nature neuroscience

Limits on the memory storage capacity of bounded synapses

Stefano Fusi & L F Abbott

Memories maintained in patterns of synaptic connectivity are rapidly overwritten and destroyed by ongoing plasticity related to the storage of new memories. Short memory lifetimes arise from the bounds that must be imposed on synaptic efficacy in any realistic model. We explored whether memory performance can be improved by allowing synapses to traverse a large number of states before reaching their bounds, or by changing the way these bounds are imposed. In the case of hard bounds, memory lifetimes grow proportional to the square of the number of synaptic states, but only if potentiation and depression are precisely balanced. Improved performance can be obtained without fine tuning by imposing soft bounds, but this improvement is only linear with respect to the number of synaptic states. We explored several other possibilities and conclude that improving memory performance requires a more radical modification of the standard model of memory storage.

The idea that memories are stored through long-lasting modifications of synaptic strengths within neural circuits has become a basic postulate of neuroscience. Experimentalists have made great strides in providing evidence for this viewpoint[1,2], and theorists have added further support[3,4]. In particular, theorists have shown that models using long-term synaptic plasticity as a storage mechanism can retain enormous numbers of memories for long periods of time. The models used to establish this result involve many simplifying assumptions and not-so-realistic features, but it has generally been believed that their impressive memory performance is not an artifact of these simplifications and would generalize to more realistic models. Unfortunately, this is not true. The remarkable memory capacity of the classic neural network models depends critically on one of their least realistic features, the fact that their synapses are unbounded. Modifying this single feature by placing reasonable bounds on synaptic efficacies reduces memory capacity in these models so drastically that it shakes the entire theoretical underpinnings supporting the link between memory (and learning) and long-lasting modifications of synaptic strength[5–7].

The dominant factor that causes memory traces to decay over time in models[5–7], experiments involving *in vitro*[8–12] and *in vivo*[12–19] synaptic modification, and psychophysical studies[20–22] is overwriting of the synaptic modifications representing the memory by ongoing plasticity. This can occur due to synaptic modifications arising from spontaneous activity, or synapses retaining a trace of one memory can be overwritten when other memories are stored that share some of the same synapses as the first memory. We use the term 'ongoing plasticity' to refer to both of these effects. In the models we consider, the rate of ongoing plasticity events, which we denote by r, is a critical parameter that sets the scale for memory lifetimes. It is difficult to know what value r takes, but it can be estimated. For example, spike pairings suitable for inducing synaptic potentiation or depression through spike timing–dependent plasticity (see, for example, ref. 23) caused by pre- and postsynaptic neurons firing at 1 Hz occur about once a minute. We keep r as an unknown factor that sets the scale for the memory lifetimes we compute, but we estimate that $1/r$ lies somewhere between 10 and 100 s.

To study how quickly memories are forgotten, we focus on one particular memory and track its mnemonic trace in the presence of ongoing plasticity. We isolate the synapses that are modified during storage of the specific tracked memory, and we determine the rate at which these synapses return to an equilibrium distribution and the memory trace is lost. To simplify this calculation, we assume that ongoing plasticity that is due to spontaneous activity and to the storage of new memories is uncorrelated with the plasticity that stored the particular memory we are tracking.

Previous work has shown that ongoing plasticity causes an exponential degradation of memory traces if synaptic efficacies are restricted to lie within realistic ranges[5–7,24]. We refer to the time constant governing this exponential decay as the memory lifetime and denote it by τ. Memories can only be recovered from a population of n synapses for a time of order $\tau \ln(n)$ after they have been stored[5–7] (see also Methods). Because the lifetime τ is proportional to $1/r$, which we estimate to be around 1 min, and the logarithm is such a slowly increasing function, memory capacity is extremely small even if n includes every synapse in the brain. Two methods come to mind for evading this catastrophe. One is to require synapses to traverse many different states before they reach their bounds. It has been suggested previously that this might improve memory performance significantly[6]. Here we show that this improvement is not robust. The other, which we also explore, is to modify how the bounds that limit the range of synaptic efficacy are implemented. It should be stressed

Center for Neurobiology and Behavior, Kolb Research Annex, Columbia University College of Physicians and Surgeons, 1051 Riverside Drive, New York, New York 10032-2695, USA. Correspondence should be addressed to L.F.A. (lfa2103@columbia.edu).

Received 27 October 2006; accepted 29 January 2007; published online 11 March 2007; doi:10.1038/nn1859

that neither of these changes affects the logarithmic dependence of memory capacity on the number of synapses used to store the memories. Changing this requires a more dramatic shift in the way synapses store memories[25]. Instead, we explore here whether the coefficient τ multiplying the logarithmic dependence can be increased enough to compensate for the severe limits on memory capacity in models without such metaplasticity.

