ARTICLE Communicated by Mark Goldman

Optimizing One-Shot Learning with Binary Synapses

Sandro Romani
sandro.romani@gmail.com
Human Physiology, Università di Roma La Sapienza, Rome 00185, Italy

Daniel J. Amit
amit@marx.uchicago.edu
INFM, Università di Roma, Rome 00185, Italy, and Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel

Yali Amit
amit@marx.uchicago.edu
Departments of Statistics and Computer Science, Chicago, IL 60637 U.S.A.

A network of excitatory synapses trained with a conservative version of Hebbian learning is used as a model for recognizing the familiarity of thousands of once-seen stimuli from those never seen before. Such networks were initially proposed for modeling memory retrieval (selective delay activity). We show that the same framework allows the incorporation of both familiarity recognition and memory retrieval, and estimate the network's capacity. In the case of binary neurons, we extend the analysis of Amit and Fusi (1994) to obtain capacity limits based on computations of signal-to-noise ratio of the field difference between selective and non-selective neurons of learned signals. We show that with fast learning (potentiation probability approximately 1), the most recently learned patterns can be retrieved in working memory (selective delay activity). A much higher number of once-seen learned patterns elicit a realistic familiarity signal in the presence of an external field. With potentiation probability much less than 1 (slow learning), memory retrieval disappears, whereas familiarity recognition capacity is maintained at a similarly high level. This analysis is corroborated in simulations. For analog neurons, where such analysis is more difficult, we simplify the capacity analysis by studying the excess number of potentiated synapses above the steady-state distribution. In this framework, we derive the optimal constraint between potentiation and depression probabilities that maximizes the capacity.

1 Introduction

A novel model for a realistic account of the outstanding ability of the brain in distinguishing (recognizing) never-seen from familiar images,

even those seen only once, was presented in Amit, Yakovlev, Romani, and Hochstein (2008). The model was initially motivated by a surprising finding in the psychophysics of monkeys (Amit, Bernacchia, & Yakovlev, 2003; Yakovlev, Bernacchia, Orlov, Hochstein, & Amit, 2005; Amit et al., 2008), which are more effective in recognizing the repetition of an image in a random sequence (multiple delay-to-sample task) for images seen only once before than for regular images (seen many times before). This observation was associated with the psychophysics findings on humans (Standing, 1973), exhibiting an enormous familiarity memory capacity (approximately 10,000) for images seen but once.

The new model took off from a study of the memory capacity of networks endowed with stochastic learning of sparsely coded stimuli. It is a recurrent neural network, with built-in dynamical synapses. An attractive feature of this study is the use of plausible (stochastic) learning in binary synapses (Amit & Fusi, 1992, 1994). It turns out that for one-shot presentations of random patterns (representing visual stimuli), the learning generates sufficient synaptic changes to produce a realistic neural signal discriminating familiarity from novelty.

The theoretical analysis of capacity presented in Amit and Fusi (1994) was based on the signal-to-noise ratio of the afferent fields to selective versus nonselective neurons on presentation of a learned stimulus in a network of binary neurons. In other words, it monitored the requirement that a stimulus once presented, following learning, would reproduce itself after removal of the external input. Randomness is due to the stochastic learning process and the pattern generation process, but noise due to the network dynamics is disregarded. Furthermore, the analysis did not confront the signal produced in the network in the presence of the stimulus, seen once before, and the presence of a stimulus never before seen. In contrast, in the experiments, both monkeys and humans respond while the test image is present. In other words, the system should express a persistent signal in presence of the once seen as opposed to the novel stimulus. This motivated, in Amit et al. (2008), the monitoring of the signal difference generated by these two types of stimuli in simulated networks of analog neurons obeying the same stochastic learning as in Amit and Fusi (1994).

We first extend the signal-to-noise analysis of Amit and Fusi (1994), for binary neurons, to include the presence of an external selective current injected by the stimulus. This current is lower than the threshold used in the noiseless network dynamics. Familiarity is tested in a stationary state of the network, starting from an initial condition corresponding to the pattern presented, in addition to the external current, which is maintained throughout. Attractors (delay activity) are tested starting from the stationary state reached during the pattern presentation and monitoring the stationary state when the stimulus (external current) is removed. The estimated familiarity and attractor memory capacity are compared to the simulation results.

It turns out to be possible to avoid the ad hoc initial conditions for the dynamics in moving from binary to analog neurons. The stationary states are close to those of the network of discrete neurons, while much is lost in the stability and robustness properties of the network. This imposes the introduction of inhibition, which allows robust stationary states with intermediate rates, even for very simple implementation of the inhibition. But it complicates the analysis in terms of the signal-to-noise ratio, because both signal and noise in these cases depend on the rates in the stationary state, which in turn is to be determined by the dynamics of the network. The same difficulties, and more, are to be expected for networks of spiking neurons.

Motivated by the previous results, we keep the stochastic learning process proposed in Amit and Fusi (1994) and maintain the analysis in the asymptotic regime of a large number of learned random patterns. However, the properties of the network in the presence of learned patterns are analyzed by measuring the excess of potentiated synaptic connections associated with a pattern relative to the asymptotic synaptic potentiation. Specifically we estimate the memory storage capacity P_c, the maximal number of patterns that can be learned and still be all of recognizable as familiar by setting a lower limit (arbitrary, ab initio) on the fraction of useful synapses (Q) for the oldest learned stimulus (see also Fusi & Abbott, 2007).

The conclusions remain qualitatively unaltered: For low coding levels f, the average fraction of selective cells per pattern, and synaptic transition probabilities of $O(1)$, the storage capacity increases as $1/f$. A constraint on the ratio between the depression transition probability (q_-) and the potentiation transition probability (q_+), $q_- = \alpha f q_+$ ($\alpha \sim O(1)$), leads to a storage capacity of $O(1/f^2)$. When P_c is maximized with respect to α and q_+, one finds $\alpha = 1$ (i.e., $q_- = f q_+$) exactly. Moreover, this value is independent of the limit imposed on the number of useful synapses to be retained, provided the latter is lower than $1/(2e)$ of the total number of potentiable synapses. The value q_+ is constrained by the fact that there is a limited number of synapses available for potentiation (i.e., initially depressed) connecting active neurons. This limit is exhausted when $Q > 1/(2e)$, at which point $q_+ = 1$. Capacity decreases monotonically with increasing Q.

We show in simulations that familiarity memory does indeed follow the expected values for the capacity and exhibits the predicted dependence on the coding level f and the probability ratio α. Persistent activity in the presence of the stimulus (familiarity) can be achieved with lower values of Q, whereas persistent activity after stimulus removal, delay activity, requires higher values. In the setting of networks of binary neurons, Q can be determined in terms of the network parameters. In more complex networks, the value needs to be determined empirically.

Memory capacity for familiarity discrimination has been previously addressed in modeling studies (e.g., Bogacz, Brown, & Giraud-Carrier, 2001; Bogacz & Brown, 2003), but this study is different in several essential aspects. First, at the functional level, while previous studies addressed novelty

discrimination, here we focus on familiarity discrimination. There may be some difficulty differentiating these two mechanisms behaviorally, but the corresponding neural activity is quite distinct. The previous models predict lower rates for familiar stimuli than for novel ones; the model presented here behaves in the opposite way: familiar stimuli elicit higher average activity than novel ones. The previous results would correspond to physiological observations *in-vivo* in monkey perirhinal cortex (Xiang & Brown, 1998), while the new results would account for observations in prefrontal cortex (Xiang & Brown, 2004), which show a higher response for familiar stimuli.

