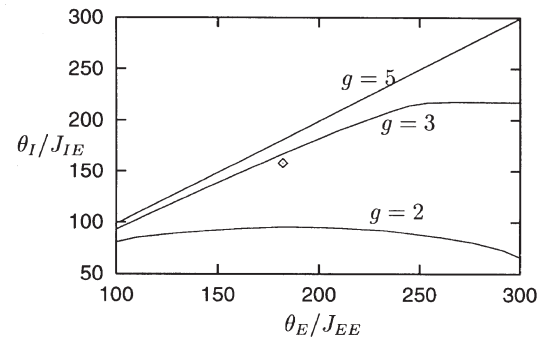

Figure 4. Stability region of spontaneous activity, in the space of threshold-to-EPSP ratios $\theta_E/J_{EE} - \theta_I/J_{IE}$ for three values of the relative inhibitory efficacy g (indicated in the figure). For each value of g the low rate state is stable below the corresponding curve. The diamond represents the numerical example discussed in text.

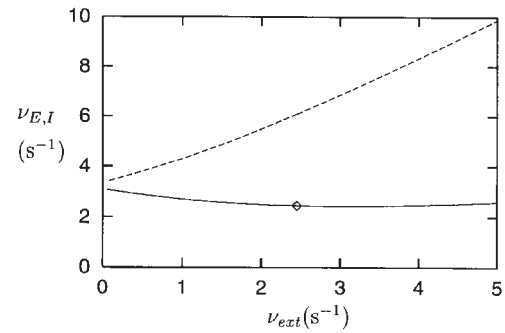

Figure 5. Dependence of spontaneous activity on variation of external afferents. Stable rates ν_E (full curve) and ν_I (dashed curve) as a function of the external rate vext. : $\nu_E = \nu_{ext}$. Fixed network parameters $\theta_E/J_{EE} = 136$, $\theta_I/J_{IE} = 85$, $g = 4$, $\tau_I = 0.4\tau_E$.

neurons three times as efficient as excitatory ones), $\theta_E/J_{EE} \sim 180$ and $\theta_I/J_{IE} \sim 160$ (marked by a diamond in Fig. 4), there is a spontaneous activity attractor with $\nu_E = 2\ s^{-1}$ and $\nu_I = 5.4\ s^{-1}$.

Effect of Variation of the External Currents. We turn next to the effect of variation of the external currents on the stability of the spontaneous activity state. We start from the stable state in which $\nu_E = \nu_{ext}$, and fix all parameters. Then the external rate ν_{ext} is varied, leading to a readjustment of the equilibrium rates in the network (ν_E,ν_I). Figure 5 presents the variation of the excitatory and inhibitory stable spontaneous rates upon variations of the external rate ν_{ext} for fixed parameters. The low rate solution is stable in the entire range ν_{ext} between 0 and $5\ s^{-1}$. The inhibition adjusts itself to the variation of the external rates, leaving the network in a spontaneous activity state of almost constant excitatory rates. When the external rate is zero, the network has another stable solution, the quiescent state $\nu_E = \nu_I = 0$. However, a major fluctuation would be required to reach this state.

Global Stability. Global stability can be considered in different ways. One may consider a sudden fluctuation in the spontaneous rates of all cells in the cortex. To have stability against such an eventuality, significantly larger inhibition would be required. If one allows for such global fluctuations, the system may also land in the global quiescent state, which is stable. But that is a very unlikely situation. The more relevant eventuality would be the spreading of a local fluctuation in rates. However, from the preceding arguments one may conclude that

our system is stable against such spreading. If we consider a particular module, and the rates fluctuate in a second module which has excitatory projections into the first, the local inhibition in the first module would be sufficient to suppress the effect at a cost perhaps of a small change in the inhibitory rate.

Learning and Selective Activity

The next major issue is whether a network characterized by the same set of parameters defining neuronal physiology and anatomy, which provided a stable spontaneous activity (even at relatively high rates), can sustain stable local delay activity by the mere modification of synapses between excitatory neurons.

We imagine the local, structured distribution of synaptic efficacies to have been generated by stimuli arriving at the module and being learned (e.g. Amit and Fusi, 1994; Amit and Brunel, 1995). Each stimulus activates a given subset of cells in the local module significantly above the spontaneous level (the visually responsive cells) and leaves the others at spontaneous rates (i.e. binary stimuli). The excitatory synapses in the local network have Hebbian plasticity: a pair of neurons simultaneously activated by a given stimulus will enhance the efficacy of the connecting synapse while a pair of neurons with anti-correlated activities reduce the efficacy of the synapse connecting them. All synapses involving inhibitory neurons remain unstructured. This is admittedly a very schematic representation of learning, but it suffices for the discussion of the issue raised here which concerns the possible stability of local, selective activity attractors on the background of the global spontaneous activity.

Suppose that p stimuli had been learned. Each stimulus activates an equal fraction f (coding level or sparseness) of the excitatory neurons in the local network. In order to obtain an intuition into the behavior of a network in which stimuli choose their selective neurons randomly and independently, we revert to an approximate situation in which stimuli are subject to the constraint that no neuron responds to more than one stimulus (non-overlapping stimuli). This case is amenable to detailed analysis (Brunel, 1994) and provides a reasonable approximation for the unconstrained case, provided the coding level is low and the number of learned stimuli is not too high, i.e. we must have $fp < 1$. The analysis of the case of non-overlapping stimuli is simplified by the fact that the excitatory network can be divided into two subpopulations: one of pfN neurons excited by one of the p learned stimuli (selective neurons); and the remaining $(1 - pf)N$ non-selective neurons, i.e. those not activated by any stimulus. We take $f = 0.01$, which may correspond to the situation in IT cortex (e.g. Brunel, 1994), and $p = 50$. For the same values of p and f, if stimuli were composed of neurons selected randomly and independently, ~9 of the neurons in the module would be activated by more than one stimulus.

The distribution of excitatory synapses on neurons of each of the two population is as follows.