RESULTS

Studying memory storage, maintenance and retrieval by neural circuits raises numerous questions and problems for both theoretical and experimental investigation. We attempted to reduce the number of issues and uncertainties by taking an 'ideal observer' approach[25]. Rather than constructing a hypothetical neural network that stores and recovers memories through synaptic modification, we directly track the states of the synapses that such a network would use. In other words, we model how synapses are modified during both memory storage and ongoing activity and ask whether a memory trace can be detected within this population of synapses by examining the synapses directly, rather than by examining the output of network neurons. Of course, we are not proposing that neural circuits directly measure synaptic strengths as we do. Rather, our approach extracts an upper bound on memory performance that optimal circuits may approach but that no circuit can exceed. The advantage of this approach is that it allows us to determine the strength and lifetime of memory traces without having to speculate about how neural circuits actually recover and reveal those traces.

We characterize the strength of each synapse by a weight parameter, w, that falls between 0 and 1. Thus, we measure synaptic strength in units of the maximum possible synaptic efficacy. When synaptic plasticity occurs, the synapse is modified according to

$$w \rightarrow w + q_+(w) \text{ or } w \rightarrow w - q_-(w) \qquad (1)$$

for potentiation and depression, respectively. The expressions $q_+(w)$ and $q_-(w)$ represent weight-dependent factors that determine the size of the potentiation and depression modifications.

In addition to the rate of ongoing plasticity, r, we needed to know the relative amounts of potentiation and depression making up the ongoing plasticity. We denote the probability of plasticity being potentiating as f_+ and the corresponding probability for depression as f_-, with $f_+ + f_- = 1$. The rates of events that cause potentiation and depression are thus rf_+ and rf_-, respectively.

With these parameters defined and set to particular values, we can proceed to examine the memory storage properties of populations of n synapses. We subject these synapses to continuous modifications at the rates given above until an equilibrium configuration is achieved. This equilibrium is characterized by constant distributions of synaptic efficacy, although modification of individual synapses is still proceeding. We denote the average value of synaptic efficacy at equilibrium by $\bar{w}$. We then introduce a particular memory that we will track over time to study features of memory storage. It should be stressed that this particular memory is no different from any of the other memories whose storage is represented by the ongoing synaptic modifications, nor is it the only memory being stored by the system. It is just a particular memory that we happen to choose to monitor.

The tracked memory is initially stored by imposing a pattern of synaptic potentiations and depressions on the population of synapses. At any later time, we define the memory trace or 'signal' as the average distance separating the potentiated and depressed synapses from their equilibrium values. In other words, the memory signal for n synapses is

$$S = \frac{1}{n}\left(\sum_{i=\text{pot}} (w_i - \bar{w}) - \sum_{i=\text{dep}} (w_i - \bar{w})\right) \qquad (2)$$

where the sums are over all the synapses potentiated (first sum) and depressed (second sum) by the memory event being tracked. Note that the average value of the synaptic strengths does not have to be altered by memory storage. The tracked memory is detectable only because we have divided the population of synapses into two groups on the basis of whether they were potentiated or depressed during the storage of this particular memory.

The signal given by equation (2) must be detected despite the fact that synapses are continually changing their values as a result of ongoing plasticity, which causes the memory signal to fluctuate and degrade. The fluctuations introduce a 'noise' N, defined as the s.d. of the signal (see Methods). The detectability of the memory trace is determined by the signal-to-noise ratio, S/N. Specifically, we use two quantities to characterize the quality of memory storage[25]: the initial signal-to-noise ratio right after the tracked memory is stored, labeled S_0/N_0, and the memory lifetime τ, which is the time constant describing the exponential decay of the memory signal-to-noise ratio over time[6,7]. S_0/N_0 gives the strength of the initial memory trace and τ controls how long it is detectable, as discussed in the introduction. Our primary interest is in the memory lifetime, so we concentrate on that quantity, but we report S_0/N_0 values as well to verify that the memory trace for which we are reporting a given lifetime is detectable.

To compute the memory signal given by equation (2), we need to split the population of synapses into two groups: those potentiated and those depressed by the tracked memory. This is what we do for all the mean-field simulations (see Methods). However, when we discuss and compute memory lifetimes, we keep track of only one of these two groups, the group potentiated by the tracked memory. We do this because the analysis for the depressed group of synapses is identical to that for the potentiated group, except for obvious reversals of the roles of potentiation and depression. In particular, the lifetimes we report for the potentiated group of synapses apply equally to the depressed set if the parameters for depression and potentiation are interchanged.

