

Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons

Stefano Panzeri^{1,4}, Edmund T Rolls², Francesco Battaglia³
and Ruth Lavis¹

¹ Department of Psychology, University of Newcastle upon Tyne, Ridley Building,
Newcastle upon Tyne NE1 7RU, UK

² Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK

³ ARL-NSMA-University of Arizona, Life Sciences North Bldg, Tucson, AZ 85724-5115, USA

Received 30 March 2000, in final form 24 May 2001

Published 15 August 2001

Online at stacks.iop.org/Network/12/423

Abstract

The speed of processing in the visual cortical areas can be fast, with for example the latency of neuronal responses increasing by only approximately 10 ms per area in the ventral visual system sequence V1 to V2 to V4 to inferior temporal visual cortex. This has led to the suggestion that rapid visual processing can only be based on the feedforward connections between cortical areas. To test this idea, we investigated the dynamics of information retrieval in multiple layer networks using a four-stage feedforward network modelled with continuous dynamics with integrate-and-fire neurons, and associative synaptic connections between stages with a synaptic time constant of 10 ms. Through the implementation of continuous dynamics, we found latency differences in information retrieval of only 5 ms per layer when local excitation was absent and processing was purely feedforward. However, information latency differences increased significantly when non-associative local excitation was included. We also found that local recurrent excitation through associatively modified synapses can contribute significantly to processing in as little as 15 ms per layer, including the feedforward and local feedback processing. Moreover, and in contrast to purely feed-forward processing, the contribution of local recurrent feedback was useful and approximately this rapid even when retrieval was made difficult by noise. These findings suggest that cortical information processing can benefit from recurrent circuits when the allowed processing time per cortical area is at least 15 ms long.

⁴ Corresponding author.

1. Introduction

The speed of processing is a fundamental constraint that must be incorporated into computational models of vision and of cortical processing in general. An analysis of response latencies indicates that there is sufficient time for only 10–20 ms per processing stage in the visual system. In the primate cortical ventral visual system the response latency difference between neurons in layer $IVC\beta$ of V1 and inferior temporal cortical cells is approximately 60 ms [1–3]. For example, the latency of the responses of neurons in V1 is approximately 40 ms [4], and in the temporal cortex visual areas approximately 80–110 ms [5, 6]. Given that there are approximately six stages of processing in the ventral visual system from V1 to the anterior inferior temporal cortex, the difference in latencies between each ventral cortical stage is on this basis approximately 10 ms [7]. Information theoretic analysis of the responses of single visual cortical cells in primates reveal that much of the information that can be extracted from neuronal spike trains is often found to be present in periods as short as 20–30 ms [8–10]. Backward masking experiments indicate that each cortical area needs to fire for only 20–30 ms to pass information to the next stage [10, 11]. Rapid serial visual presentation of image sequences shows that cells in the temporal visual cortex are still face selective when faces are presented at the rate of 14 ms/image [12]. Finally, event-related potential studies in humans provide strong evidence that the visual system is able to complete some analyses of complex scenes in less than 150 ms [13].

Drawing together evidence of this type, several authors were led to conclude that such rapid visual processing must be based *almost entirely on a feedforward* mechanism [7, 13–17]. In their view, since the signal has to pass so rapidly through the large number of hierarchically organised cortical processing stages involved in primate vision, the information could only flow through the unidirectional chain formed by the synaptic relays connecting one area to the next. They suggest that information flow through feedback loops based on recurrent connections within each cortical area would involve an additional number of synaptic steps in each area, and thus recurrent connections could not contribute to such rapid visual processing. However, the intuition that recurrent processing is too slow is mainly fostered by results obtained with early models of autoassociative attractor networks. These networks use associative learning in the recurrent collateral connections between the neurons to store information, and to retrieve it through feedback processing [18–20]. In these early models the dynamics of retrieval were usually simulated with discrete timesteps, with one timestep for each feedback cycle and simultaneous update of the outputs of the neurons, each of which represented some short-time average of the neuron's firing rate. With these simulations of associative networks, typically several processing timesteps are needed for the network to perform pattern retrieval through the recurrent processing, and to begin to settle into a dynamical attractor corresponding to a correct memory retrieval state [21]. Discrete timestep recurrent processing of this type is therefore far too slow to contribute to rapid visual cognition, given that a biophysical estimate of the time for a discrete step of the artificial model would be in the order of the neuronal membrane time constant or the typical inter-spike interval (both of which are in the 10–30 ms range [22]).

However, work done in recent years has shown that recurrent processing in the brain might be considerably faster than that of early models with discrete dynamics. In fact, it has been shown that more realistic models of recurrent networks of spiking neurons incorporating the dynamics of membrane potential and synaptic conductance changes in continuous time can process information much faster than the discrete timestep models. In particular, networks of leaky integrate-and-fire neurons with recurrent connections can respond very quickly to a change in the input current [23, 24]. If the recurrent collaterals are endowed with associative learning, the network can retrieve information about patterns that are stored in the associative

synapses within the order of 1–3 time constants of the recurrent synapses (that is within 10–30 ms if excitatory synaptic transmission during retrieval is mediated by AMPA receptors with an inactivation time constant of 10 ms) [25, 26]. The way in which networks with continuous dynamics (such as networks made of real neurons in the brain, and networks modelled with integrate-and-fire neurons) can be conceptualized as settling so fast into their attractor states is that spontaneous activity in the network ensures that some neurons are close to their firing threshold when the retrieval cue is presented, so that the firing of these neurons is influenced within a few ms by the retrieval cue. These neurons then influence other neurons through the modified recurrent collateral synapses which store the information within ms (given the point that some other neurons will be close to threshold). In this way, the neurons in networks with continuous dynamics can influence each other within a fraction of the synaptic time constant, and retrieval can be very rapid [21]. This work challenges the hypothesis that only feedforward processing is relevant for fast visual cognition, and provides a foundation for understanding rapid cortical information processing.

The studies of the continuous time dynamics of networks of spiking neurons described above focused mainly on single networks. A full understanding of the contribution of recurrent collaterals to fast visual processing would also require a study of the dynamics of hierarchically organized networks, in which the signal is transmitted between areas by feedforward connections, and is at the same time processed within each area by recurrent connections. In fact, there are aspects of the dynamics of a network with a complex architecture, such as the interaction between recurrent and feedforward processing when the signal from the previous area is poor, that may be overlooked or difficult to address when focusing on single modules. We therefore present in this paper a simulation study of the dynamics of a four-layer network of integrate-and-fire neurons, with associative feedforward connections between each layer and the next, and with excitatory recurrent collateral connections in each layer, as well as local inhibition. Each layer models a single cortical area in the visual hierarchy. Since it is of interest to test how rapidly multistage systems with different functional architectures can transmit information, we compared the dynamics of information retrieval in cases where the recurrent connections are associative, to situations in which the recurrent collaterals are either removed or are uniform producing purely feedforward information processing. The main finding was that the processing time of the multilayer system can be as fast as approximately 15 ms per layer even when feedback information processing is required within each layer. Purely feedforward processing can be much faster than if the signal provided to each layer is information-rich, that is in low-noise conditions. These results provide a foundation for understanding the rapidity with which feedback processing within each layer of a multilayer system can contribute to rapid information processing. They also emphasize the importance of the feedback processing when the signals to each stage are noisy.

2. Methods

2.1. Network structure and single neuron models

The model consisted of four layers of excitatory and inhibitory populations (see figure 1). Excitatory neurons within each layer were connected by excitatory recurrent collateral synaptic connections, implementing local feedback and denoted as RC in the following. The RC synapses were set up by associative (Hebbian) learning in some of the simulations. The excitatory populations within one layer were connected to the excitatory cells in the next layer by Hebbian feedforward (FF in the following) projections. Non-associative inhibition within each layer was also provided.