In addition, the higher rates for familiar stimuli are also a lever for strengthening the encoding of a stimulus that has been observed several times, possibly indicating something of importance to be learned, as opposed to most random stimuli that are seen only once. Such strengthening can eventually lead to the consolidation of this stimulus in memory. Again this is in contrast to novelty detection, where the same stimuli are observed large numbers of times and the goal is to detect a new one.

There is another fundamental difference in the new modeling effort. Here we implement a conservative, realistic Hebbian plasticity (as opposed to the anti-Hebbian one in Bogacz et al., 2001, and Bogacz & Brown, 2003) and employ a simple, common network structure. In this, our model shares its elements with many studies on attractor networks, in both the use of binary synapses with stochastic learning rules (Tsodyks, 1990; Amit & Fusi, 1994; Brunel, Carusi, & Fusi, 1998; Amit & Mongillo, 2003; Del Giudice, Fusi, & Mattia, 2003; Senn & Fusi, 2005; Bernacchia & Amit, 2007; Fusi & Abbott, 2007) and analysis of the neural dynamics (Wilson & Cowan, 1972; Hopfield, 1984; Salinas & Abbott, 1996). We propose a unified framework that is able to explain both familiarity discrimination and identification (identified with attractor working memory, WM). This plasticity mechanism integrates naturally with the ideas mentioned above: that higher rates for familiar stimuli produce more potentiation and hence stronger encoding, and eventually selective delay activity. Finally, our model exhibits a strong dependence of memory capacity on the coding level, in contrast with the results of Bogacz and Brown (2002).

2 Recapitulation of One-Shot Learning in Binary Synapses

2.1 The Synaptic Distribution. The original studies of learning with binary neurons and binary synapses (Willshaw, Buneman, & Longuet-Higgins, 1969; Tsodyks, 1990) have been extended to a palimpsest memory, that is, to a network that can learn indefinitely, replacing old memories by new ones (Amit & Fusi, 1994). In this extension, stimuli to be learned are random patterns of N bits $(0, 1)$ of coding level f. We study two settings: *fixed coding size*, where each pattern has a subset of exactly fN 1-neurons, selected at random, and *random coding size*, where each neuron of each pattern is set

to 1 independently with probability f. These two settings exhibit different behavior in terms of capacity, as shown in Nadal (1991) and Brunel (1994).

Each synapse can assume one of two stable values: J_- depressed or J_+ potentiated. The presentation of a stimulus imposes the corresponding pattern of binary activities on the neurons of the network, and learning is expressed as a Markov chain whose state space is the set of the two stable synaptic efficacies. The stimuli are presented to the network one at a time, and the activity state of each pair of neurons determines the new state of the synapse connecting them.

The quantity useful for tracing the effect of later stimuli on the synaptic trace generated by previous ones, and hence for estimating the capacity of the network's memory, is $\rho_J^P(\xi, \tilde{\xi})$—the conditional probability of finding a synapse in the state J following the presentation of P stimuli, given that the first (oldest) stimulus in the sequence imposed the activity pair $(\xi, \tilde{\xi})$ to that synapse (Amit & Fusi, 1994). This quantity follows the stochastic dynamics imposed by the successive stimuli (see also Fusi, Annunziato, Badoni, Salamon, & Amit, 2000), which is,

$$\rho_J^P(\xi, \tilde{\xi}) = \sum_{K=J_-,J_+} \rho_K^1(\xi, \tilde{\xi})(M^{P-1})_{KJ}. \tag{2.1}$$

M_{KJ} is the transition matrix of the chain, defined by the following learning process.

Upon presentation of a given stimulus:

- If a synapse sees an activity pair $(1, 1)$ and is in its depressed state J_-, then $J_- \to J_+$, with probability q_+.
- If the activity is $(0, 1)$, that is, the presynaptic neuron is on and the postsynaptic neuron is off, and the synapse is in its potentiated state J_+, then $J_+ \to J_-$, with probability q_-.
- In all other cases the synapse is unmodified.

The resulting transition matrix M is

$$\begin{pmatrix} 1 - f(1-f)q_- & f(1-f)q_- \\ f^2 q_+ & 1 - f^2 q_+ \end{pmatrix}.$$

Its subleading eigenvalue (the maximal being 1) is

$$\lambda_M = 1 - f^2 q_+ - f(1-f)q_-.$$

This eigenvalue determines the rate of decay of memory via equation 2.1. If the sequence of stimuli becomes infinite, $\rho_J^P(\xi, \tilde{\xi})$ tends to

$$\rho^\infty = (\pi_+, \pi_-) = \left(\frac{f^2 q_+}{f^2 q_+ + f(1-f)q_-}, \frac{f(1-f)q_-}{f^2 q_+ + f(1-f)q_-} \right), \tag{2.2}$$

whose components are, respectively, the asymptotic fractions of potentiated and depressed efficacies. These are used as the natural initial distribution on top of which memories are learned (see, e.g., Amit & Fusi, 1994).

Equation 2.1 can then be rewritten as

$$\rho^P_{J_+}(\xi,\tilde{\xi}) = \lambda_M^{P-1}\left[\rho^1_{J_+}(\xi,\tilde{\xi}) - \pi_+\right] + \pi_+. \tag{2.3}$$

For the various activity pairs, the fraction of potentiated synapses on presentation of the earliest stimulus, $\rho^1_{J_+}(\xi,\tilde{\xi})$, reads,

$$\begin{aligned}
\rho^1_{J_+}(1,1) &= \pi_+ + \pi_- q_+ \\
\rho^1_{J_+}(0,1) &= \pi_+(1-q_-) \\
\rho^1_{J_+}(0,0) &= \pi_+ \\
\rho^1_{J_+}(1,0) &= \pi_+.
\end{aligned}$$

We reiterate that these probabilities are computed assuming the network starts with synaptic strengths distributed according to the stationary state. Note that here, we have opted for the asymmetric depression dynamics. In the symmetric case, we would have had $\rho^1_{J_+}(1,0) = \rho^1_{J_+}(0,1)$. There are no significant consequences for the following results.