Selective Neurons

Each of these neurons receives $\sim fxC_{EE}$ = 100 synaptic contacts from other neurons in the subpopulation which is activated by the same stimulus. These synaptic efficacies are characterized by a new (potentiated) probability distribution $P_+(J)$ with mean $J_+ > J$, where J ($= J_{EE}$) is the average, excitatory, unstructured synaptic efficacy sustaining the spontaneous activity. The enhanced mean of this distribution expresses the potentiation (LTP) due to coactivation of pre and post synaptic neurons by the same stimulus (e.g. Bliss and Collingridge, 1993). The structured local synaptic efficacies have an SD $J_+\Delta$. A selective neuron has $(1 - f)xC_{EE}$ synaptic contacts with other excitatory neurons which are not excited by this stimulus. These neurons are either neurons selective for a different stimulus or non-selective neurons. The corresponding synaptic contacts are characterized by a distribution $P_-(J)$, with mean $J_- < J$ and SD $J_-\Delta$, expressing depression (LTD) in these synapses, due to anti-correlation of the activities of pre- and post-synaptic neurons when the stimulus is present (e.g. Artola and Singer, 1993).

Non-selective Neurons

These neurons receive $pfxC_{EE}$ = 5000 contacts from selective neurons, which are drawn from the depressed distribution $P_-(J)$. They also receive $(1 - pf)xC_{EE}$ = 5000 contacts from other non-selective neurons. These synapses are not affected by learning and therefore are drawn from the original distribution $P(J)$.

The distributions $P_\pm$ characterize the learning of the set of stimuli in the network. We suppose that the mean excitatory synaptic efficacy in the entire module is not changed by the learning process. So on average synaptic depressions compensate for potentiations (see in this context Stanton and Sejnowski, 1989; Amit and Fusi, 1994). This is expressed as

$$pf^2 J_+ + pf\left[2 - f(p+1)\right]J_- + (1 - pf)^2 J = J$$

which is a statement that in the local module the mean potentiation plus the mean depression plus the mean unmodified efficacies gives the pre-learning mean. J_- can therefore be expressed as a function of J_+, i.e.

$$\frac{J_-}{J} = \frac{2 - f(p + J_+/J)}{2 - f(p+1)} \qquad (26)$$

Usually, the performance of an attractor network, in terms of the number of memories it can associatively recall, increases as the coding level decreases (Willshaw *et al.*, 1969; Tsodyks and Feigelman, 1988; Amit and Fusi, 1994). However, in a network with a fixed, finite connectivity one cannot lower f too much, since a selective neuron of a given stimulus will eventually get too few inputs from the other selective neurons in the same stimulus. For a recurrent connectivity of 10 000 and $f \sim 0.01$, εϖερψ νευρον ρεχειϖεσ ~100 potentiated synaptic contacts from other selective neurons, which is a number that may suffice to keep activity going. This is an attractive coincidence since, as was mentioned above, both numbers are biologically plausible.

We may now investigate the effect of learning, represented by the distributions $P_\pm(J)$, on the stable rates sustained by the local module. We consider two situations: (i) no stimulus-related activity in the local network, corresponding to stimulation by an unfamiliar stimulus; and (ii) stimulus-related persistent activity in the local network, corresponding to stimulation by a familiar stimulus. In both cases we consider the situation following the removal of the stimulus, much like in the delay period of delayed response experiments (e.g. Miyashita and Chang, 1988; Wilson *et al.*, 1993).

What makes the problem amenable to detailed solution is the combination of the facts that (i) each neuron's behavior is fully determined by the mean and the variance of the synaptic input to the depolarization at the cell body and (ii) that when there is no sharing of selective neurons between different stimuli, both quantities are uniform in each class of neurons and easily

computable. To determine the stable rate distributions in different stimulation scenarios one computes the mean and the variance of the distribution of the synaptic inputs arriving to a neuron of every type in the local network. Then one solves the equations for self-reproducing rate distributions, i.e. those rates in the different populations of neurons in the network that when acting as inputs to neurons give rise to the same set of rates. Finally, one checks the stability of the solution. The expressions for the means and variances, in terms of the rates in the various populations, are given in Appendix C.

No Stimulus-related Activity in the Local Network. The question is whether following learning a solution exists in which all selective neurons are equivalent and all non-selective neurons are equivalent in terms of rates and inputs. Selective neurons would have equal rates ν_+ and receive equal inputs of mean μ_+ and SD σ_+ . Non-selective neurons would have rates ν_0 and inputs with mean μ_0 and SD σ_0. Inhibitory neurons would have rates ν_I and input with mean μ_I and SD σ_I. The stable rates are given by the system of three coupled equations

$$\nu_+ = \phi_E(\mu_+,\sigma_+) \tag{27}$$

$$\nu_0 = \phi_E(\mu_0,\sigma_0) \tag{28}$$

$$\nu_I = \phi_I(\mu_I,\sigma_I) \tag{29}$$

where $\phi_{E,I}$ are the transduction functions, given by equation (25), for the excitatory and the inhibitory neurons respectively. The means μ_+, μ_0, μ_I and the SDs σ_+, σ_0, σ_I depend in turn on the rates ν_+, ν_0, ν_I via equations (42–46) of Appendix C. This makes the above equations a closed set of equations for the rates in the different populations.

Equations (27–29) are studied numerically for the three stable, self-reproducing rates ν_+, ν_0, ν_I. The parameters are $\tau_I = 0.2\tau_E$, $J_{EI} = 1.5\ J$, $J_{IE} = J_{II}$, $\theta_E \sim 560\ J$, $\theta_I \sim 140\ J_{IE}$. (Recall that the first subscript is for the type of the post-synapstic cell and the second for the pre-synaptic.) Prior to learning $J_+ = J$, there is stable spontaneous activity of $\nu_E = 3\ s^{-1}$ and $\nu_I = 4.2\ s^{-1}$.

Next the solutions of these equations, ν_+, ν_0, ν_I, are studied as a function of J_+/J, the enhancement due to learning of the mean synaptic efficacy. The observed stable rates for the three populations of neurons are plotted as full curves in Figure 6. Note that as J_+/J increases, the stable rate of the selective neurons increases, from 3 s^{-1} at $J_+/J = 1$ (no learning) to 5.5 s^{-1} at $J_+/J = 5$ (the average LTP synaptic value is five times the background one). Correspondingly, there is a decrease of the rates of the non-selective neurons, which become as low as $\nu_0 = 1\ s^{-1}$ when $J_+/J = 5$. The inhibitory neurons have a milder variation of their rates, and stay between 4 and 5 s^{-1}.