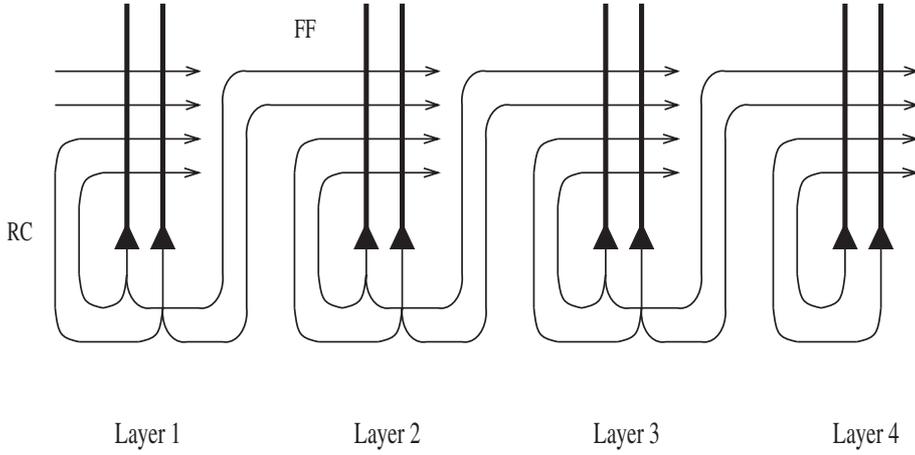


Figure 1. The structure of the excitatory connections in the network. There are feedforward (FF) connections between each layer and the next, and excitatory recurrent collaterals (RC) in each layer. Inhibitory connections are also present within each layer, but they are not plotted in this figure.

Each layer was composed of $N_E = 800$ excitatory units, and 400 inhibitory units, divided into $N_S = 200$ ‘shunting’ inhibitory units and $N_H = 200$ ‘hyperpolarizing’ inhibitory units. Each unit was modelled as a leaky integrate-and-fire device, and it represented a neuron as a single branch, compartmented, dendrite, and a point-like soma where spikes were generated. (The neuronal model is similar to that of [26].) The compartmental model was used in order to implement shunting inhibition, which is necessary for stable pattern retrieval in a recurrent system [26]. The current flowing from each compartment to the external medium was expressed as

$$I(t) = g_{\text{leak}}(V(t) - V^0) + \sum_j g_j(t)(V(t) - V_j), \quad (1)$$

where g_{leak} is a constant passive leakage conductance, V^0 the membrane resting potential, $g_j(t)$ the value of the j th synapse conductance at time t , and V_j the reversal potential of the j th synapse. $V(t)$ is the potential in the compartment at time t . A list of all the parameter values used is reported in table 1. We note here that the RC integration time constant of the membrane of excitatory cells was 20 ms long for the simulations presented. The synaptic conductances decayed exponentially in time, obeying the equation

$$\frac{dg_j}{dt} = -\frac{g_j}{\tau_j} + \Delta g_j \sum_k \delta(t - \Delta t - t_k^j), \quad (2)$$

where τ_j is the synaptic decay time constant, Δt is a delay term summarizing axonal and synaptic delays, and Δg_j is the amount the conductance is increased when the presynaptic unit fires a spike. Δg_j thus represented the (unidirectional) coupling strength between the pre-synaptic and the post-synaptic cell. t_k^j is the time at which the pre-synaptic unit fires its k th spike. For consistency with the single layer study of [26], we decided to present mainly results of simulations not including an axonal conduction delay and using a slightly conservative value of 10 ms for all the excitatory synapses⁵. This was to make sure that the time course

⁵ The inactivation time constant of AMPA channels is generally reported to be in the range 5–8 ms [27].

Table 1. Parameters used in the simulations. Parameter values were taken from [22,26]. Inhibitory parameters were then adjusted in order to get similar mean firing levels across the different conditions tested (see text). The symbols †, ¶, and § are placed near parameter values that are used only in simulations of the ‘Hebbian RC’, ‘no RC’, and ‘Hom. RC’ type respectively. Note that the index E in the parameter symbols refers to excitatory units, the index S indicates shunting units, and the index H is used for indicating hyperpolarizing units. Voltage parameters in this table are relative to the resting value of the neuronal membrane potential, which was set to zero. Parameters reported here correspond to the ‘no delay’ simulations. Parameters changes for the ‘delay’ case are discussed in the main text.

Quantity	Symbol	Value
# Excitatory cells	N_E	800
# Hyperpolarizing Inh. cells	N_H	200
# Shunting Inh. cells	N_S	200
Corruption level	ρ	0.3
# Dendritic compartments	N_{cmp}	10
Dendritic compartment leakage conductance	G_0^d	6.28×10^{-14} (S)
Somatic compartment leakage conductance	G_0^s	5×10^{-9} (S)
Dendritic–dendritic axial conductance	G_0^{dd}	2.25×10^{-7} (S)
Exc. somatic capacitance	$C_{\text{som},E}$	1×10^{-10} (F)
Inh. somatic capacitance	$C_{\text{som},S}, C_{\text{som},H}$	5×10^{-12} (F)
Cue current	I_{cue}	0.25 (nA)
Firing threshold potential (exc.)	Θ_E	32 (mV)
Firing threshold potential (inh.)	Θ_S, Θ_H	25 (mV)
After-spike hyperpolarization potential	V_{ahp}	−15 (mV)
Excitatory–excitatory connectivity level	C_{EE}	1
Excitatory–inhibitory connectivity level	C_{ES}, C_{EH}	0.25
Inhibitory–excitatory connectivity level	C_{SE}, C_{HE}	0.25
Inhibitory–inhibitory connectivity level	C_{SS}, C_{HH}	0.5
Feed–forward connectivity level	C_{FF}	0.5
Exc.–exc. syn. conductance par. (Hebb RC)	g_{EE}	$1.2 \times 10^{-7} \dagger$ (S)
FF Hebbian Syn. conductance par	g_{FF}	1.2×10^{-7} (S)
Exc.–exc. synaptic conductance (Hom. RC)	Δg_{EE}	$0 \ddagger; 4.1 \times 10^{-10} \S$ (S)
Exc.–inh. synaptic conductance	$\Delta g_{ES}, \Delta g_{EH}$	4×10^{-9} (S)
Shunting–exc. synaptic conductance	Δg_{SE}	1.42×10^{-8} (S)
Hyperpolarizing–exc. synaptic conductance	Δg_{HE}	$1.5 \times 10^{-8} \dagger \ddagger; 10^{-13} \S$ (S)
Inh.–inh. synaptic conductance	$\Delta g_{SS}, \Delta g_{HH}$	9×10^{-10} (S)
Exc. synaptic reversal potential	V_E	65 (mV)
Shunting synaptic reversal potential	V_S	0 (mV)
Hyperpolarizing synaptic reversal potential	V_H	−40 (mV)
Exc.–exc. synaptic time constant	τ_{EE}	10 (ms)
Exc.–inh. synaptic time constant	τ_{ES}, τ_{EH}	1 (ms)
Inh.–exc. synaptic time constant	τ_{SE}, τ_{HE}	1 (ms)
Inh.–inh. synaptic time constant	τ_{SS}, τ_{HH}	1 (ms)

reproduced by the model was not artificially fast⁶. For other control simulations (see figure 6), we set the excitatory inactivation time constant to the more realistic value of 7 ms and included a transmission delay $\Delta t = 3$ ms. The reason for choosing the same value of 3 ms for both RC and FF axonal conduction delays is that local excitatory axons, though shorter, have in general

⁶ Note that in the mean field approximation the retrieval time near the attractor settling is proportional to the synaptic inactivation time constant [25]. This was verified by simulations for a single layer network in [26] and was fully confirmed in the FF case in this study (data not shown).

slower conduction speed than between-area connections [2]. The inhibitory system was set to work faster than the excitatory one, as this was helpful for preventing destabilization of the attractor retrieval dynamics due to neuronal synchronization (see [28] for detailed studies of regimes of inhibitory couplings leading to synchronization).