A basic issue we explore is whether memory lifetimes can be increased by reducing the amount by which each potentiation and depression modifies synaptic efficacies. The idea is that smaller changes are less destructive to a stored memory[6,7,26,27]. We use a parameter α to characterize the magnitude of each potentiation or depression, and many of our results are expressed in terms of this parameter. There are a number of ways in which α can be interpreted. For the description we use here, the entire allowed range of synaptic efficacy, from the weakest allowed synapse to the strongest, defines 1 unit of synaptic strength. In these units, α sets the scale of the efficacy change induced by a single plasticity event. We consider the case when the range of synaptic efficacies is kept fixed and the plasticity step size, controlled by α, is modified. Alternatively, the plasticity step size could be held fixed and the allowed range of efficacies could be expanded proportional to $1/\alpha$. The results we report for memory lifetimes apply to either case. Our results for the initial signal-to-noise ratio, however, should be multiplied by $1/\alpha$ if this latter interpretation is used. In either case, the number of unitary plasticity events of a given type (potentiation or depression) required to move the synapse from one limit of its efficacy range to the other is proportional to $1/\alpha$. We typically treat synaptic efficacy as a continuous variable taking any value in the allowed range, but our results apply equally to cases in which synaptic strength only takes a number m of discrete values corresponding to m synaptic states.

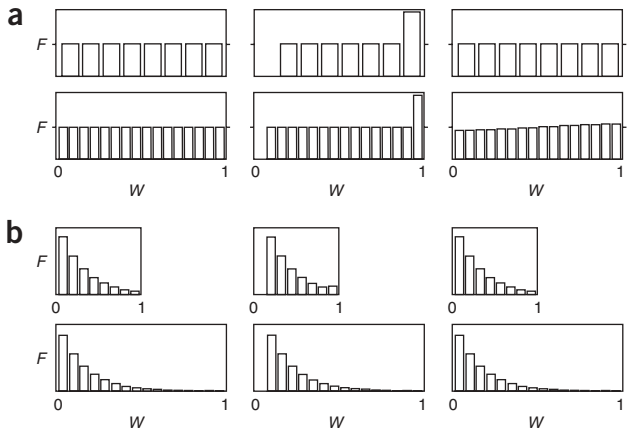

Figure 1 Distributions F of strength for synapses potentiated by the tracked memory and constrained by hard bounds. In both **a** and **b**, the upper row of panels corresponds to $\alpha = 1/8$ and the lower row to $\alpha = 1/16$. The bin sizes are set equal to the value of α, so potentiation shifts synapses one bin to the right in both cases. The left panels show the synaptic strength distribution at equilibrium, before the tracked memory is stored. The distribution is uniform and the tick marks outside the frames indicate the height of the bins at equilibrium. The middle panels show the distribution immediately after the tracked memory has been stored, and the right panels show the distribution at a time $t = 50/r$ after memory storage. The scale for F is such that the total area under the histogram is 1 in all cases. (**a**) Balanced potentiation and depression ($f_+ - f_- = 0$). Memory lifetime is longer for smaller values of α (lower row of panels). (**b**) Unbalanced potentiation and depression ($f_+ - f_- = -0.2$). Memory lifetime is no longer improved by making α smaller. In the upper row of panels, the x axis has been scaled by a factor of 2.

In this case, $1/\alpha$ is equivalent to m. In other words, reducing α can be interpreted as increasing the number of synaptic states.

Having described the framework we use, we now explore how different plasticity modification functions, $q_+(w)$ and $q_-(w)$, affect memory performance. Recall that the poor memory performance in the class of models we are studying arises from the presence of bounds on the synaptic strengths that store the memories. We will focus on two aspects of plasticity: the magnitudes of $q_+(w)$ and $q_-(w)$ and their dependence on the synaptic strength w. Studying the first allows us to see whether memory performance can be improved by forcing the synapse to step through many states before it reaches a boundary. The second aspect relates to the question of whether modifying the way in which bounds on synaptic strength are imposed can improve memory performance.

Hard bounds

We begin by considering the so-called hard boundary case in which the step sizes for potentiation and depression are constant, independent of the synaptic weight,

$$q_+(w) = q_-(w) = \alpha \qquad (3)$$

within the range $0 < w < 1$, and the boundaries at $w = 0$ and $w = 1$ are imposed by truncating any transitions that would take a synaptic weight outside the allowed range between 0 and 1. (We could allow for different values of α for potentiation and depression, but any such differences can be absorbed into the factors f_+ and f_- that we have introduced, so we do not consider this further.) Memory storage under these conditions is illustrated in **Figure 1**, which shows histograms for the population of synapses potentiated by the tracked memory.