We conclude that the spontaneous attractor is robust against the synaptic variations induced by learning, though it acquires some structure on the local level. The spontaneous rates carry some indication of the learning history. It is a precursor of the impending selective attractor. Neurons that had been strongly activated by external stimuli during learning will have enhanced rates, while those that have not been activated at all will have their rates depressed relative to the bare spontaneous activity. Note, however, that the enhanced spontaneous activity is non-selective. In other words, neurons partaking in any of the stimuli will have somewhat elevated spontaneous rates simultaneously, irrespective of the stimulation.

Stimulus-related Activity in the Local Network. We now test

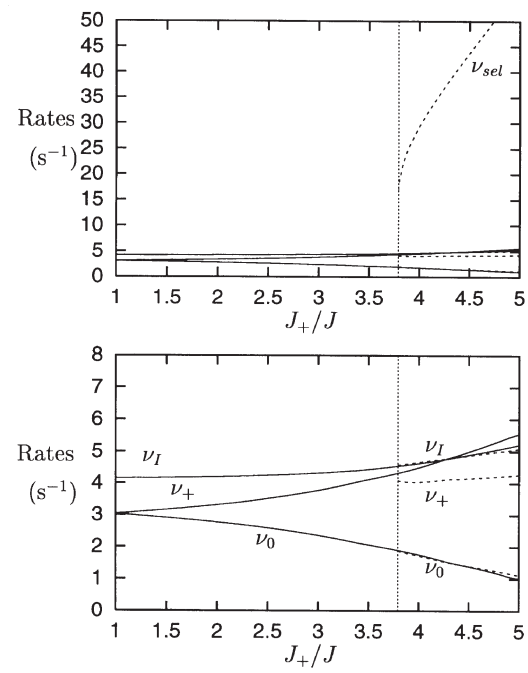

Figure 6. Stable, self-reproducing rates for selective (ν_+ and ν_{sel}), non-selective (ν_0) and inhibitory (ν_I) neurons versus J_+/J, the relative mean synaptic potentiation due to learning. To the left of the dotted vertical line is a single solution (full curves) — spontaneous activity. To the right is an abrupt development of a second stable solution (dashed curves) — selective delay activity. The group of neurons selective for the current stimulus has a higher, selective rate, marked ν_{sel}. The bottom figure is a close-up of the lower part of the top figure. The spontaneous activity, non-selective state, coexists with the selective one to the right of the dotted line. To the left of the dotted line the network sustains the spontaneous state following presentation and removal of any stimulus. To the right of the dotted line, if the stimulus is similar to one of those learned, the network will sustain a selective delay activity following removal of the stimulus. If the stimulus is unfamiliar, the network will return to the spontaneous state.

the option that following the presentation and removal of a stimulus similar to one previously learned the network can sustain a stimulus-selective delay activity. In the simplest case such a selective delay activity state will consist of the subset of neurons responsive to the particular stimulus maintaining an elevated spike rate. In this case there may be four types of rates: of neurons selective of the presented stimulus, ν_{sel}; of neurons selective of other learned stimuli, ν_+ ; of non-selective excitatory neurons, ν_0; and of inhibitory neurons, ν_I.

The stable, self-reproducing rates are determined by the stable solutions of the four coupled equations

$$\nu_{sel} = \phi_E(\mu_{sel},\sigma_{sel}) \tag{30}$$

$$\nu_+ = \phi_E(\mu_+,\sigma_+) \tag{31}$$

$$\nu_0 = \phi_E(\mu_0,\sigma_0) \tag{32}$$

$$\nu_I = \phi_I(\mu_I,\sigma_I) \tag{33}$$

in which the rate in each population is determined by the mean and the variance of the corresponding input, itself given in terms of the four rates. The explicit expressions are given in Appendix C. When J_+ is close to J, the only solution of equations (30–33)

are the spontaneous rates ($\nu_+^*, \nu_0^*, \nu_I^*$), which were found as solutions of equations (27–29) above, i.e.

$$\nu_{sel} = \nu_+ = \nu_+^*, \quad \nu_0 = \nu_0^*, \quad \nu_I = \nu_I^*$$

When J_+/J crosses a critical value (~3.78), equations (30–33) develop abruptly an additional solution over and above the spontaneous rate. The full scenario is displayed in Figure 6, where all the stable frequency solutions are plotted as a function of the relative potentiation due to learning J_+/J. To the left of the dotted vertical line, the critical value of J_+/J, there is a single stable solution for each of the three neuron types: selective, non-selective excitatory and inhibitory. In the delay activity there is no selectivity to any particular stimulus.

Upon crossing the dotted vertical line there is a bifurcation: a new branch appears. Neurons selective to the particular afferent stimulus develop an alternative stable rate, represented by the dashed curve in the top figure marked ν_{sel}. This selective rate is maintained by the learned synaptic structure after removal of the stimulus. It varies in a rather sensitive way with J_+/J, from 17 s^{-1} at $J_+/J \sim 3.78$ to >50 s^{-1} when $J_+/J = 5$. The neurons in the other classes remain at low rates, shown in the dashed curves in the bottom figure, which expands the low rate part of the top figure. For these neurons, the main difference with the unselective attractor is that the rates of neurons selective to other learned stimuli (ν_+) are significantly reduced. Non-selective and inhibitory neurons are hardly affected by the presence or absence of the selective structured activity. This is because selective activity affects only mildly the overall activity in the excitatory network, due to the fact that the coding level is low, and only 1% of the excitatory neurons have significantly higher activity, combined with the fact that the selective activity is not much higher than the spontaneous rates. For $J_+/J \sim 4$, the selective rates are between 20 and 30 s^{-1}, and thus only 4–6 times the spontaneous rates.

At the risk of repetition, we recall that both types of states – spontaneous and selective – are stable. Which of the two is actually manifested in the local module depends on how the module was stimulated. If the stimulation was similar to one of the memories, the high selective branch is maintained after the stimulus is removed. If the stimulation was uncorrelated with any of the previously learned stimuli, the network relaxes to the spontaneous activity state.