For each timestep of 1 ms, the cable equation for the dendrite was integrated [22] with a finer time resolution of 0.1 ms, and the somatic potential was compared with the spiking threshold V^{thr} . When this was exceeded, the post-synaptic conductances were updated and the somatic potential was reset to the after-hyperpolarization value V^{ahp} throughout the neuron.

The intralayer connections from excitatory to inhibitory, from inhibitory to excitatory, and between inhibitory units were taken to be homogeneous, that is, all of the same strength. The intralayer connectivity was 0.25 between excitatory and inhibitory populations, and 0.5 within each inhibitory population. There were no connections between shunting and hyperpolarizing cells. To build the connections, each cell was connected to a fraction of the cells of the receiving population chosen at random according to the level of connectivity. In contrast, the excitatory units within a layer were all connected to each other. This high connectivity was necessary to produce sufficient statistical averaging in the synaptic input to each unit because of the small size of the simulated network. The connectivity of the FF excitatory connection between excitatory units of successive layers was set to 0.5. This was to take into account the fact that cortical cells receive more excitatory inputs from neurons in the same region than from axons coming from more distant parts of the cortex.

Both RC and FF excitatory-to-excitatory connections encoded in their strength p memorized patterns of activity $\eta_i^{L;\mu}$, consisting of binary words (with elements 0 or 1), the fraction of active units in the pattern being $a = 0.1$. The p patterns (indexed by $\mu = 1, \dots, p$) $\eta_i^{L;\mu}$ to be stored in each layer were chosen at random independently for each layer L . Encoding was implemented through a Hebb rule as follows, with a procedure ensuring that the synaptic weights (conductances g in the model) are always positive. All conductances were initially set to zero and then, for each pattern, the synapse from the i th to the j th unit was modified by an (associative, Hebbian) covariance term, which is reported below for the RC and FF case (see table 1 for the values of the parameters):

$$\Delta g^{\text{RC}} = \frac{g_{\text{EE}}}{pN_{\text{E}}} \left(\frac{\eta_i^{L;\mu}}{a} - 1 \right) \left(\frac{\eta_j^{L;\mu}}{a} - 1 \right), \quad (3)$$

$$\Delta g^{\text{FF}} = \frac{g_{\text{FF}}}{pN_{\text{E}}} \left[\left(\frac{\eta_i^{L;\mu}}{a} - 1 \right) \left(\frac{\eta_j^{L+1;\mu}}{a} - 1 \right) + \sigma_{\text{FF}} \delta_{\text{FF}} \right], \quad (4)$$

δ_{FF} in the equation above is a noise term of zero mean and uniformly distributed between -0.5 and 0.5 . It was used to modify the information content of the FF (signal) input to each layer. Both δ_{FF} and σ_{FF} are dimensionless quantities that modify the pattern-dependent term multiplying the synaptic conductance term⁷. If, after summing each pattern, the synaptic conductance became negative, it was reset to zero. Memories were therefore stored through a ‘random walk with one reflecting barrier’ procedure. The barrier acted as a ‘forgetting’ mechanism [26,29], as whenever the conductance value bumped into the barrier, it lost memory about the previously presented patterns. For computational reasons, the network was tested at low memory loading ($p = 10$).

In order to test the speed of processing in the purely FF case, the RC connections were switched off in some of the simulations by setting $\Delta g^{\text{RC}} = 0$. The inhibition was reduced accordingly to compensate for the decreased level of intralayer excitation, thus achieving

⁷ To get a feeling of the meaning of σ_{FF} values, one should compare them with the magnitude of synaptic strength changes enforced by each pattern (taking into account that a was set to 0.1 in our simulations).

similar mean rates as in the associative (Hebbian) RC case. In another control condition, used to test the specific effect of the plasticity of RC, we set all the conductances of the recurrent collateral synapses to a constant value, chosen again to achieve, in each layer, the same mean rates as in the case of Hebbian synapses.

The excitatory synapses impinged on the distal end compartment of the post-synaptic dendrite, and they had a positive reversal potential (with respect to the resting membrane potential). Inhibitory synapses of the shunting type were distributed uniformly along the dendritic body, and they had a reversal potential equal to the resting membrane potential. Inhibition of this type leads to a mainly divisive effect on the post-synaptic firing rate [30]. Inhibitory synapses of the hyperpolarizing type were instead located at the distal end of the dendritic tree, colocalized with excitatory inputs, and had a reversal potential lower than the resting membrane potential. This last type of inhibition is predominantly subtractive in nature.

2.2. Simulation protocol

Once the connection matrix was constructed, a test of the retrieval dynamics was performed according to the following protocol:

2.2.1. Layer 1 (L1) Layer 1 was a recurrent network responding to a cue current, and was used to set the stage for retrieval in successive layers. L1 provided a pattern-specific cue signal, which initiated retrieval in the successive layers. The structure of L1 was essentially identical to the autoassociative network studied in [26], to which we refer for full details. In brief, in the first 100 ms (for $-100 \leq t < 0$, t being the simulation time), a current was injected into a random fraction $a = 0.1$ of the units of all types. This generated non-specific activity, modelling the effects of spontaneous firing or of a previous visual stimulus. At simulation time $t = 0$ ms, the retrieval process was started by replacing the random current with a *cue* current, injected in a fraction $a + \varrho(1 - a)$ of the L1 units active in the pattern being tested and in a fraction $a(1 - \varrho)$ of the units inactive in the pattern. ϱ was the average correlation between pattern and cue. It was set to 0.3 for all the simulations presented. The cue current lasted for another 300 ms.

2.2.2. Successive layers (L2–L4) Each of the successive layers (L2–L4) was activated by the incoming spikes, received by the previous layers through FF excitatory projections. As for L1, spontaneous activity was generated initially (only for $-100 \leq t < 0$) by injecting a current in a random fraction $a = 0.1$ of the units. Then another 300 ms of simulation followed, in which the network processed the pattern-specific cue received by the previous layer.

2.3. Quantification of the performance of retrieval

The time course of processing and of retrieval of the stored patterns was quantified by using the mutual information about which pattern was presented in L1 that was carried at different times by the responses of small populations of excitatory neurons in each layer. There are several advantages in using an information theoretic measure. First, mutual information is the only measure of correlation that is consistent with simple and plausible requirements [31]. Second, mutual information naturally quantifies how significant the differences in the responses to different stimuli are with respect to variability or spontaneous activity. Third, unlike simpler measures such as pattern overlap, the same information measure can be applied to neurophysiological recordings of small neuronal populations [32, 33], thereby facilitating the comparison between theory and experiment. The mutual information was computed as follows.

The retrieval simulation protocol was repeated for 25 ‘trials’ for each stored memory. A population of 100 randomly selected excitatory units was sampled for the number of spikes they fired in a 10 ms long time window. The population size and the time window length were limited to the values reported because the information measure used here may suffer problems for too large populations or too small time windows, see [33]. The window was moved across the entire simulated time course with steps of 2 ms. The population firing rate vector was constructed at every timestep of each trial and was *decoded*. Decoding consisted in [32] predicting the pattern presented on each trial from the average response and its distribution across all the patterns. The predicted memory pattern s^p was the one whose mean response was closest (using a ‘Euclidean distance’ [33] measure in the response space) to the response on the particular trial. The result of decoding all the trials was a probability table $Q(s^p|s)$ containing the relation between the actual cue pattern s that had been presented on a trial and the cue pattern s^p that was predicted to have been presented by measuring the neuronal response. The *mutual information* I between the actual and decoded pattern, was calculated as follows:

$$I(s, s^p) = \sum_s \frac{1}{p} \sum_{s^p} Q(s^p|s) \log_2 \frac{Q(s^p|s)}{Q(s^p)}. \quad (5)$$

A term correcting for the limited number of trials was then subtracted from equation (5), as explained in [34]. We verified that the number of trials used was sufficient to obtain solid information estimation as follows. First, we made sure that the number of trials per stimulus was larger than the number of stimulus classes. This condition, as shown by computer simulations in [34], ensures that the bias corrections are reasonably accurate. Second, we replicated the control analysis introduced in [6]. Namely, we recomputed the information values using random subsets of 15 and 20 trials respectively. We verified that, *after* subtracting the bias correction, the results obtained using fewer trials were not statistically correlated with the number of trials. In the few cases where there were differences with the number of trials, the differences were in the range of the timestep with which the sliding window was moved.