2.2 Signal-to-Noise Analysis of Capacity. To estimate the network's storage capacity under this learning process, we follow the standard procedure (Willshaw et al., 1969; Nadal & Toulouse, 1990; Nadal, 1991) of comparing the distribution of the field $h_i^P = 1/N\sum_{j\neq i} J_{ij}^P \xi_j$ (current) to selective cells, which should be active (state 1), to that of the field to the nonselective ones, which should be inactive (0). If we set $J_- = 0$, then the averages of the two recurrent fields with respect to the randomness in both the synapses and the patterns are

$$\begin{aligned}
\langle h^P\rangle_1 &= E\left[h_i^P \middle| \xi_i = 1\right] = f J_+ \rho^P_{J_+}(1,1) \\
\langle h^P\rangle_0 &= E\left[h_i^P \middle| \xi_i = 0\right] = f J_+ \rho^P_{J_+}(0,1),
\end{aligned} \tag{2.4}$$

where $E[h_i^P \mid \xi_i = \xi]$ is the conditional expectation of the field to the cells with given activity ξ of the first learned stimulus. The signal is defined as

$$S_r^P = \langle h^P\rangle_1 - \langle h^P\rangle_0. \tag{2.5}$$

The asymptotic variance (large P) is the same for selective and nonselective neurons. Ignoring the correlations between synapses connected to the same neuron, we write the variance decomposition conditioning on the

pattern ξ,

$$\mathrm{Var}\left(\frac{1}{N}\sum_{j\neq i} J_{ij}\xi_j\right) = \frac{1}{N^2}\left[E\left[\mathrm{Var}\left(\sum_{j\neq i} J_{ij}\xi_j \mid \xi\right)\right] + \mathrm{Var}\left(E\left[\sum_{j\neq i} J_{ij}\xi_j \mid \xi\right]\right)\right]$$
$$= \frac{1}{N^2}\left[\mathrm{Var}(J)E\left[\sum_{j\neq i}\xi_j\right] + (EJ)^2\mathrm{Var}\left(\sum_{j\neq i}\xi_j\right)\right]. \quad (2.6)$$

For random coding-size stimuli to first order in f, this becomes,

$$R^2 = \frac{f}{N}E[J^2] = \frac{f}{N}\pi_+ J_+^2.$$

Note that for fixed coding size, the second term in equation 2.6 is zero, so that

$$R^2 = \frac{f}{N}\mathrm{Var}(J) = \frac{f}{N}\pi_+(1-\pi_+)J_+^2,$$

and the variance is smaller.

The existence of a gap between the two distributions (including the fluctuations) would allow a threshold that reconstructs the pattern presented. The ratio of the squared signal and the variance is found to be

$$\frac{\left(S_r^P\right)^2}{R^2} = \lambda_M^{2(P-1)} Nf \frac{(q_+\pi_- + q_-\pi_+)^2}{\pi_+}. \quad (2.7)$$

One then assumes a lower limit, $S_r^P/R > A$, to ensure the persistence of the oldest (first learned) pattern, arriving at a capacity (P_c) for the reproduction of the 0-1 patterns:

$$P_c \approx \frac{1}{-2\ln(\lambda_M)} \ln\left[\frac{fN(\pi_- q_+ + \pi_+ q_-)^2}{A^2\pi_+}\right]. \quad (2.8)$$

Since for low coding ($f \ll 1$),

$$-\ln\lambda_M \approx q_+ f^2 + q_- f, \quad (2.9)$$

at fixed q_+ and q_-, P_c is $\mathcal{O}(1/f)$. An additional "balancing" condition: $q_- = \alpha f q_+$ with $\alpha \sim O(1)$ (Amit & Fusi, 1994), yields

$$\pi_+ = \frac{1}{1+\alpha}, \quad \pi_- = \frac{\alpha}{1+\alpha}.$$

Keeping only terms of order f and assuming $J_+ = 1$, the eigenvalue λ_M can be approximated as $\lambda_M \approx 1 - q_+(\alpha+1)f^2$, and

$$\langle h^P \rangle_1 \approx f\left(\pi_+ + \pi_- q_+ \lambda_M^{P-1}\right); \quad h_0 \equiv \langle h^P \rangle_0 \approx f\pi_+ \tag{2.10}$$

$$R^2 \approx \frac{f\pi_+}{N}, \tag{2.11}$$

$$P_c \approx \frac{1}{2q_+(1+\alpha)f^2} \ln\left[\frac{Nfq_+^2\alpha^2}{A^2(1+\alpha)}\right], \tag{2.12}$$

so the capacity rises to $\mathcal{O}(1/f^2)$, while decreasing as $1/(1+\alpha)$.

2.3 Signal to Noise in the Presence of a Stimulus. In the presence of the stimulus, there is a constant external contribution S_e to the field of the selective neurons. The difference between the mean fields of selective and nonselective neurons becomes $S = S_e + S_r^P$, where S_r^P is the recurrent contribution of equation 2.5. If, for example, in terms of the standard deviation of the fields R, $S_e = B \cdot R$ for some $B < A$, one should have $S_r^P > (A - B) \cdot R$ to obtain persistent activity and equation 2.12 is modified to

$$P_c \approx \frac{1}{2q_+(1+\alpha)f^2} \ln\left[\frac{Nfq_+^2\alpha^2}{(A-B)^2(1+\alpha)}\right]. \tag{2.13}$$

One should note that the meaning of equation 2.13 is somewhat different from that of the corresponding equation 2.12. The latter ensures that a sufficient gap exists between fields afferent to selective and nonselective neurons to allow a threshold that renders the pattern an attractor. Equation 2.13, on the other hand, must be supplemented by the requirement that the threshold not be too low relative to the field on selective neurons of novel stimuli, which contains the contrast, S_e.

Basically, if the required gap between the signals is AR, with $A \gg 1$, then choosing the threshold at, say, $\theta \approx h_0 + (A - 1.5)R$, will put the field to the nonselective neurons, h_0, well under threshold. As long as $S_r^P > AR$, the field of a large proportion (over 90%) of the selective neurons will be above threshold, enough to sustain WM. Clearly, then, the smaller we can take A, the larger the capacity for WM.

For familiarity recognition, assume that the stimulus present has a contrast $S_e(=BR) \approx \theta - h_0$. In this case, the field to nonselective neurons, both learned and novel, will be far below θ, as before. For novel (unlearned) patterns, the signal is only $h_0 + B \cdot R$, which would be insufficient to retain the input pattern and network activity dies out. Selective neurons of learned stimuli will receive a field above threshold as long as the recurrent contribution $S_r^p > (A - B)R$, hence a much larger capacity than for WM. For familiarity alone, the smaller that $A - B$ is, the larger the capacity, but again the larger A is, the smaller the capacity for WM. Note that a key quantity is the argument of the logarithm in equations 2.12 and 2.13, which must be greater than 1; otherwise, the corresponding capacity is 0. These considerations are shown to hold in the simulations described in the next section.

2.4 Simulations for Binary Neurons. To illustrate the above considerations, we chose a network with 5000 neurons and 3000 patterns, of either random or fixed coding size. We explored two scenarios: one with fast learning, $q_+ = 1.0$, and the other with slower learning, $q_+ = 0.3$. In all cases $f = 0.02, \alpha = 1.0$, $S_e = 0.0075$, and $\theta = 0.017$. Substituting in equation 2.10, we find for both values of q_+ that

$$\pi_- \approx \pi_+ \approx 0.5,$$
$$h_0 = 0.01, \quad R = 0.0014,$$

and the eigenvalues are

$$\text{for } q_+ = 1 : \lambda = .99920; \quad \text{for } q_+ = 0.3 : \lambda = .99976.$$

To connect with the considerations preceding equation 2.9, we note that with these parameters,

$$S_e = 5R; \quad h_0 + S_e - \theta = 0.0005 = 0.33R; \quad \theta - h_0 = 0.007 = 4.7R.$$

Finally, we find $A = 6.0$, for the pattern that is at the limit of capacity. Note that the above value of R is for random coding size at the asymptotic limit. The variance for fixed coding size is smaller.