This scenario takes place either if single synapses are potentiated gradually or if synapses have only discrete stable states but on each presentation only a fraction of the synapses that should be potentiated are actually potentiated. For example, if each synapse has two stable states, J and J_P, and learning dynamics is stochastic (Amit and Fusi, 1994; Amit and Brunel, 1995), then the resulting distribution of synapses has a mean

$$J_+ = c_P J_P + (1 - c_P) J$$

where c_P is the fraction of actually potentiated synapses out of the total which Hebbian learning would have potentiated had it been deterministic. Thus, if the potentiated efficacy is $J_P/J = 5$, and the critical value for the appearance of the selective attractors is $J_+/J = 3.78$, we need a fraction $c_P = 0.70$ of potentiated synapses to obtain the corresponding attractors.

The above observation could be either a prediction for cortical regions, such as IT or pre-frontal cortex, or a test of the learning dynamics: if LTP takes place gradually, then during the process of training there should appear a phase in which neurons selective of any stimulus have enhanced spontaneous activity. This spontaneous state would be reached after presentation of any stimulus in a delayed match-to-sample experiment. But this enhancement of the spontaneous activity during learning depends on the presumed learning process. In the present picture any selective neuron sees an increase of the average excitatory synaptic efficacy on its dendrites upon learning. This is related to the requirement that the synaptic strength averaged over the whole network remain constant during learning, and to the resulting fact that efficacies of synapses on non-selective neurons decrease. If during learning the average synaptic efficacy remained constant on the dendrites of every individual neuron, the spontaneous activity of neurons would not be modified during learning. At the other extreme, if learning decreases the mean average synaptic strength, because depression is on average stronger than potentiation, the spontaneous activity of selective neurons would decrease with learning. In any case the observation of the evolution of the spontaneous activity during learning would give rather interesting indications on the learning dynamics.

It is rewarding to note that the range of values of both selective and spontaneous rates includes rates observed in experiments in IT cortex (Miyashita, 1988; Miyashita and Chang, 1988; Sakai and Miyashita, 1991) and in prefrontal cortex (Wilson *et al.*, 1993).

Discussion

The present study integrates the structureless spontaneous neural activity into the framework of attractor neural networks. It makes spontaneous activity into another stable collective property of a system of interacting neurons. The particular stable state of spontaneous activity differs from usual attractors previously considered in neural dynamics first by its global nature, second by its extra stability and third by the fact that it can sustain, on top of itself, local structured metastable states, the learned selective delay activity distributions.

Even if the present description does not deflect all criticisms, some of which we will ourselves suggest below, we consider it a significant step forward in the long road to the construction of neurobiologically plausible neural network dynamics, testable and at the same time expressing cognitively related tokens – delay activities in cognitively performing animals. It demonstrates that no external agent is required to maintain this pervasive, low rate activity and that selective delay activity can coexist with spontaneous activity even if the rates in both states are close. It connects levels of spontaneous activity to learning and identifies the appearance of selective delay activity as a bifurcation caused by the quantitative level of learning. To our knowledge this ensemble of properties in a single framework is shown here for the first time.

All the above requires an interplay between excitation and inhibition. But that interplay is rather robust: it is unstructured; it has a wide scope in the space of neural and network parameters; and the functional domain of parameters includes biologically plausible values.

Other models of networks of interacting excitatory and inhibitory neurons have been studied (e.g. Amari, 1972; Wilson and Cowan, 1972). What is new in the system analyzed here is that (i) the dynamical equations governing each population are obtained from a well-defined model of the single neuron, i.e. the integrate-and-fire neuron; (ii) anatomy and physiology are related to realistic cortical parameters; (iii) it exhibits both spontaneous as well as selective states; and (iv) the resulting stable states are

identified and confronted quantitatively with 'states' observed in neurophysiological experiments in performing primates. In fact, even in previous models of networks of spiking neurons exhibiting selective working memory properties of cortical neurons (e.g. Amit *et al.*, 1990; Zipser *et al.*, 1993), the problem of spontaneous activity was not considered.

Tsodyks and Sejnowski (1995) have recently argued that the dynamics of a network of integrate-and-fire neurons cannot be accurately described by a mean-field description because of the importance of the temporal fluctuations of the depolarization in the spike dynamics. Here the mean-field approach was extended to include precisely these fluctuations, in order to get a consistent picture of spontaneous and structured activity in a neuronal module in a cortical environment.

Softky and Koch (1993) have raised the issue of the variability of inter-spike intervals in cortical neurons. They argue that an integrate-and-fire neuron receiving a large stochastic input will have a very regular firing pattern, at odds with experimental evidence in primary visual cortex. However, they considered exclusively the regime in which rates are very high, of the order of 100 s^{-1}. Here the situation is different, since in the stable states we consider firing rates are very low compared with the inverse neuronal time constants. In this regime spikes are due mainly to fluctuations, and thus firing will be naturally very irregular, Poisson-like.

A number of issues still remain open. We start with the seemingly lighter ones. The study has focused on the stationary states of the dynamics and their stability. We have also simulated the current-rate dynamics for a network of 2000 neurons. This network was considered as a local module, the global environment being introduced by hand. The findings confirm the reported results. The main difference between the results of the simulated dynamics and the above analysis is that in the simulation the rates in populations of functionally similar neurons are not uniform. Rather, in each population there is a wide distribution of rates around a mean that is close to the rate determined by the analysis. This is due to the fact that in the simulations we allowed for variability in the connectivity on different neurons. This situation can also be covered by the analysis but for lack of space will be reported elsewhere. These wide rate distributions are, of course, more realistic.

Deeper issues are the full linearity on the input side of our neurons and the assumed relation of the time constants. Concerning the first issue, had we had our choice the scenario would be that when a couple of hundred spikes arrive independently each 10 ms and are scattered randomly on the dendrites, the non-linear cable effects are negligible. In fact, it has been argued (Softky and Koch, 1993; Bernander *et al.*, 1994) that in this situation the simple leaky integrate-and-fire neuron and a complex reconstructed layer V pyramidal cell have similar temporal dynamics.