An interesting feature in using such a decoding step to compute the information is that the same decoding procedure can also be performed, in principle, by a downstream neuron receiving the population responses through appropriate synaptic weights [33]. All the information extracted in this way is thus likely to be readable by downstream cells.

3. Results

The aim of this study is to understand the dynamics of information processing in multiple layer networks with both feedforward and local recurrent connections, and in particular to investigate the speed of the specific RC contribution. For this purpose, it is useful to compare the information processing in three different conditions. In the first condition (‘no RC’), the RC in L2–L4 are switched off, and therefore the propagation of the FF signal does not have to compete against any non-specific excitatory activity in each layer.

In the other two conditions, the FF inputs have instead to compete against local excitation. In the ‘Hebbian RC’ case, the RC in L2–L4 were trained with Hebbian learning of the patterns, and therefore they contribute to processing by making the local excitation pattern-specific.

In the last condition (‘homogeneous RC’), all RC synapses in L2–L4 are homogeneous, i.e. set to the same value which is the average of the values when trained with the Hebb rule. The information processing in this case is still all FF, as no information is stored in the RC. The RC provide in this case only non-pattern-specific activation. This type of activation is present in the inputs to a cortical neuron (as a result of spontaneous firing or of a previous

visual stimulus) at the time when the neuron is reached by the FF signal coming from the previous area, even if the hypothesis is valid that the RC would not have time to contribute to processing. Therefore this is an important control condition both to estimate realistically the speed of pure FF processing, and to understand over what time scale the RC are able to refine retrieval by modifying the local excitation depending on the signal coming from the preceding stage⁸.

We also investigated the effect of the amount of information contained in the FF signal coming from the previous layer. It is intuitive that it takes more time to process and retrieve a pattern when the cue signal is less rich or is noisier. However, the RC contribution may become important at different speeds depending on the signal quality, and we wished to investigate this.

We start by showing in figure 2(a) the speed of operation of L1. L1 operated in all the simulations in the same way as a recurrent autoassociative network responding to a cue current injected at $t = 0$ ms, so that the signal received and processed by L2–L4 was the same across conditions. L1 dynamics were studied in detail in [26]. The information about the cue becomes very significant within a time in the order of 10–20 ms from cue injection. Its time course indicates when L2 starts to receive useful inputs. For the measure of information latencies, an information value of $I = 0.5$ bits was arbitrarily chosen⁹. For L1, 0.5 bits of useful information about which pattern was presented can be read off from the firing of the neurons within 10 ms of cue injection. All the latency measures in successive layers must be interpreted in the knowledge of this reference L1 information latency value.

We begin the study of the propagation of the signal in the successive layers (L2–L4) by first considering the case in which the cue signal provided by the FF connections is information rich. This is achieved when there is little noise in the FF connections ($\sigma_{FF} = 10$), and is reported in figure 2(b)–(d). We examine first the ‘no RC’ condition, where the FF signal does not have to compete against any non-specific local excitatory activity. It is evident that in this case the FF associative connections alone are able to account for good pattern discrimination. The information plateaus are, for all layers, not far from the information level for perfect pattern recognition (i.e. $I = 3.3$ bits). The information latencies in the no RC condition are 16 ms for L2, 22 ms for L3, and 30 ms for L4. Therefore, useful information is passed between three layers in approximately 20 ms, corresponding to 6–7 ms per layer.

We examine now the propagation of the signal when the FF inputs have to compete with local excitation. When the RC are Hebbian, there is a delay of about 10 ms per layer in information latencies with respect to the ‘no RC’ case (see figures 2(b)–(d)). The total delay across 3 layers is 44 ms. This happens because the FF input meets non-specific local excitation, and therefore more time is needed to remove the effect of this noise. (Note that, given the connectivity values used here, the local recurrent excitation is twice as big as the FF inputs.) However, having been trained with the Hebb rule, the RC are able to make the local excitation more pattern specific using the local attractor dynamics, and thus the information can reach a slightly higher level of pattern discrimination than in the ‘no RC’ case. In this condition the information latency is longer than in the ‘no RC’ case. Thus the speed of information retrieval in the pure FF paradigm depends on the amount of local non-specific excitation.

⁸ We stress that we have set the inhibitory parameters such that the mean firing rates are as similar as possible across the simulations to be compared and across all the four layers.

⁹ The value of 0.5 bits of information about which stimulus was presented is a useful point at which to measure the latency of retrieval because 0.5 bits is a much higher amount than random fluctuations of information values, which corresponded to the amount of information computed in the pre-cue time and were in the range of 0–0.05 bits. Also, 0.5 bits was still far from the $\log 2(10) = 3.3$ bits necessary to discriminate perfectly the ten patterns, and therefore latency measures were not distorted by ‘ceiling’ effects.