To study the dynamics of the network following learning, we implement asynchronous dynamics, defined by

$$V_i = H\left(\frac{\sum_j J_{ij}\xi_j}{N} + C_i - \theta\right), \tag{2.14}$$

where $V_i(t)$ are the dynamic (0,1)-variables of the units of the network; H is the Heaviside function, with threshold θ. When we test familiarity,

$C_i = S_e$ if the neuron is selective and $C_i = 0$ otherwise. For each pattern, the simulation is initiated with all selective neurons in their 1-state. When WM is tested, $C_i = 0$.

It turns out that at the level of the mean fields, the theory is quite accurate. Each stimulus is presented as a set of 1's on the selective neurons, and the field is recorded one iteration time later. The standard deviation R of the fields for large (asymptotic) P is observed to be 0.0015, somewhat larger than 0.0014, estimated in equation 2.11. This discrepancy is due to the fact that the variance is computed assuming the synapses are independent, whereas they are actually correlated. For nonnegligible f, this difference is significant (see Amit & Fusi 1994).

With these parameters, we can estimate the capacity for familiarity recognition, from equation 2.12, with $A - B = 1.0$ and the capacity for attractors is obtained taking $A = 6$, $B = 0$. Figure 1 shows the results of simulations for familiarity recognition and for delay activity attractors. Black curves present results for fixed coding size and gray ones for random coding size. Each point represents the fraction of selective neurons of the presented stimulus that are found in the 1-state once the network reaches a stationary state. This is averaged over a sliding window of 50 consecutive stimuli and over 5 simulation trials. Values lower than 1 primarily represent cases in which some stimuli, in the 50-stimuli window, led to an all-0 state. Other stationary states have nearly all (>95%) of the selective neurons in their 1-state.

For estimating the familiarity memory capacity from simulations, we smoothed the data even further by averaging over a window of 500 stimuli (compared to the 50 used to generate the figure). The memory capacity is estimated as the pattern at which the signal reaches 0.5, which, assuming no novel patterns are activated (see below), corresponds to an error rate of 25% (see also section 4.1). Figure 1A presents the familiarity signal for slow learning, $q_+ = 0.3$. For random coding size patterns, the observed familiarity capacity is 2670. The predicted familiarity capacity from equation 2.12 is 3133. In this case, no attractors are found, or expected, because the argument of the logarithm in equation 2.12, with $B = 0$, is negative. Figure 1B is a familiarity signal for fast learning $q_+ = 1$, and in Figure 1C, the corresponding delay activity attractors. The observed and predicted familiarity capacities for random coding size at $q_+ = 1$, are, respectively, 2220 and 2445. The memory capacities for attractors are 115 (observed) and 205 (predicted). For these networks of 0-1 neurons, approximately 97% of the novel stimuli lead to all-0 states. Hence, a familiar stimulus reaching a stationary state with some neurons in the 1-state has a clear familiarity signal.

We make the following comments:

- In all cases, fixed coding produces a higher average fraction of active selective neurons for both WM (where it exists) and familiarity. This implies that a higher fraction of these stimuli are recognized. For example, for $q_+ = 1$, for fixed coding, 100% of the 2000 most recent

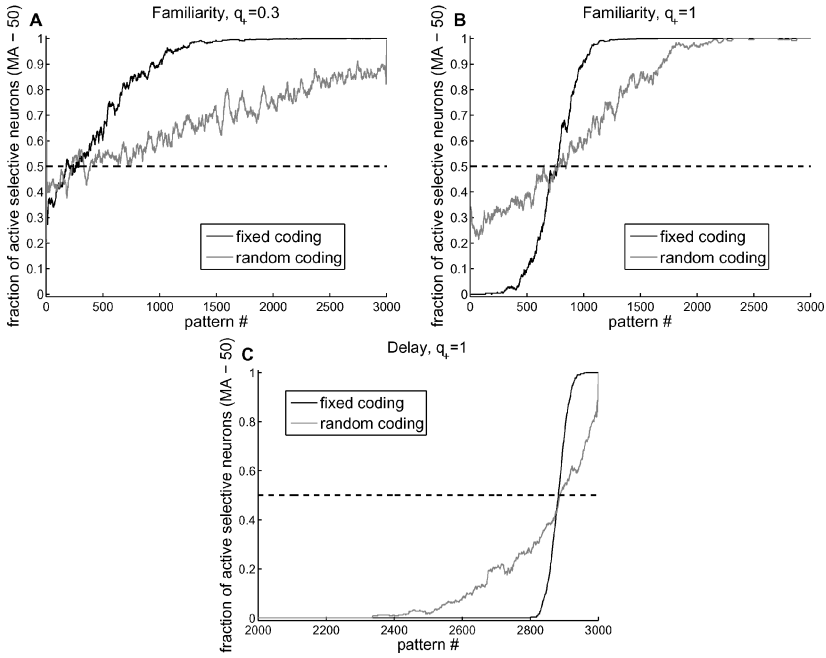

Figure 1: Familiarity recognition and attractor states in networks of binary neurons. Population-averaged activity (fraction of selective units in the 1-state) in a steady state corresponding to each stimulus versus stimulus index in the learning sequence (0, oldest stimulus; 3000, most recent). Gray: random coding size; black: fixed coding size. (A) Familiarity, slow learning, $q_+ = 0.3$. There is no working memory for any pattern. (B) Familiarity, fast learning, $q_+ = 1$. (C) Working memory attractors (no external stimulus, after initial state). Each point is the fraction of the selective neurons, found in the 1-state, in the stationary state reached for that stimulus, averaged over a sliding window of 50 consecutive stimuli (MA-50) and over 5 runs. Values lower than 1 are due primarily to cases in which some stimuli, in the 50-stimuli window, led to an all-0 state. Parameters: $N = 5000$, $P = 3000$, $f = 0.02$, $\alpha = 1.0$, $J_+ = 1.0$, $J_- = 0.0$, $S_e = 0.0075$, $\theta = 0.017$. Dynamics is asynchronous. Initial state is all selective units in 1-state, others in 0-state.

stimuli express familiarity recognition, while the number is more than halved for random coding.

- The situation is reversed for older stimuli, when random coding becomes more effective due to coding fluctuations.
- The total number of recognized stimuli is near equal for fixed and random coding size.

- For WM, at $q_+ = 1$, for fixed coding size, all most recent 100 stimuli have delay activity attractors, while the randomly coded stimuli start missing a significant fraction right from the most recent.
- For $q_+ = 0.3$ there is no working memory for random or fixed coding stimuli. At $q_+ = 1$, WM appears, which is in line with the idea that the type of learning proposed here, when repeated, would lead to an effective increase in q_+ and the appearance of delay activity, working memory (Amit et al., 2008).
- Note that the asymptotic no-error condition $S^2/R^2 \gg \ln N$, together with equation 2.7, implies the well-known constraint, $f > \ln N/N$ (Willshaw et al., 1969).
- Finally, an interesting fact is exposed by the simulations: the fixed and random performance curves intersect in the vicinity recognition rate of 0.5.

These subjects will be addressed in more detail in a future report.

3 Familiarity Recognition in Networks of Analog Neurons

In more realistic situations, neurons emit in a continuous range of rates, which can be modeled in various levels of detail (see, e.g., Amit & Brunel, 1997). The simplest is an extension of the response function of the 0-1 neurons discussed so far. This implies ascribing a width, w, which defines a transition of the neural rate, from 0 to 1, given the transition of its afferent current (field), from below to above threshold (Wilson & Cowan, 1972; Salinas & Abbott, 1996). Larger values of w correspond to higher levels of noise in the neural dynamics and lead to higher fluctuations in the rates (see below).