Even in the presence of non-linearities it may be the case that one can approximate the distribution of depolarization at the soma by a Gaussian, whose first moments can be calculated as a single-valued function of the afferent rates. The calculation then can be redone to obtain the conditions for self-reproduction and stability, though the resulting conditions may be different. Such is the case if the underlying synaptic dynamics includes reversal potentials but no dendritic geometrical structure. We have indications that in that case the system is even more stable, i.e. the acceptable range of parameters expands. Beyond the stabilization of the spontaneous activity, the depolarization due to the learned structured stimuli can be considered marginal and hence linear, in the spirit of Amit and Tsodyks (1992).

Concerning the issue of the finite time constants associated with the synaptic conductance changes, we have preliminary results from analysis and from simulations indicating that their introduction does not have severe effects on the stability of the network.

As mentioned in the introduction, the above analysis holds in the absence of correlations between neuronal spike activities. This question cannot be investigated in the framework of the theory presented here, because the independence of the spike emission processes underlies the conversion from spikes to rates. The issue is being investigated by extensive simulations.

Though not selective, spontaneous activity has some implications for computational phenomena. It has the advantage of placing cortical neurons close to threshold, rather than at their resting potential. Thus the response of a spontaneously active system to a stimulus will be much faster than the one of a system at rest. Furthermore, since neurons responsive to recently learned stimuli have higher spontaneous activity, the response of the system to a learned stimulus will be even faster than it is for unfamiliar stimuli. This suggests that spontaneous activity may play a major role in the speed of processing in cortex.

The network, upon stimulation by a familiar, learned stimulus, goes into the corresponding selective delay activity (attractor) state, rather quickly. In principle, this state can persist for an arbitrary long time, even in absence of the eliciting stimulus. There are several natural ways in which the network may abandon a given selective attractor. It will happen upon a presentation of a strong new stimulus. A new stimulus destabilizes the attractor, and the network will switch from one structured state to another, corresponding to the incoming stimulus. It is the effect of the interplay between excitation and inhibition: the neurons belonging to the foreground of the new stimulus provoke a higher inhibitory activity, which in turn 'turns off' the activity of the neurons belonging to the foreground of the previous stimulus (see also Amit and Brunel, 1995). Another scenario for terminating a delay activity is a variation in the level of non-selective external afferents. In this case the network will tend to go into a spontaneous activity state. This may be a mechanism to implement attention. Finally, if learning has not been strong enough, a network with a relatively low fraction of cells with enhancd rates may leave an attractor due to a fluctuation. It may then find itself in a spontaneous activity state or in another learned attractor. As learning proceeds, the time for such a transition would become longer. This again is an experimental prediction, testable by single unit recordings.

Selective delay activity in various working memory tasks has been observed both in unimodal (e.g. IT cortex for visual stimuli) and multimodal (e.g. prefrontal cortex) association cortices (Fuster, 1995). The behavioral role of delay activity appears rather clear in the experiments on the oculomotor delayed-response task (Goldman-Rakic *et al.*, 1990; Williams and Goldman-Rakic, 1995) as well as in other tasks implying retention of the cue information (Fuster, 1995). The observation by Miyashita (1988) that behavioural performance is the same for new stimuli (for which he finds no delay activity in IT cortex) as it is for learned ones (for which delay activity in IT cortex is present) has raised some question as to the behavioral relevance of delay activity in the delayed match-to-sample tasks. It should be pointed out that in the latter tasks the correlation between the level of delay activity, when present, and the correct

performance has not been measured. On the other hand, when delay activity has not yet been formed, the task may have to be performed by alternative strategies (e.g. Baylis and Rolls, 1987). There may, for example, be other cortical areas where delay activity is learned in a single shot, while delay activity in IT takes a large number of trials to form. Yet somewhere something must keep active the information of the last stimulus in a delay task (e.g. Amit, 1995).

Appendix A: Stability in Network of Simplified Neurons

In this paper $\alpha \sim 2$ (output rates of the order of 1 s^{-1}), one can approximate Pr(ν) (equation 9), by (Abramowitz and Stegun, 1970)

$$\Pr(\nu) \approx \frac{1}{\sqrt{2\pi}\alpha(\nu)} \exp\left(-\frac{\alpha(\nu)^2}{2}\right)$$

with $\alpha = (\theta - \mu)/\sigma$ (equation 5). The equation for the self-reproducing rate is

$$\nu^* \approx \frac{1}{\sqrt{2\pi}\alpha(\nu^*)} \exp\left(-\frac{\alpha(\nu^*)^2}{2}\right)$$

The derivative of Pr(ν) is

$$\frac{d\Pr(\nu)}{d\nu} = -\frac{d\alpha(\nu)}{d\nu}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\alpha(\nu)^2}{2}\right) \approx -\alpha(\nu)\nu \frac{d\alpha(\nu)}{d\nu} \quad (34)$$

and the derivative of α(ν) with respect to ν is given by

$$\frac{d\alpha(\nu)}{d\nu} = -\frac{1}{\sigma(\nu)}\frac{d\mu(\nu)}{d\nu} - \frac{\alpha(\nu)}{2\sigma^2(\nu)}\frac{d\sigma^2(\nu)}{d\nu} \quad (35)$$

To calculate the derivatives of μ and σ^2, we note that both are a sum of two contributions – the local one, which is linear in ν, and the external one, which is fixed (see equations 10 and 11 for excitation only, and equation 19 when inhibition is included). Thus $\nu d\mu/d\nu = \mu_I$, the mean of the local input, and $\nu d\sigma^2/d\nu = \sigma_I^2$, the variance of the local input, and we have from equations (34) and (35)

$$\frac{d\Pr(\nu)}{d\nu} \approx \alpha\frac{\mu_I}{\sigma} + \alpha^2\frac{\sigma_I^2}{\sigma^2} \quad (36)$$

In the absence of inhibition, $\mu_I = xC\nu\tau J$ is of order of a few hundred J, while σ is of order a few tens. On the other hand, the two variances σ_I^2 and σ^2 may differ only by a factor of 2, and $\alpha \sim 2$. So the first term on the right-hand side of equation (36) dominates since $\mu_I >> \sigma$.