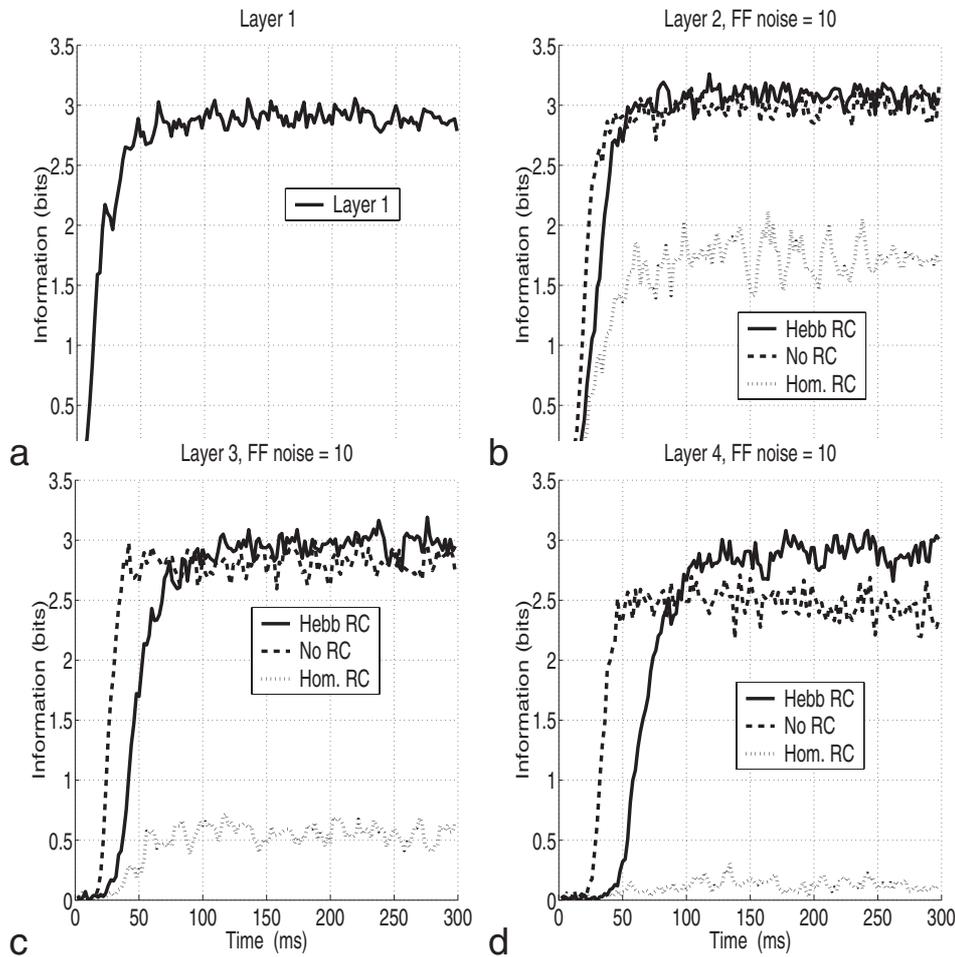


Figure 2. The information time course of the average information carried by the responses of a population of 100 excitatory neurons in each layer. In the simulations considered here there is little noise in the FF synapses ($\sigma_{FF} = 10$). Layer 1 is tested in just one condition. Layers 2–4 are tested in three different conditions: no RC, Hebbian RC, homogeneous RC (a) Layer 1; (b) layer 2; (c) layer 3; (d) layer 4.

Now we consider the ‘homogeneous RC’ case. The information latency is slower than in the Hebbian RC case, and the total information reached is much lower, as shown in figure 2(d). For example, already at $t = 50$ ms, considerably more information is available in L4 when the RC are Hebbian as compared with homogeneous. Therefore we can conclude that by 50 ms the associative processing in the RCs is already improving the quality of retrieval. Taking into account the L1 latency, this corresponds to less than 15 ms per layer.

It is also important, however, to understand the dynamics of the system when the signal provided by the preceding stage is poorer, so that the recurrent collateral attractor processing becomes more critical for good retrieval. Results on this are shown in figure 3. The difference with respect to figure 2 is that now during training a larger amount of noise is included in the FF projections ($\sigma_{FF} = 21$). Figure 3 shows that, when the FF carry an incomplete signal, some information is still transmitted successfully in the ‘no RC’ condition, and with a relatively

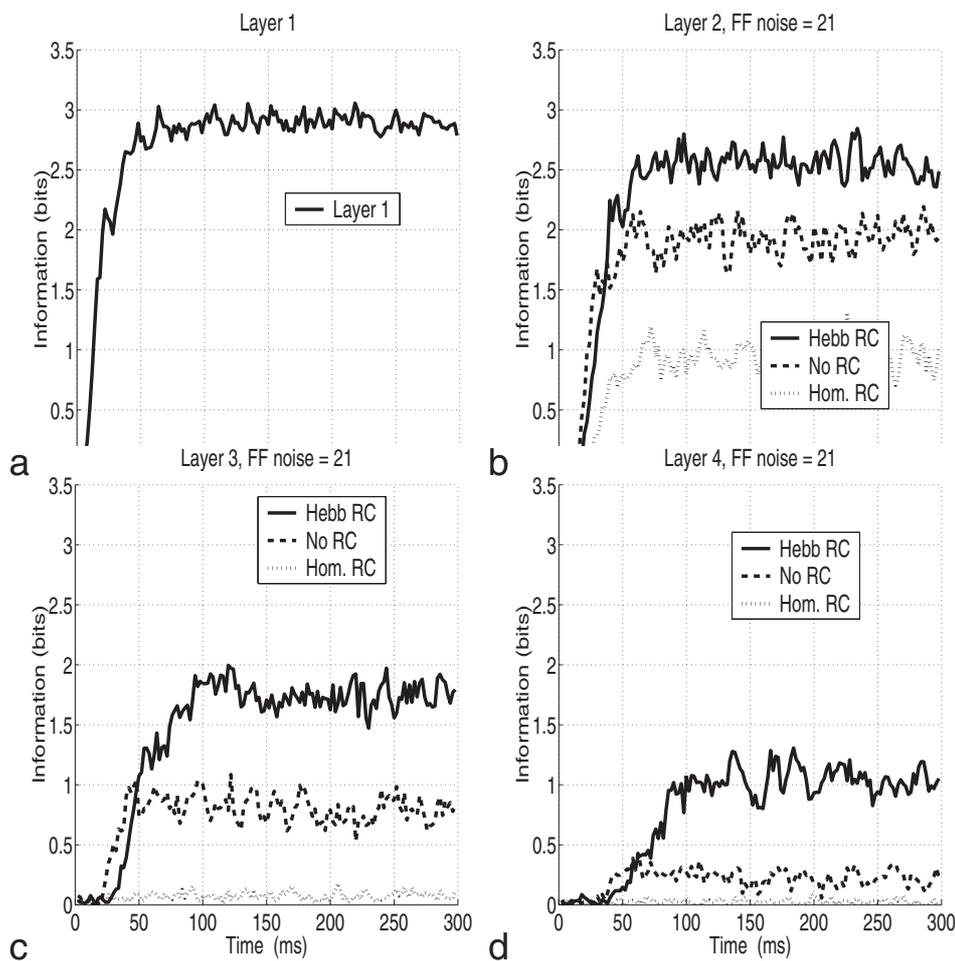


Figure 3. The information time course in each layer when there is much more noise in the FF synapses ($\sigma_{FF} = 21$). Symbols as in figure 2 (a) Layer 1; (b) layer 2; (c) layer 3; (d) layer 4.

short latency. However, the noise term in the FF synaptic strengths makes the retrieval fail more and more layer by layer.

When the RC are present and operating after Hebbian training, the amount of information retrieved is now much higher, because the RC are able to correct a good part of the erroneous information injected into the neurons by the noisy FF synapses. In L4, 66 ms after cue injection in L1, the information in the Hebbian RC case is 0.2 bits higher than that provided by the FF connections in the 'no RC' condition. This shows that the RC are able to retrieve information in L4 that is not available by any other pure FF mechanism after only roughly 50–55 ms from the time when L1 responds. (This corresponds to 17–18 ms per layer.) When the recurrent excitation is uniform, the retrieved information is significantly less than in the corresponding Hebbian RC case for each layer and almost every timestep. This indicates that the RC are, in this case, almost immediately effective in turning the local excitation from non-specific to specific.

We now consider the time course of the neuronal firing, since its study can help understand the mechanisms of information retrieval. Examples of the network activity in L2 and L4 are