3.1 Network Dynamics. The network contains N excitatory neurons. Neuron i expresses, at time t, emission rate $\nu_i(t)$. The total excitatory recurrent current afferent to neuron i at any given time is

$$\mu_i^r = \sum_{j=1}^{N} J_{ij}\nu_j.$$

The total current to cell i is

$$\mu_i = \sum_{j=1}^{N} J_{ij}\nu_j + S_e\xi_i - A_I \sum_{j=1}^{N} \nu_j,$$

where $S_e\xi_i$ is a term describing the external current due to the presentation of the stimulus. The last term represents an inhibitory contribution proportional to the global network activity. The total afferent current μ_i is

Table 1: Parameters Used in the Simulations.

Parameters	Values
Network parameters	
f: Coding level	0.02
p: Number of coded stimuli	3000
N: Number of neurons	5000
Synaptic parameters	
J_-: Depressed efficacy	0
J_+: Potentiated efficacy	1
q_+: $J_- \to J_+$ transition probability	0.3–1
q_-: $J_+ \to J_-$ transition probability	$f \cdot q_+$
A_I: Inhibitory weight	0.5
Other parameters	
θ: Transduction function threshold	0.016
S_e: Contrast on selective cells	0.015
w: Transduction function width (0 for binary neurons)	0.004
$\Delta t/\tau$: Analog dynamics time step	0.5
ϵ: Relative tolerance	10^{-3}

transduced into a rate via

$$\nu_i = \Phi(\mu_i).$$

The evolution of the rates in the network is described by

$$\tau \frac{d\nu_i(t)}{dt} = -\nu_i(t) + \Phi(\mu_i(\nu_1, \ldots, \nu_N)). \tag{3.1}$$

The gain function used is

$$\Phi(\mu) = \frac{1}{2}\left(1 + \tanh\left(\frac{\mu - \theta}{w}\right)\right) \in [0, 1]$$

The dynamics is solved using the explicit Euler method, with time step Δt (see Table 1). This choice of dynamics (see equation 3.1) is purely for convenience.

To simulate the dynamics of the network, we first generate patterns as in the binary case, from which the synaptic matrix is constructed by the one-shot learning process described there. Given the matrix, we initialize the neurons with $\nu_i(0) = 0$, and then, for each pattern, the stimulus is imposed by injecting the current S_e. At this point we solve equation 3.1 numerically until a stationary state is reached. This set of stationary rates is used for the familiarity signal. The stationary state corresponding to a given pattern is then used as the initial condition for investigating WM. This is done by

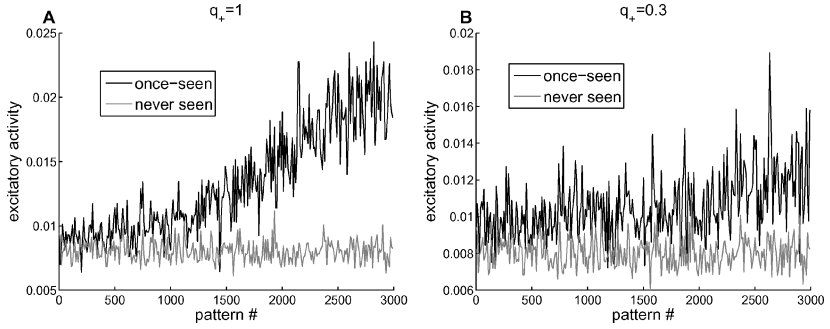

Figure 2: Familiarity signal in a network of analog units with inhibition. Rates in stationary states, averaged over all the excitatory units versus the stimulus presentation order. Black: familiar stimuli (1-shot learned); gray: novel stimuli. Every point is the average rate in the network for a single stimulus, sampled each 10 stimuli. Familiarity signal is the difference between the two curves. (A) Fast learning, potentiation probability, $q_+ = 1$. (B) Slow learning, $q_+ = 0.3$. Both networks are able to discriminate familiar from novel stimuli. The signal is stronger for recent images if $q_+ = 1$ but lasts further into the past if $q_+ = 0.3$. Random coding size; other parameters as in Table 1.

eliminating the stimulus ($S_e = 0$) and inspecting the new stationary state reached by the same dynamics, equation 3.1.

Stability is not affected by the chosen combination of tolerance ϵ (used to define convergence to the stationary state) and time step Δt. This has been verified in a subset of simulations by using lower values for both epsilon and Δt.

3.2 Simulation Results. We first observe that in the absence of inhibition, as w increases, the familiarity and attractor memory capacity decreases (data not shown). At the same time, the percentage of patterns leading to an all-1-state increases, and eventually all nontrivial stationary states disappear. Thus, the range of w producing meaningful network responses is quite narrow ($|\frac{\mu-\theta}{w}| \gg 1$). In this regime, the network behaves essentially as a network of binary units.

This problem is overcome by introducing inhibition in the network's dynamics (see, e.g., Amit et al., 2008). It tames the activity explosions and allows a wider parameter range: $|\frac{\mu-\theta}{w}| \sim 1$. Figure 2 shows that when the network's average rates are shown as a signals, the network can discriminate between once-seen (black) and never-seen (gray) patterns. The population-averaged emission rates in the network in stationary states for a given stimulus are sampled every 10 stimuli. Note that in contrast to Figure 1, for the 0-1 neurons, here the actual rates are plotted. The two panels confront the results for fast ($q_+ = 1$, Figure 2A) versus slow ($q_+ = 0.3$,

Figure 2B) learning. The difference between the two curves is the familiarity recognition signal. The large rate fluctuations are due to fluctuations in coding size and the stochastic learning.

The observed familiarity signal for recently learned stimuli is higher for fast learning than for slow learning. On the other hand, for slow learning, one observes (see Figure 2B) a larger capacity for discrimination of older stimuli, which is due to the slower decay of the signal (see section 4).

4 Capacity Optimization in One-Shot Learning

The signal-to-noise consideration is "static" in that it ignores the real dynamics of the network. Even for binary neurons, in the absence of inhibition, the network can easily move to the state where all neurons are active due to small fluctuations. Once inhibition is introduced, it is much harder to estimate the mean and variance of the fields at the steady state of the network. This becomes all the more difficult when moving to networks of integrate- and fire neurons, with analog rates.