Appendix B: Stability in IF Network with Inhibition

The dynamic behavior of the neurons is determined by the mean and the variance of the depolarization. In fact (Tuckwell, 1988), one can obtain the mean interspike interval (the inverse rate) in terms of the mean and variance of the 'free' depolarization, i.e. in the absence of a threshold. In a steady state the mean and the variance are directly deducible from the afferent rates. If excitatory and inhibitory inputs are independent, the mean of the depolarization is the balance of the depolarizing and hyperpolarizing terms. The variance is the sum of the depolarizing and hyperpolarizing input variances (equations 15–18). In this appendix we choose simplifying units, taking the average EPSP on both types of neurons $J_{EE} = J_{IE} = 1$, i.e depolarizations are measured in units of the average EPSP on the corresponding neuron. We take $\tau_E = 1$, so time is measured in units of τ_E. The evolution of the moments $\mu[V_E]$ and $\sigma[V_E]^2$ is given by

$$\partial_t \mu[V_E] = -\mu[V_E] + C_{EE}[x\nu_E + (1-x)\nu_{ext}] - C_{EI}J_{EI}\nu_I \quad (37)$$

and

$$\partial_t(\sigma[V_E]^2) = -2\sigma[V_E]^2 + \lambda(C_{EE}[x\nu_E + (1-x)\nu_{ext}] + C_{EI}J_{EI}^2\nu_I) \quad (38)$$

and we have similar equations for the inhibitory neurons. These are minor extensions of equations 9.146 and 9.147 of (Tuckwell (1988). Note that in equation (7) it is $\sigma_E = \sqrt{2}\sigma[V_E]$ that enters. We rewrite equation (38) in terms of these quantities, e.g.

$$\partial_t(\sigma_E^2) = -2\sigma_E^2 + 2\lambda(C_{EE}[x\nu_E + (1-x)\nu_{ext}] + C_{EI}J_{EI}^2\nu_I) \quad (39)$$

The effect of the perturbation in the rates $(\nu_i) \to (\nu_i + \delta\nu_i)$ on the moments μ and σ is determined from these equations as

$$\partial_t\delta\mu_E = -\delta\mu_E + xC_{EE}\delta\nu_E - C_{EI}J_{EI}\delta\nu_I$$

$$\partial_t\delta\sigma_E = -2\delta\sigma_E + \frac{\lambda}{\sigma_E}(xC_{EE}\delta\nu_E + C_{EI}J_{EI}^2\delta\nu_I)$$

for the excitatory neurons, and for the inhibitory cells

$$\partial_t\delta\mu_I = -\frac{1}{\tau_1}\delta\mu_I + xC_{IE}\delta\nu_E - C_{II}J_{II}\delta\nu_I$$

$$\partial_t\delta\sigma_I = -\frac{2}{\tau_1}\delta\sigma_I + \frac{\lambda}{\sigma_I}(xC_{IE}\delta\nu_E + C_{II}J_{II}^2\delta\nu_I)$$

The shifts in the rates are given via the corresponding ϕ_i of equation (25), i.e.

$$\delta\nu_i = \frac{\partial\phi_i}{\partial\mu_i}\partial\mu_i + \frac{\partial\phi_i}{\delta\sigma_i}\delta\sigma_i$$

taken at the self-reproducing rates satisfying equations (23) and (24). These solutions will be stable if all the eigenvalues of the following matrix have negative real parts

$$\begin{pmatrix} -1+xX_E & xY_E & -J_{EI}X_I & -J_{EI}Y_I \\ xX_E\left(\frac{\lambda}{\sigma_E}\right) & -2+xY_E\left(\frac{\lambda}{\sigma_E}\right) & J_{EI}^2X_I\left(\frac{\lambda}{\sigma_E}\right) & J_{EI}^2Y_I\left(\frac{\lambda}{\sigma_E}\right) \\ xX_E & xY_E & -(1/\tau_I)-J_{II}X_I & -J_{II}Y_I \\ xX_E\left(\frac{\lambda}{\sigma_I}\right) & xY_E\left(\frac{\lambda}{\sigma_I}\right) & J_{II}^2X_I\left(\frac{\lambda}{\sigma_I}\right) & -\left(\frac{2}{\tau_1}\right)J_{II}^2Y_I\left(\frac{\lambda}{\sigma_I}\right) \end{pmatrix} \quad (40)$$

in which

$$X_i = C_i\frac{\partial\phi_i}{\partial\mu};\quad Y_i = C_i\frac{\partial\phi_i}{\partial\sigma};\quad i = E,I$$

In the simplified case $\tau_I = 1$ (equal inhibitory and excitatory time constants) and $J_{EI} = J_{II} = g$ (equal average inhibitory efficacies on both types of neurons), we have $\sigma_E = \sigma_I$, the matrix (40) has two simple eigenvalues, –1 and –2, and the other two are negative if

$$xX_E - gX_I + \left(xX_E + g^2Y_I - xg(g+1)\left[X_EY_I - X_IY_E\right]\right)\frac{\lambda}{2\sigma} < 1 \quad (41)$$

which is the extension of equation (21) in the presence of inhibition. If, for example, $\theta_E = \theta_I$, excitatory and inhibitory transduction functions become equivalent and equation (41) becomes

$$(xC_E - gC_I)\phi'_\mu + \left(xC_E + g^2C_I\right)\phi'_\sigma\frac{\lambda}{2\sigma} < 1$$

In other cases one has to find numerically the eigenvalues of (40).

Appendix C

Since the theory depends only on the mean and the variance of the input to neurons of every class, either for the spontaneous attractor or for selective delay attractors, we provide these quantities in terms of the relevant spike rates.