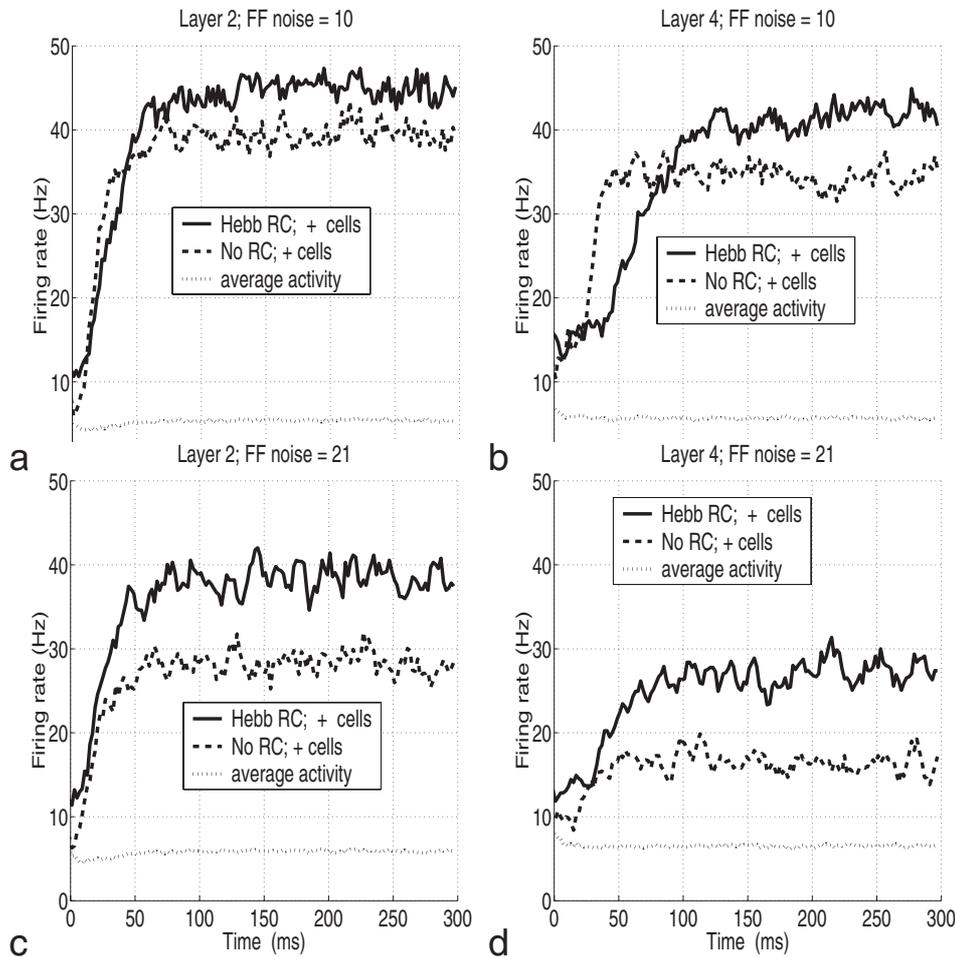


Figure 4. Time course of the mean firing rate of the excitatory neurons in the network. The average activity of all neurons (obtained in the Hebbian RC case) is plotted against the average activity of all the neurons that should be active in the pattern (indicated as ‘+’ cells in the legends). The mean activity in the ‘no RC’ case (not plotted) is very close to the mean activity in the RC case. (a) Layer 2, $\sigma_{FF} = 10$; (b) layer 4, $\sigma_{FF} = 10$; (c) layer 2, $\sigma_{FF} = 21$; (d) layer 4, $\sigma_{FF} = 21$.

reported in figure 4. The noise values are the same as those used for the information results plotted in figure 2 and 3. The mean firing rates of the excitatory neurons in the network were computed averaging over all trials and over all the excitatory cells and using a 2 ms long sliding integration window. Since the inhibitory parameters were set to make the mean activity of excitatory neurons as constant as possible across conditions, we report the mean activity level only for the RC case (the mean activity in the ‘no RC’ case is similar within 5% for all timesteps and simulations). We then selected, for each pattern presentation, what we called the ‘+’ neurons, i.e. the neurons that were active in the stored pattern (and that should hence be firing during correct pattern retrieval), and we averaged their activity separately. The firing rates of ‘+’ neurons gradually increases above the baseline activity. The time course of increase of rate of the neurons active in the pattern is clearly related to the information time course. In particular, 0.5 bits of information are reached when the mean activity of ‘+’ neurons

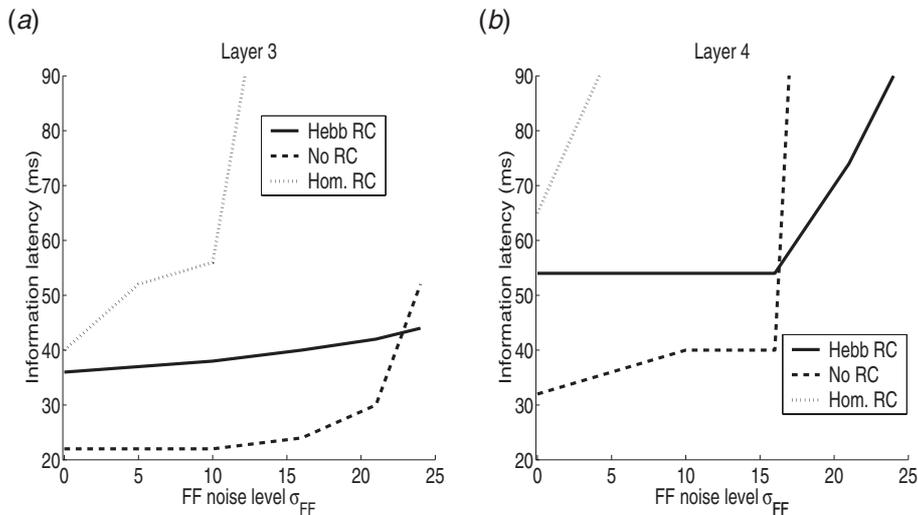


Figure 5. Response latencies are plotted for different values of the noise in the FF synaptic strengths. (a) Layer 3 latencies, (b) layer 4 latencies.

is higher than approximately 22–23 Hz, i.e. at this rate there are enough spikes from ‘+’ neurons to start making the pattern identification more correct. Note also that the plateau of information values reached for long times correlates to the plateaus reached by the firing rates of ‘+’ neurons—the larger the separation between the rate of ‘+’ neurons and the average, the higher the information reached¹⁰. Hence the firing rate analysis confirms the timescales found with the information analysis.

The response latencies in L3 and L4 for different values of the noise in the FF synaptic strengths are reported in figure 5. In all conditions, the network is faster for zero noise. For $\sigma_{FF} = 0$, information latencies are 22 ms for L3 and 32 ms for L4 in the ‘no RC’ case, and 36 ms in L3 and 54 in L4 for the Hebb RC condition. Thus purely FF processing is significantly delayed when noise is injected in the FF synaptic strengths. The interesting result is that the latency for the Hebbian RC case is much less delayed in the presence of noise. This indicates that the RC can contribute to information retrieval within a short time frame, and that in fact their presence may be particularly important in noisy conditions.

To directly model conduction time delays that might affect the overall time per layer of the network, we ran additional simulations with a conduction time delay of 3 ms both for the recurrent collateral axons and for the axons connecting to the next layer, and with an AMPA time constant for the recurrent collateral synapses of 7 ms (rescaling the synaptic conductances so that the total charge passed by a single synaptic event is the same as in the 10 ms case). Under the simple hypothesis that the retrieval time is approximately proportional to the inactivation time constant of the Hebbian excitatory collateral synapses (as is the case in the mean-field approximation [25]), added to the axonal transmission delays, the two situations should give very similar information latency results. We found that the information latencies obtained with transmission delays of 3 ms and a reduced synaptic inactivation time constant of 7 ms confirmed the conclusions reached with the simulations without delays and with longer synaptic inactivation time constant of 10 ms, although there were some differences.

¹⁰ The small difference between rates of ‘+’ neurons and the mean at time $t = 0$ is due to the finite number of pattern stored, and it does not carry any information (see [26], and figures 2 and 3).