We therefore simplify the considerations for memory capacity by estimating the expected fraction of synapses connecting selective neurons of a stimulus of age P that were potentiated on its presentation and remain so following the presentation of the subsequent $(P-1)$ successive stimuli. This is the fraction $\mathcal{E}^P(1,1)$ of potentiated synapses, in the sector of the synaptic matrix connecting pairs of selective neurons of stimulus P, that is among $(fN)^2$ synapses, over and above the background level, π_+ (see equation 2.2). From equation 2.3 we get

$$\mathcal{E}^P(1,1) = \rho^P(1,1) - \pi_+ = \lambda_M^{P-1}\pi_- q_+.$$

Imposing a lower limit Q on this fraction and substituting equation 2.2 for π_-, this yields the constraint

$$\mathcal{E}^p(1,1) = \frac{f(1-f)q_-q_+}{f^2q_+ + f(1-f)q_-}\lambda_M^{p-1} > Q. \tag{4.1}$$

Setting $q_- \equiv \alpha f q_+$, with $\alpha = O(1)$, to leading order in f, the above inequality can be written as

$$\lambda_M^{P-1} > Q\frac{(fq_+ + (1-f)q_-)}{(1-f)q_-q_+} \sim Q\frac{(1+\alpha)}{\alpha q_+},$$

and using equation 2.9, we obtain for the capacity estimate

$$P_c < \frac{1}{f^2q_+(1+\alpha)}\left[\ln\frac{q_+\alpha}{Q(1+\alpha)}\right]. \tag{4.2}$$

Maximizing P_c with respect to α and q_+ gives

$$\ln(Q) + \ln\left(\frac{1+\alpha}{\alpha}\right) - \ln(q_+) = -1/\alpha \tag{4.3}$$

$$\ln(Q) + \ln\left(\frac{1+\alpha}{\alpha}\right) - \ln(q_+) = -1. \tag{4.4}$$

Comparing equations 4.3 and 4.4 where both equations possess a solution (see below) gives

$$\alpha_c = 1,$$

which is in qualitative agreement with Amit and Fusi (1994) but provides a definite proportionality constant in section 2.1 and is not dependent on a balancing argument.

The maximizing q_+ is

$$q_{+,c} = 2eQ, \tag{4.5}$$

and the critical capacity, equation 4.2, evaluated at $(\alpha_c, q_{+,c})$, becomes

$$P_c = \frac{1}{4ef^2Q}.$$

The capacity scales, with the coding level, as $P_c \sim \frac{1}{f^2}$, as in Amit and Fusi (1994) and Willshaw et al. (1969). Since $q_+ \leq 1$, equation 4.5 implies a constraint: $Q < 1/(2e) \sim 0.18$. Recall that the maximum value of Q is $1 - \pi_+ = \pi_-$ when all available synapses selective to a given stimulus have been potentiated.

In fact, P_c, equation 4.2, has a single extremum in the entire two-dimensional space (α, q_+), which is a global maximum. Solutions are permissible only in the square $0 \leq (q_+, q_-) \leq 1$, and for $Q > 1/2e$ the maximum is outside this square. Since there is a unique extremum in the whole plane, the permissible maximum must lie on the square's boundary. Excluding the trivial cases $(q_- = 0, q_+ = 0)$, we have verified that for $q_- = 1$, P_c is a monotonic increasing function of q_+. Consequently P_c finds its maximum at the boundary: $q_{+,c} = 1$. The corresponding value of α is determined by the solution of

$$F(\alpha) \equiv \frac{\alpha}{1+\alpha} e^{-1/\alpha} = Q; \quad (1/2e < Q < \pi_-). \tag{4.6}$$

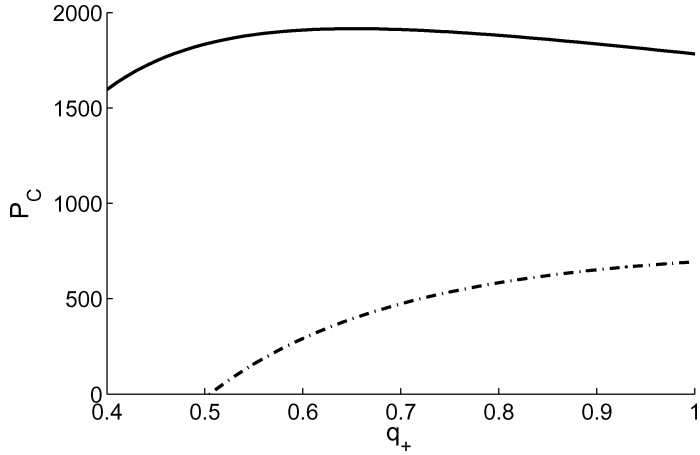

Figure 3: Memory capacity versus transition probability q_+. Solid curve: memory capacity for fixed $Q = 0.12 < 1/(2e)$ and $\alpha = 1$, $f = 0.02$. Dashed curve: capacity at $Q = 0.3 > 1/(2e)$, with $\alpha = \alpha_Q$ (solution of equation 4.6). When $Q > 1/(2e)$, the global maximum lies outside the permissible region, and the curve is monotonic.

The function $F(x)$ is a monotonic increasing function between 0 and 1, when $0 < \alpha < \infty$. For Q in the interval $[1/2e, \pi_-)$, $q_{-,c} = \alpha_Q f$, with α_Q given by the solution of equation 4.6. Given that $\alpha = O(1)$, it follows that $q_- = O(f) \ll 1$. Figure 3 presents the memory capacity as a function of q_+. It changes from concave (for $Q < 1/(2e)$, at $\alpha = 1$) to monotonic increasing ($Q > 1/(2e)$, α_Q solution of equation 4.6).

The capacity in the two regimes is given by

$$P_c = \begin{cases} \frac{1}{4eQf^2}, & \text{if } Q \le \frac{1}{2e} \\ \frac{1}{\alpha_Q(1+\alpha_Q)f^2}, & \text{if } Q > \frac{1}{2e} \end{cases}. \tag{4.7}$$

4.1 Simulations. To check the validity of these computations, we simulated networks of 5000 analog neurons, each storing one-shot learned $P = 10,000$ patterns, with the parameters of Table 1, varying only α and q_+. We ran 10 simulations for each pair of parameters chosen. Given a pair α, q_+, for each of the 10 resulting synaptic matrices, we stimulated the network with learned stimuli number $0, 50, 100, \ldots, 9950$ (for economy, see also Standing, 1973) and recorded the overall average rate at the stationary state of the dynamics (see equation 3.1). These are denoted $\nu_L^i(k), k = 1, \ldots, 200,\ i = 1, \ldots, 10$. Similarly we recorded the average rates in stationary states of the network stimulated by 200 random (unseen) stimuli, denoted $\nu_U^i(k)$. A stimulus is recognized as familiar if the stationary rate

for the seen pattern is larger than that of the corresponding unseen pattern: $\nu_L^i(k) \geq \nu_U^i(k)$. If the rate of the unseen pattern is higher, we have an error, as in a two-alternatives forced-choice test (Standing, 1973; see also Amit et al., 2008). For each simulation, we built a binary vector, $E^i(k) = 1(0)$, if there was (was not) an error, where higher k corresponds to more recent patterns. We estimate the probability of an error at a given pattern age by averaging the error vector $E^i(k)$ over the 10 runs: $E(k) = \sum_i E^i(k)/10$. The error probabilities $E(k)$ (smoothed using a centered moving average) are shown in Figure 4.

Familiarity recognition capacity is defined as the point where the smoothed error curves of Figure 4 reach 25%. This choice of a performance threshold is motivated by the fact that the error curve follows a sigmoidal shape, with values from 0 (no error, perfect performance) to 0.5 (50% of error responses, chance level). Note that frequently a 0-1 scale is used in the estimate of error probabilities, where 0 corresponds to perfect performance and 1 to chance level. With this scale S, which is simply $S = 2E$, the threshold would be 50%.