No Stimulus-selective Activity in the Local Network

The mean recurrent excitatory synaptic input to a selective neuron in its integration time τ is the sum of three contributions: $xC_{EE}fJ_+\nu_+\tau$, from other neurons selective for the same stimulus; $xC_{EE}(p-1)fJ_-\nu_+\tau$, from neurons selective for other stimuli; and $xC_{EE}(1-pf)J_-\nu_0\tau$, from non-selective neurons. After adding the external excitatory afferent (with mean $\mu_{E,\text{ext}}$ and SD $\sigma_{E,\text{ext}}$) and the local inhibition (with mean μ_{EI} and SD σ_{EI}), the total average synaptic input, in an interval corresponding to the integration time of the membrane potential, becomes

$$\mu_+^2 = xC_{EE}\tau\left[\left(fJ_+ + (p-1)fJ_-\right)\nu_+\right] + (1-pf)J_-\nu_0 + \mu_{E,\text{ext}} - \mu_{EI} \quad (42)$$

and its variance is

$$\sigma_+^2 = \lambda xC_{EE}\tau\left[\left(fJ_+^2 + (p-1)fJ_-^2\right)\nu_+ + (1-pf)J_-^2\nu_0\right] + \sigma_{E,\text{ext}}^2 + \sigma_{EI}$$

A non-selective neuron receives $xC_{EE}[pfJ_-\nu_+]\tau$ recurrent excitation from selective cells; $xC_{EE}(1-pf)J\nu_0\tau$ from other non-selective cells. Its total mean input is:

$$\mu_0 = xC_{EE}\tau\left[pfJ_-\nu_+ + (1-pf)J\nu_0\right] + \mu_{E,\text{ext}} - \mu_{EI} \quad (43)$$

and

$$\sigma_0^2\lambda xC_{EE}\tau\left[pfJ_-^2\nu_+ + (1-pf)J^2\nu_0\right] + \sigma_{E,\text{ext}}^2 + \sigma_{EI}^2 \quad (44)$$

An inhibitory unit receives a mean synaptic input of

$$\mu_I = xC_{IE}J_{IE}\tau_I\nu_E + \mu_{I,\text{ext}} - \mu_{II} \quad (45)$$

with variance

$$\sigma_I^2 = \lambda\tau_I xC_{IE}J_{IE}^2\nu_E + \sigma_{I,\text{ext}}^2 + \sigma_{II}^2 \quad (46)$$

where ν_E is the average rate of an excitatory neuron in the local network, i.e.

$$\nu_E = pf\nu_+ + (1-pf)\nu_0$$

Stimulus-selective Activity in the Local Network

The mean synaptic input to a neuron selective to the present stimulus, in an interval corresponding to the integration time τ, is:

$$\mu_{\text{sel}} = xC_{EE}\tau\left[fJ_+\nu_{\text{sel}} + (p-1)fJ_-\nu_+ + (1-pf)J_-\nu_0\right] + \mu_{E,\text{ext}} - \mu_{EI}$$

where the different terms correspond to the synaptic contributions coming from, respectively, the neurons selective to the same stimulus; the other selective neurons; the non-selective excitatory neurons in the module; the input from spontaneous activity external to the module; and the inhibitory input. For the corresponding variance we have

$$\sigma_{\text{sel}}^2 = \lambda xC_{EE}\tau\left[fJ_+^2\nu_{\text{sel}} + (p-1)fJ_-^2\nu_+ + (1-pf)J_-^2\nu_0\right] + \mu_{E,\text{ext}}^2 - \mu_{EI}^2$$

Neurons selective of other stimuli have a mean input:

$$\mu_+ = xC_{EE}\tau\left[fJ_-\nu_{\text{sel}} + f\left[J_+ + (p-2)J_-\right]\nu_+ + (1-pf)J_-\nu_0\right] + \mu_{E,\text{ext}} - \mu_{EI}$$

and a variance

$$\sigma_+^2 = \lambda xC_{EE}\tau\left[fJ_-\nu_{\text{sel}} + f\left[J_+^2 = (p-2)J_-^2\right]\nu_+ + (1-pf)J_-^2\nu_0\right] + \sigma_{E,\text{ext}}^2 - \sigma_{EI}^2$$

The mean synaptic input to a non-selective neuron is:

$$\mu_0 = xC_{EE}\tau\left[fJ_-\left[\nu_{\text{sel}} + (p-1)\nu_+\right] + (1-pf)J\nu_0\right] + \mu_{E,\text{ext}} - \mu_{EI}$$

with variance:

$$\sigma_0^2 = \lambda x C_{EE} \tau \left[f J_-^2 \left[\nu_{\text{sel}} + (p-1)\nu_+ \right] + (1-pf) J \nu_0 \right]$$

$$+ \sigma_{E,\text{ext}}^2 + \sigma_{EI}^2$$

Finally, an inhibitory neuron has its mean input and variance given again by equations (45) and (46), but with ν_E replaced by

$$\nu_E = f[\nu_{\text{sel}} + (p-1)\nu_+] + (1-pf)\nu_0.$$

Notes

We have benefited from discussions with Simone Sergi and Stefano Fusi. We are grateful to Carl van Vreeswijk and Marc Potters for a critical reading of the manuscript, and particularly to Misha Tsodyks for his many detailed comments. We also thank two anonymous referees for their comments which helped to improve significantly the manuscript. This work was supported in part by a grant from the Consiglio Nazionale delle Ricerche and by a Human Capital grant of the EEC. N.B. was supported by Programme Cognisciences, CNRS, France.

Address correspondence to: Daniel J. Amit, Racah Institute of Physics, Hebrew University, Jerusalem, Israel.

References

Abeles M (1991) Corticonics. New York: Cambridge University Press.

Abeles M, Bergman H, Margalit E, Vaadia E (1993) Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. J Neurophysiol 70:1629.

Abramowitz M, Stegun IA (1970) Tables of mathematical functions. New York: Dover Publications.

Amari S (1972) Characteristics of random nets of analog neuron-like elements. IEEE Transact SMC 2 5:55.

Amit DJ (1989) Modeling brain function. New York: Cambridge University Press.

Amit DJ, Brunel N (1995) Learning internal representations in an attractor neural network with analogue neurons. Network 6:359.

Amit DJ, Brunel N, Tsodyks MV (1994) Correlations of cortical Hebbian reverberations: experiment vs theory. J Neurosci 14:6445.

Amit DJ, Evans MR, Abeles M (1990) Attractor neural networks with biological probe neurons. Network 1;381.

Amit DJ, Fusi S (1994) Dynamic learning in neural networks with material sysnapses. Neural Computat 6;957.

Amit DJ, Tsodyks MV (1991) Quantitative study of attractor neural network retrieving at low spike rates I: Substrate – spikes, rates and neuronal gain. Network 2:259.