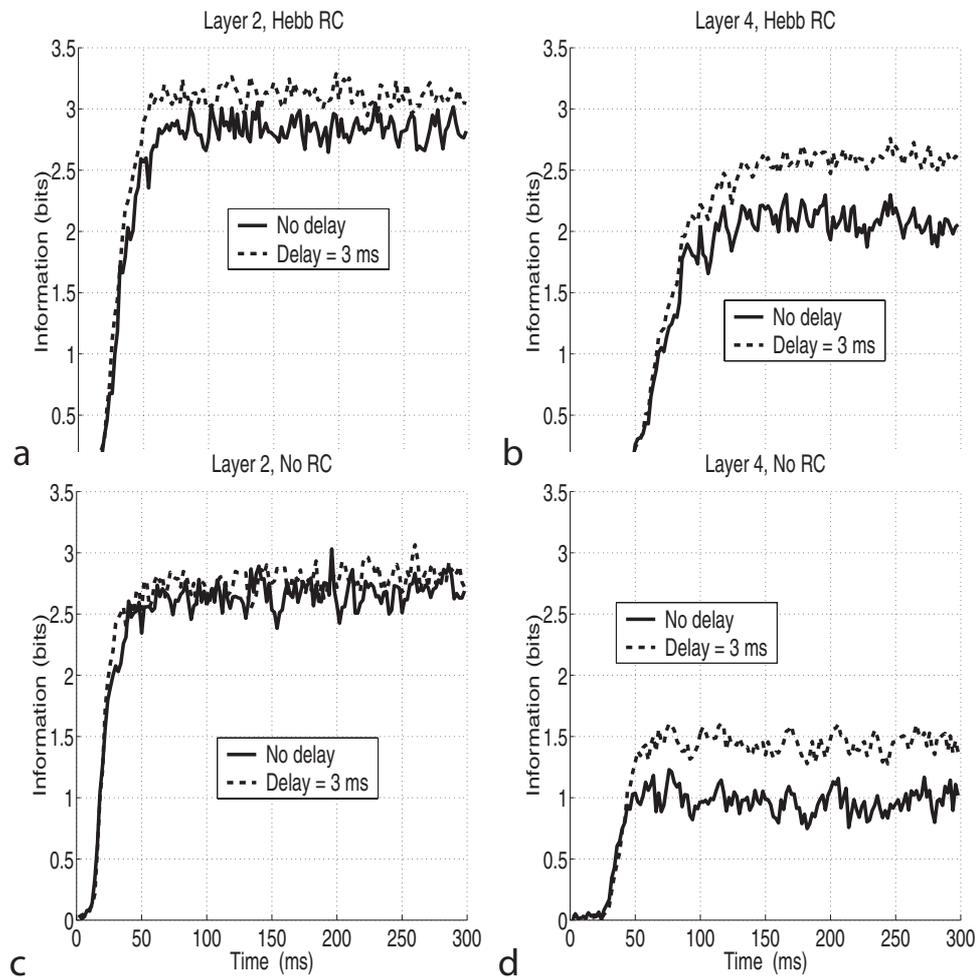


Figure 6. A comparison of the information time course obtained in the simulations with and without conduction delays. An inactivation time constant of 10 ms was used for the excitatory synapses in the no-delay case, and a delay of 3 ms and a synaptic inactivation time constant of 7 ms was instead used in the delay case. A value of $\sigma_{FF} = 16$ was used for the noise parameter. (a) layer 2, Hebb RC (b) layer 4, Hebb RC (c) layer 2, no RC (d) layer 4, no RC.

An example of the dynamics with delays, for a noise value intermediate to those presented in figures 2 and 3, is illustrated in figure 6. It is apparent that the onset of retrieval follows a similar pattern in both conditions, although the information plateau reaches a different value due to differences balance of excitation and inhibition reached in the two conditions (we rescaled only the excitatory synaptic conductance, leaving the inhibition parameters unchanged).

4. Discussion

We simulated the dynamics in continuous time of a multi-layer network of integrate-and-fire neurons when it is retrieving the information stored in both the feedforward and within-layer recurrent synaptic connections. The novelty of this paper, with respect to previous studies [25,26,35] considering a single cortical module, is that we have addressed more directly

the timescale of the specific contribution of recurrent feedback processing within layers along a multilayer pathway with forward connections between layers.

We found that, when the information content of the FF signal is very good, then pure FF processing is very fast when local excitation is absent. In this condition, the signal can be propagated across layers within 6–7 ms per layer. Given that we used inactivation time constants of 10 ms for the excitatory synapses, and an integration time constant of 20 ms for the membrane of excitatory neurons, the FF processing speed in continuous time is much faster than that predicted by a discrete timestep model. In this latter case at least one timestep would be required for each layer. We also found that the speed of pure FF information is slowed down when the FF signal has to compete with local non-specific excitation, particularly in conditions where the FF signal is information poor. This local non-specific excitation would be present in the cortex, where it would reflect either the spontaneous firing of the neurons, or the firing of the neurons to a preceding, different, stimulus.

We found that when the within-layer excitation involves feedback through Hebbian synapses, it can contribute to substantial information retrieval within 15 ms per layer, and information retrieval remains relatively fast even when the cue provided to the layer is very noisy. A comparison of the Hebbian RC case with the homogeneous RC case shows that the RC are able to make local excitation pattern-specific in a time of approximately 10 ms per layer, thereby facilitating the information transmission to the next area in a very short time.

Our results hence confirm the hypothesis that FF processing can be very fast, 6–7 ms per cortical stage, at least in noiseless conditions. However, if there is noise in the system (related for example to imperfect processing by earlier stages, or a noisy signal) then FF processing is slowed. Our study also shows that 15 ms per cortical area may be enough for performing *both* FF and recurrent computations. Hence, RC synapses can contribute significantly to stimulus processing, in a given area, within 10 ms of the initiation of the FF sweep of activity in that area. Recurrent processing is often suggested to be responsible for more sophisticated computations, such as visual attention, curve tracking and other phenomena that usually involve changes in a neuronal response happening on a timescale of several tens of ms per area (see e.g. [17,36]), and are hence not attributable to FF synapses only. However, the results presented here show that recurrent processing can be performed in an area within only 10 ms of the initial FF sweep of activity under noiseless conditions. Therefore, feedback processing is sufficiently fast that it can contribute to a variety of visual computations based on changes on neuronal activity on the timescale of 15 ms of processing time per layer. Hence, recurrent processing may play a role in phenomena such as rapid selectivity of face cell neurons under backward masking [10,11], rapid stimulus presentation [12], or rapid change from coarse to fine stimulus-related information in IT [6]. FF synapses in layered systems with continuous dynamics are part of the mechanism by which rapid cortical information processing is possible, and can certainly account for the rapid performance of some cognitive tasks [7,13,17]. What we have shown in this paper is that feedback processing within a layer can be sufficiently rapid to contribute to many other tasks, and indeed may actually be needed if the system is to work well under noisy conditions.

A simplification of the model is the strictly sequential architecture of the network used to model successive stages of the visual system. Although analysis of the anatomical connectivity suggests that the primate visual system is organized hierarchically [37,38], there are also direct connections from V1 to several non-primary areas at different levels of the visual hierarchy [37]. For this reason, a fraction of the visual processing in each stream might be done in parallel. Indeed, although response latencies in the ventral stream present a precise hierarchical organization, response latencies along the dorsal pathway are largely overlapping and consistent with more parallel processing [1,3]. We note that, if part of the visual processing

is done in parallel, there is more time for each area to process the signal within the 150 ms needed for fast visual cognition [13], and therefore the constraints that the speed of recurrent processing has to satisfy would be much less stringent. Therefore the timescales for recurrent processing met by our sequential model might be taken as a ‘worst-case’ scenario in which there is only serial, hierarchical processing. The possibility that part of visual function may be performed in parallel also suggests that further anatomical and neurophysiological studies may be necessary to define more precisely the processing time allowed per visual area.

We were able to address the speed of local (within area) feedback processing, but not the speed of top-down processing, involving back projections from higher areas. However, by the very same principles of continuous dynamics involved here, one would expect that top-down effects can arise in a much shorter time than would be expected from discrete timestep dynamics.

In this study we estimated the processing time per layer by taking the processing time required by the whole four-layer network and by dividing it by the number of layers. However, it is possible that the processing time required by higher layers is a little more than that required by lower layers. Indeed, this effect was evident in our simulations. For small noise values, the fourth layer took a few ms more than the third layer to provide 0.5 bits of information. For example, in the Hebb RC case, the layer 4 latency for low noise was 18 ms (to be compared with the average of 16 ms across layers). In the ‘no RC’ case, the layer 4 latency for low noise was 8 ms (to be compared with the average value of 6 ms across layers). Increasing the noise, we found this cumulative effect was much more marked for pure FF processing than for the Hebb RC case (see figures 3 and 5). From this last finding, from the small magnitude of the effects found, and from the fact that the number of processing stages of the primate ventral visual system is approximately six, we conclude that our estimate that recurrent circuits in each layer can contribute within 10 ms of the initial sweep of FF information is however robust.