The capacities as a function of q_+ for different values of α are plotted in Figure 5. The maximum lies around $\alpha \approx 1$, $q_+ \approx .3$, and the corresponding observed maximal capacity is around 3800. The estimated $\alpha \approx 1$ is in agreement with the theoretical results described in this section. But the knowledge of the estimated optimal α and q_+ is not in itself sufficient to compare the estimated capacity with the one predicted by equation 4.2 because the value of Q is unknown. We therefore assume that the estimated optimal $q_+ \approx 0.3$ is produced from the correct value of Q (see equation 4.5), yielding $Q = q_+/(2e) = .055$, and a capacity of 4167. We have verified that when different methods are used to estimate performance (e.g., a discriminating threshold between seen and unseen patterns), the results are qualitatively unaltered.

5 Discussion

For binary neurons, fast learning (high plasticity probability per presentation) can produce a synaptic matrix expressing both very large familiarity capacity, and a moderate working memory capacity for the most recent patterns if the threshold and the stimulus contrast are properly chosen. This indicates that while learning is occurring at a slow rate, a "deviation" in which a pattern appears several times not too far away in the presentation sequence may lead to synaptic modifications allowing retrieval of this pattern in working memory for identification (see, e.g., Amit et al., 2008). In other words, the network serves to filter out most random stimuli that are seen only once and should not affect the long-term memory of the system. Those that are seen repeatedly may be important to retain in memory. One can imagine proposing a stochastic process of pattern repetitions, allowing a quantitative analysis of the network functioning in this

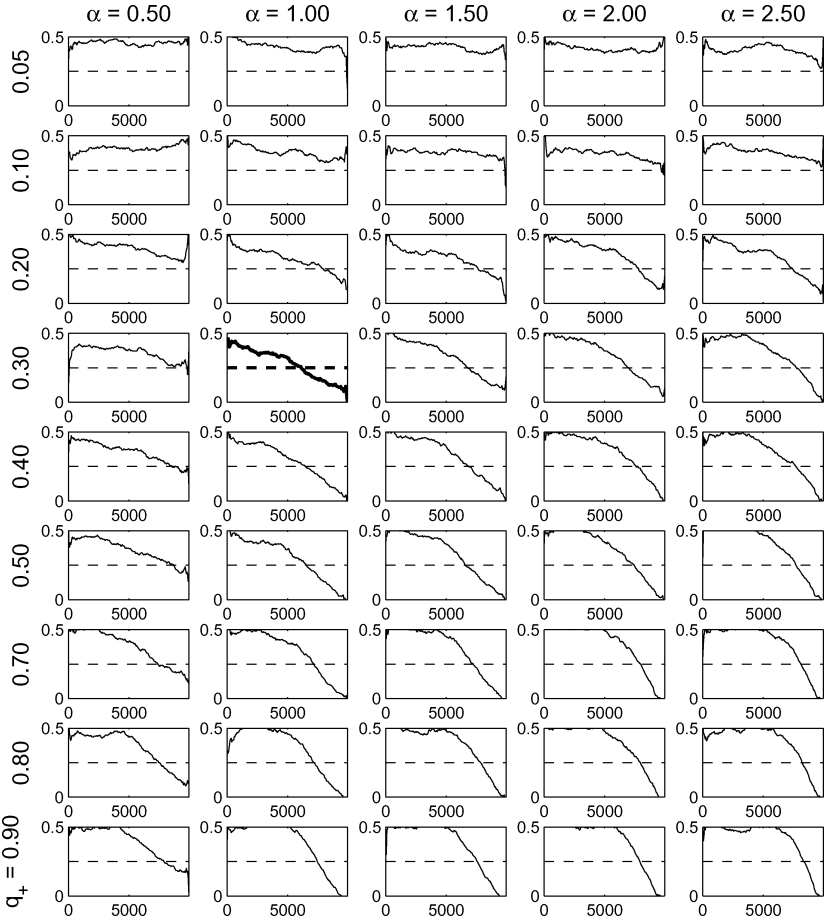

Figure 4: Familiarity recognition errors at pairs of α, q_+. An error is recorded when the stationary rate of a seen pattern is lower than that of a random unseen pattern (to which it is paired). We sample rates each 50th pattern. Curves: Fraction of errors among 10 trials. Curves are smoothed with a centered moving average over 50 points. Horizontal dashed line: performance threshold of 25% errors. Example panel with thick curves ($\alpha = 1, q_+ = 0.3$): Low fraction of errors for recent patterns; most of the once-seen stimuli produce network-averaged rates greater than those produced by novel stimuli. For older images, the rates are similar, so that on average, half of the time, there is an error. In this case, the error curve intersects the threshold near $P_c = 3800$. Total number of patterns $P = 10^4$. Other parameters as in Table 1.

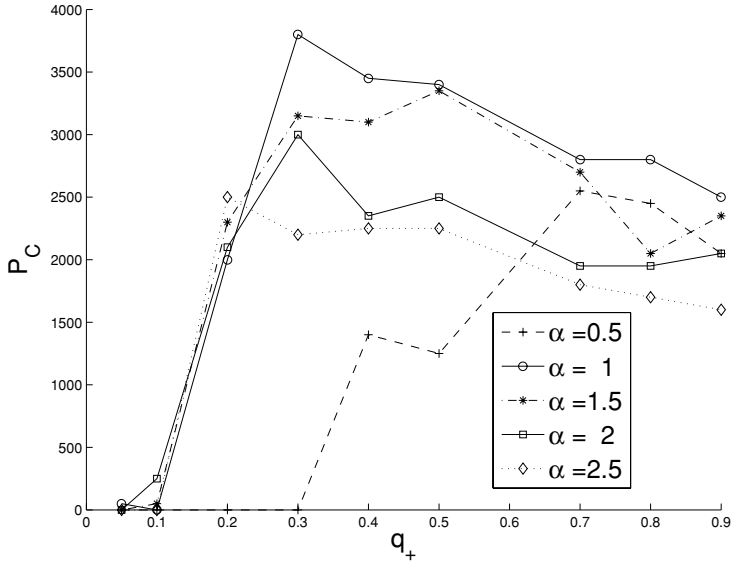

Figure 5: Estimated familiarity capacity from networks of analog neurons. The age of the pattern at which the error rate reaches 25%, estimated versus q_+ from Figure 4, at several values of α. A maximum for the memory capacity is found near $\alpha = 1$, $q_+ = 0.3$.

situation. An alternative interpretation of large values of q_+ could be a motivational enhancement of plasticity, which generates, perhaps even in one shot, identification memory (see, e.g., in the psychological context in McGaugh, 2000).

The capacity analysis for binary neurons is based on a fixed value of the variance of the input fields. This is a coarse analysis since this variance depends on the age of the patterns. A more detailed analysis could yield an explanation for the relationship between the error curves for fixed and random size coding (see Figure 1). This topic will be discussed in a future report.

The simplified capacity analysis produces good qualitative agreement with simulation data for networks of analog neurons. It may be worth pointing out that the relative robustness of the networks functioning, at fixed threshold and contrast, at varying constitutive parameter values (α, q_+), as opposed to the case for binary neurons, is due to two factors: the width of the response function and the presence of inhibition, which effectively renders the threshold rate dependent. However, there are several points that still require clarification. We have observed that for coding of fixed size, the optimal parameters occur at lower values of α and q_+, closer to the

boundary of the range of acceptable values (not shown). This fact may be related to the sensitivity of the network to the correlations implied in the fixed coding size case. Alternatively, it may be that for fixed-size coding, the analog network is more sensitive to the choice of threshold and contrast.