Amit DJ, Tsodyks MV (1992) Effective neurons and attractor neural networks in cortical environment. Network 3:121.

Amit DJ, Fusi S, Yakovlev V (1996) Paradigmatic working memory (attractor) cell in IT cortex. Neural Computat (in press).

Artola A, Singer W (1993) Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. Trends Neurosci 16:480.

Baylis GC, Rolls ET (1987) Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. Exp Brain Res 65:614.

Bekkers JM, Stevens CF (1989) NMDA and non-NMDA receptors are co-localized at individual excitatory synapses in cultured rat hippocampus. Nature 341:230.

Bernander O, Douglas RJ, Martin KAC, Koch C (1991) Synaptic background activity determines spatio-temporal integration in single pyramidal cells. Proc Natl Acad Sci USA 88:1569.

Bernander O, Koch C, Usher M (1994) The effect of synchronized inputs at the single neuron level. Neural Computat 6:622.

Bliss TVP, Collingridge GL (1993) A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361:31.

Braitenberg V, Schüz A (1991) Anatomy of the cortex. Berlin: Springer-Verlag.

Brunel N (1994) Dynamics of an attractor neural network converting temporal into spatial correlations. Network 5:449.

Connors BW, Gutnick MJ (1990) Intrinsic firing patterns of diverse neocortical neurons. Trends Neurosci 13:99.

Douglas R, Martin K (1991) A functional microcircuit for cat visual cortex. J Physiol (Lond) 440:735.

Frolov AA, Medvedev AV (1986) Substantiation of the 'point approximation' for describing the total activity of the brain with the use of a simulation model. Biofizika 31:304 (Engl transl (1986) Biophysics 31:332).

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331.

Fuster JM (1995) Memory in the cerebral cortex. Cambridge, MA: MIT Press.

Fuster JM, Jervey JP (1981) Inferotemporal neurons distinguish and retain behaviourally relevant features of visual stimuli. J Neurophysiol 36:61.

Goldman-Rakic PS, Funahashi S, Bruce CJ (1990) Neocortical memory circuits. Symposia on Quantitative Biology, Vol LV, p 1025. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hestrin S, Sah P, Nicoll R (1990) Mechanisms generating the time course of dual component excitatory synaptic currents recorded in hippocampal slices. Neuron 5:247.

Koch KW, Fuster JM (1989) Unit activity in monkey parietal cortex related to haptic perception and temporary memory. Exp Brain Res 74:139.

Komatsu Y, Nakajima S, Toyama K, Fetz E (1988) Intracortical connectivity revealed by spike-triggered averaging in slice preparations of cat visual cortex. Brain Res 442:359.

Mason A, Nicoll A, Stratford K (1991) Synaptic transmission between individual pyramidal neurons of the rat visual cortex *in vitro*. J Neurosci 11:72.

McNaughton B, Barnes C, Andersen P (1981) Synaptic efficacy and EPSP summation in granule cells of rat fascia denta studied *in vitro*. J Neurophysiol 46:952.

Miyashita Y (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature 335:817.

Miyashita Y, Chang HS (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature 331:68.

Nakamura K, Kubota K (1995) Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. J Neurophysiol 74:162.

Nelken I (1988) Analysis of the activity of single neurons in stochastic settings. Biol Cybern 59:201.

Rall W (1977) Core conductor theory and cable properties of neurons. In Handbook of physiology (Kandel ER, ed) Vol I. Bethesda, MD: American Physiology Society.

Rapp M, Yarom Y, Segev I (1992), The impact of parallel background activity on the cable properties of cerebellar Purkinje cells. Neural Computat 4:518.

Ricciardi LM (1977) Diffusion processes and related topics on biology. Berlin: Springer-Verlag.

Sakai K, Miyashita Y (1991) Neural organization for the long-term memory of paired associates. Nature 354:152.

Sayer R, Friedlander M, Redman S (1990) The time course and amplitude of EPSPs evoked at synapses between pairs of FP3/FP1 neurons in the hippocampal slice. J Neurosci 10:826.

Shadlen MN, Newsome WT (1994) Noise, neural codes and cortical organization. Curr Opin Neurobiol 4:569.

Simons DJ (1978) Response properties of vibrissa units in rat SI somatosensory neocortex. J Neurophysiol 41:798.

Softky WR (1993) Sub-millisecond coincidence detection in active dentritic trees. Neuroscience 58:13.

Softky WR, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. J Neurosci 13:334.

Stanton PK, Sejnowski TJ (1989) Associative long-term depression in the hippocampus induced by Hebbian covariance. Nature 339:215.

Thompson A, Girdlestone D, West D (1988) Voltage dependent currents prolong single-axon postsynaptic potentials in layer III pyramidal neurons in rat neocortical slices. J Neurophysiol 60:1896.

Traub RD, Miles R (1991) Neuronal networks of the hippocampus. Cambridge: Cambridge University Press.

Tsodyks M, Feigelman MV (1988) The enhanced storage capacity in neural networks with low activity level. Europhys Lett 46:101.

Tsodyks M, Sejnowski T (1995) Rapid state switching in balanced cortical network models, Network 6:111.

Tuckwell HC (1988) Introduction to theoretical neurobiology. Cambridge: Cambridge University Press.
Usher M, Stemmler M, Koch C, Olami Z (1994) Network amplification of local fluctuations causes high spike rate variability, fractal firing patterns and oscillatory local field potentials. Neural Computat 6:795.
Williams GV, Goldman-Rakic PS (1995) Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. Nature 376:572.
Willshaw D, Buneman OP, Longuet-Higgins H (1969) Non-holographic associative memory. Nature 222:960.
Wilson FAW, Scalaidhe SPO, Goldman-Rakic PS (1993) Dissociation of object and spatial processing domains in primate prefrontal cortex. Science 260:1955.
Wilson HR, Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. Biophys J 12:1.
Zador A, Koch C, Brown TH (1990) Biophysical model of a Hebbian synapse. Proc Natl Acad Sci USA 87:6718.
Zipser D, Kehoe B, Littlewort G, Fuster J (1993) A spiking model of short–term active memory. J Neurosci 13:3406.