It is of interest that part of the capability studied in some connectionist networks is obtained by virtue of the discrete timestep approach. One example is the connectionist type of recurrent network, which can produce a particular sequence of outputs by virtue of counting through its timesteps (see e.g. [39]). Given that the continuous time dynamics of networks of spiking neurons seem to be rather different from that of discrete-time models, with no discrete states in ‘timesteps’ on the way, the issue of how sequences of discrete states are produced becomes more complicated, and cannot be solved in this unrealistic way.

Finally, we note that attractor network processes operating by feedback may operate in many brain areas with the types of dynamical processing described in this paper. The feedback may be implemented by for example recurrent collateral connections, or by associatively modifiable forward and back-projections between cortical areas. The brain systems to which this may apply include the cerebral cortex, for example to implement short term memories; and the CA3 cells of the hippocampus to store and later retrieve information rapidly [21]. The generic results on the dynamics of feedback systems implemented with integrate-and-fire neurons that are described in this paper are expected to apply to all these systems.

Acknowledgments

We thank A Treves and M P Young for helpful discussions. This research was supported by the Wellcome Trust and the EC (SP), and by the Medical Research Council grant PG9826105 and the Human Frontier Science Program (ETR).

References

- [1] Bullier J and Nowak L G 1995 Parallel versus serial processing: new vistas on the distributed organization of the visual system *Curr. Opin. Neurobiol.* **5** 497–503
- [2] Nowak L G and Bullier J 1997 The timing of information transfer in the visual system *Cerebral Cortex: Extrastriate Cortex in Primate* ed K Rockland, J Kaas and A Peters (New York: Plenum) p 870
- [3] Schmolesky M T, Wang Y, Hanes D P, Thompson K G, Leutgeb S, Schall J D and Leventhal A G 1998 Signal timing across the macaque visual system *J. Neurophysiol.* **79** 3272–7
- [4] Celebrini S, Thorpe S J, Trotter Y and Imbert M 1993 Dynamics of orientation coding in area v1 of the awake primate *Vis. Neurosci.* **10** 811–25
- [5] Baylis G C, Rolls E T and Leonard C M 1987 Functional subdivisions of temporal lobe neocortex *J. Neurosci.* **7** 330–42
- [6] Sugase Y, Yamane S, Ueno S and Kawano K 1999 Global and fine information coded by single neurons in the temporal visual cortex *Nature* **400** 869–73
- [7] Oram M W and Perrett D I 1994 Modeling visual recognition from neurophysiological constraints *Neural Netw.* **7** 945–72
- [8] Tovée M J, Rolls E T, Treves A and Bellis R P 1993 Information encoding and the response of single neurons in the primate temporal visual cortex *J. Neurophysiol.* **70** 640–54
- [9] Heller J, Hertz J A, Kjaer T W and Richmond B J 1995 Information flow and temporal coding in primate pattern vision *J. Comput. Neurosci.* **2** 175–93
- [10] Rolls E T, Tovee M J and Panzeri S 1999 The neurophysiology of backward visual masking: information analysis *J. Cogn. Neurosci.* **11** 300–11
- [11] Kovacs G, Vogels R and Orban G A 1995 Cortical correlate of pattern backward masking *Proc. Natl Acad. Sci. USA* **92** 5587–91
- [12] Keysers C, Xiao D K, Foldiak P and Perrett D I 2001 The speed of sight *J. Cogn. Neurosci.* **13** 90–101
- [13] Thorpe S J, Fize D and Marlot C 1996 Speed of processing in the human visual system *Nature* **381** 520–2
- [14] Thorpe S J and Imbert M Biological constraints on connectionist models 1989 *Connectionism in Perspective* ed R Pfeifer, Z Schreier and F Fogelman-Soulie (Amsterdam: Elsevier) pp 63–92
- [15] Oram M W and Perrett D I 1992 Time course of neuronal responses discriminating different views of face and head *J. Neurophysiol.* **68** 70–84
- [16] Riesenhuber M and Poggio T 1999 Are cortical models really bound by the ‘binding problem’ *Neuron* **24** 87–93
- [17] Lamme V A F and Roelfsema P R 2000 The distinct modes of vision offered by feedforward and recurrent processing *Trends Neurosci.* **23** 571–9
- [18] Hopfield J J 1982 Neural networks and physical systems with emergent collective computational abilities *Proc. Natl Acad. Sci. USA* **79** 2554–8
- [19] Rumelhart D E and McClelland J L (ed) 1986 *Parallel Distributed Processing* (Cambridge, MA: MIT Press)
- [20] Amit D J 1989 *Modeling Brain Function* (Cambridge: Cambridge University Press)
- [21] Rolls E T and Treves A 1998 *Neural Networks and Brain Function* (Oxford: Oxford University Press)
- [22] Koch C and Segev I (ed) 1998 *Methods in Neuronal Modelling* 2nd edn (Cambridge, MA: MIT Press)
- [23] Tsodyks M V and Sejnowski T J 1995 Rapid state switching in balanced cortical models *Network* **6** 111–24
- [24] Gerstner W 2000 Population dynamics of spiking neurons: fast transients, asynchronous states and locking *Neural Comput.* **12** 43–89
- [25] Treves A 1993 Mean-field analysis of neuronal spike dynamics *Network* **4** 259–84
- [26] Battaglia F P and Treves A 1998 Rapid stable retrieval in high-capacity realistic associative memories *Neural Comput.* **10** 431–50
- [27] McCormick D A 1998 Membrane properties and neurotransmitter actions *The Synaptic Organization of the Brain* ed G M Shepherd (Oxford: Oxford University Press) pp 37–75 ch 2
- [28] VanVreeswijk C, Abbott L F and Bard Ermentrout G 1994 When inhibition not excitation synchronizes neuronal firing *J. Comput. Neurosci.* **1** 313–21
- [29] Parisi G 1986 A memory which forgets *J. Phys. A: Math. Gen.* **19** L617–9
- [30] Abbott L F 1991 Realistic synaptic inputs for model neural networks *Network* **2** 245–58
- [31] Shannon C E and Weaver W 1949 *The Mathematical Theory of Information* (Urbana, IL: University of Illinois Press)
- [32] Rolls E T, Treves A, Robertson R G, Georges-François P and Panzeri S 1998 Information about spatial view in an ensemble of primate hippocampal cells *J. Neurophysiol.* **79** 1797–813
- [33] Panzeri S, Treves A, Schultz S and Rolls E T 1999 On decoding the responses of a population of neurons from short time windows *Neural Comput.* **11** 1553–77
- [34] Panzeri S and Treves A 1996 Analytical estimates of limited sampling biases in different information measures *Network* **7** 87–107

-
- [35] Treves A 1993 Local neocortical processing: a time for recognition *Int. J. Neuronal Syst.* **3** 115–9
 - [36] Roelfsema P R, Lamme V A F and Spekreijse H 1998 Object based attention in the primary visual cortex of the macaque monkey *Nature* **395** 376–81
 - [37] Felleman D J and VanEssen D C 1991 Distributed hierarchical processing in the primate cerebral cortex *Cereb. Cortex* **1** 1–47
 - [38] Hilgetag C C, O'Neill M A and Young M P 1996 Indeterminate organization of the visual system *Science* **217** 776–7
 - [39] Jordan M 1986 An introduction to linear algebra in parallel distributed processing *Parallel Distributed Processing* ed D E Rumelhart and J L McClelland (Cambridge, MA: MIT Press) pp 365–422