The robustness of the coexistence of familiarity memory with working memory, at $q_+ = 1$, is more fragile for analog neurons. Yet the phenomenology remains essentially the same, in the sense that we find sets of parameters (e.g., $\theta = 0.0045$, $S_e = 0.005$; other parameters as in Table 1) for which slow learning produces only familiarity signals, while for $q_+ = 1$, there are some attractors. In more realistic networks of spiking neurons, achieving such attractors would be even more problematic because of activity fluctuations.

As is well known, the tanh transduction function though attractive in its simplicity, and hence useful for exhibiting certain qualitative features (Salinas & Abbott, 1996), does not lead to networks of overall satisfactory operation (see, e.g., Amit & Tsodyks, 1991; Amit & Brunel, 1997). They have to be rather fine tuned and do not have stationary states at plausibly low rates. The main difficulty is in the sensitivity of working memory attractors to the threshold or, in turn, the average amplitude of the potentiation. At $q_+ = 1$, the decay of the synaptic trace is faster, and an attractor state for very recent stimuli will become destabilized rapidly with stimulus age.

This situation can be remedied with the transduction function proposed in Amit and Tsodyks (1991), motivated by Tuckwell (1988), which corresponds to a mean field description of spiking neurons. This allows low-rate stationary attractor states to become stable (Amit & Brunel, 1997; van Vreeswijk & Somplinsky, 1998). But when actual spiking neurons are reintroduced, attractor stability becomes fragile again, and NMDA receptors appear to be necessary (Wang, 1999; Mongillo, Curti, Romani, & Amit, 2005). In fact, familiarity capacity should be tested in more realistic networks of spiking neurons as in Mongillo et al. (2005).

Acknowledgments

D. J. A. and S. R. are indebted to S. Hochstein and V. Yakovlev for frequent discussions on the experimental aspects of familiarity recognition. D. A. was supported in part by the Changing Your Mind Center of Excellence of the Israel Academy of Science and by the SMC Center of Excellence, INFM, Rome. Y. A. was supported in part by NSF ITK DMS-0219016.

Yali Amit and Sandro Romani dedicate this article in memory of their coauthor, Daniel Amit, beloved father and teacher, who passed away on November 3, 2007.

References

Amit, D., Bernacchia, A., & Yakovlev, V. (2003). Multiple working memory model for behavioral experiment. *Cerebral Cortex*, *13*, 435–443.

Amit, D. J., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex*, *7*, 237–252.

Amit, D. J., & Fusi, S. (1992). Constraints on learning in dynamic synapses. *Network*, *3*, 443–464.

Amit, D. J., & Fusi, S. (1994). Learning in neural networks with material synapses. *Neural Comp.*, *6*, 957–982.

Amit, D. J., & Mongillo, G. (2003). Spike-driven synaptic dynamics generating working memory states. *Neural Comp.*, *15*, 565–596.

Amit, D. J., & Tsodyks, M. V. (1991). Quantitative study of attractor neural networks retrieving at low spike rates. II: Low-rate retrieval in symmetric networks. *Network*, *2*, 275–294.

Amit, D., Yakovlev, V., Romani, S., & Hochstein, S. (2008). Universal synaptic plasticity mechanism for familiarity and recognition memory. *J. Neurosci.*, *28*, 239–248.

Bernacchia, A., & Amit, D. J. (2007). Impact of spatiotemporally correlated images on the structure of memory. *Proc. Natl. Acad. Sci. USA*, *104*(9), 3544–3549.

Bogacz, R., & Brown, M. W. (2002). The restricted influence of sparseness of coding on the capacity of familiarity discrimination networks. *Network*, *13*, 457–485.

Bogacz, R., & Brown, M. W. (2003). Comparison of computational models of familiarity discrimination in the perirhinal cortex. *Hippocampus*, *13*, 494–524.

Bogacz, R., Brown, M. W., & Giraud-Carrier, C. (2001). Model of familiarity discrimination in the perirhinal cortex. *J. Comput. Neurosci.*, *10*, 5–23.

Brunel, N. (1994). Storage capacity of neural networks: Effect of the fluctuations of the number of active neurons per memory. *J. Phys. A*, *27*, 4783–4789.

Brunel, N., Carusi, F., & Fusi, S. (1998). Slow stochastic Hebbian learning of classes of stimuli in a recurrent neural network. *Network*, *9*, 123–152.

Del Giudice, P., Fusi, S., & Mattia, M. (2003). Modelling the formation of working memory with networks of integrate-and-fire neurons connected by plastic synapses. *J. Physiol. Paris*, *97*(4–6), 659–681.

Fusi, S., & Abbott, L. F. (2007). Limits on the memory storage capacity of bounded synapses. *Nat. Neurosci.*, *10*, 485–493.

Fusi, S., Annunziato, M., Badoni, D., Salamon, A., & Amit, D. J. (2000). Spike-driven synaptic plasticity: Theory, simulation, VLSI implementation. *Neural Comp.*, *12*, 2227–2258.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *PNAS*, *81*, 3088–3092.

McGaugh, J. L. (2000). Memory—a century of consolidation. *Science*, *287*, 248–251.

Mongillo, G., Curti, E., Romani, S., & Amit, D. J. (2005). Learning in realistic networks of spiking neurons and spike-driven plastic synapses. *Eur. J. Neurosci.*, *21*, 3143–3160.

Nadal, J. (1991). Associative memory: On the (puzzling) sparse coding limit. *J. Phys. A Math. Gen.*, *24*, 1093–1101.

Nadal, J.-P., & Toulouse, G. (1990). Information storage in sparsely coded memory nets. *Network*, *1*, 61–74.

Salinas, E., & Abbott, L. F. (1996). A model of multiplicative neural responses in parietal cortex. *Proc. Natl. Acad. Sci. USA*, *93*, 11956–11961.

Senn, W., & Fusi, S. (2005). Convergence of stochastic learning in perceptrons with binary synapses. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, *71*(6 Pt. 1), 061907.

Standing, L. (1973). Learning 10,000 pictures. *Q. J. Exp. Psychol.*, *25*, 207–222.

Tsodyks, M. (1990). Associative memory in neural networks with binary synapses. *Modern Physics Letters B*, *4*, 713.

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

van Vreeswijk, C., & Somplinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Computation*, *10*, 1321–1371.

Wang, X.-J. (1999). Synaptic basis of cortical persistent activity: The importance of NMDA receptors to working memory. *J. Neurosci.*, *19*, 9587–9603.

Willshaw, D., Buneman, O. P., & Longuet-Higgins, H. (1969). Nonholographic associative memory. *Nature (London)*, *222*, 960–962.

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*, 1–24.

Xiang, J. Z., & Brown, M. W. (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology*, *37*, 657–676.

Xiang, J. Z., & Brown, M. W. (2004). Neuronal responses related to long-term recognition memory processes in prefrontal cortex. *Neuron*, *42*, 817–829.

Yakovlev, V., Bernacchia, A., Orlov, T., Hochstein, S., & Amit, D. (2005). Multi-item working memory—a behavioral study. *Cereb. Cortex*, *15*, 602–615.

Received October 3, 2007; accepted December 27, 2007.