We distinguish two cases in **Figure 1**: the balanced case when potentiation and depression are equally likely ($f_+ = f_-$, **Fig. 1a**), and the unbalanced case in which they are not ($f_+ \neq f_-$, **Fig. 1b**). In the balanced case, the equilibrium configuration for synaptic strengths before storage of the tracked memory is uniform (**Fig. 1a**, left), indicating that synaptic strengths are equally likely to take any of the allowed values. In the distribution immediately after this set of synapses has been potentiated during the storage of the tracked memory (**Fig. 1a**, middle), potentiation of the synapses has emptied the left-most bin and moved the excess into the right-most bin. At a time $t = 50/r$ after the tracked memory was stored, the distribution of synaptic weights has returned to its equilibrium configuration, and no trace of the stored memory remains. Recalling that we estimate $1/r$ to be of order 1 min, this illustrates the short memory lifetime in models with bounded synapses.

The lower row of panels in **Figure 1a** illustrates the improvement in memory lifetime caused by reducing the potentiation step size by a factor of 2. The left, middle, and right panels show the distribution of synaptic strengths at the same times as the upper row of panels. In this case, the distribution of synaptic strengths has not yet returned to equilibrium by the time $t = 50/r$ shown in the rightmost panel, indicating that the memory trace had not yet vanished. Thus, halving the step size for potentiation improved the memory lifetime[6]. This improvement provided the motivation for our study, because we wondered how general and robust it was.

A simple argument can explain the improved memory lifetime (**Fig. 1a**). When potentiation and depression are balanced, the random walk in the space of synaptic weights due to ongoing plasticity is unbiased, so the return to equilibrium is a diffusion process. In other words, when the synapses move up and down by the same average amount, their evolution is described by a random walk without drift. This description is not valid when the random walk hits the boundaries, but we use it as an approximation to estimate the motion between the boundaries. Each new plasticity event represents a step in this random walk, and in time t there are rt such steps. The distance traveled by a random walk of step size α in rt steps is of the order of $\alpha\sqrt{rt}$. Memory storage moves a block of synapses all the way from the leftmost bin of the strength histogram to the rightmost bin (**Fig. 1a**, middle panels). Thus, the effects of the storage of the tracked memory will be erased roughly when the random walk has had time to move these synapses all the way across the histogram, a distance of 1. This occurs when $\alpha\sqrt{rt} \approx 1$, from which we find the memory lifetime to be $\tau \approx 1/(r\alpha^2)$. Thus, in the balanced case, the memory lifetime grows as $1/\alpha^2$, the inverse of the square of the plasticity step size.

When potentiation and depression are not balanced, the equilibrium distribution is not flat, but rather exponential (**Fig. 1b**, left). In the case shown, $f_+ < f_-$, so the equilibrium distribution is skewed toward the lower bound of synaptic strength. When the memory is stored (**Fig. 1b**, middle), the distributions for both values of α (upper and lower panels) shift one bin to the right. In this unbalanced case, both distributions return to equilibrium and the memory trace has disappeared by the time $t = 50/r$, corresponding to the rightmost panels. Thus, the advantage of a smaller potentiation step size that we found when potentiation and depression were balanced disappears in the unbalanced case.

When potentiation and depression are unbalanced, the random walk that erases the memory trace is biased away from the middle of the synaptic range, producing an overall drift of the system at a speed αr (the contribution of the much slower diffusion is negligible in this

case). In addition, as can be seen from the middle column of panels in **Figure 1b**, memory storage in the unbalanced case moves the histogram only one bin to the right, a distance α. The time required to undo this shift is then determined by $\alpha rt = \alpha$, which gives a memory lifetime of $\tau = 1/r$, which is independent of α. Thus, in the unbalanced case, there is no advantage in terms of the memory lifetime when the step size for plasticity is decreased.

Memory performance, quantified by the memory lifetime τ and the initial signal-to-noise ratio S_0/N_0, in both the balanced and unbalanced cases is illustrated in **Figure 2**. The memory lifetime is defined as the time constant of the slowest exponential component in the convergence to the equilibrium distribution. We determined τ both by fitting an exponential function to the signal-to-noise ratio for large times and by numerically computing the subdominant eigenvalue of the matrix of synaptic transitions (see Methods). The two results were almost identical, so we plot only the one obtained from the subdominant eigenvalue. As α decreases, τ continues increasing quadratically over the whole interval shown in **Figure 2a** only when potentiation and depression are perfectly balanced. If the balance between the effects of potentiation and depression is not maintained, the improvement due to reducing the potentiation step size becomes negligible (**Fig. 2a**,**b**). The initial signal-to-noise ratio is proportional to α in the balanced case and tends to a nonzero constant in the unbalanced case (**Fig. 2c**). The latter is due to the fact that both the signal and the noise go to zero at the same rate as α decreases.

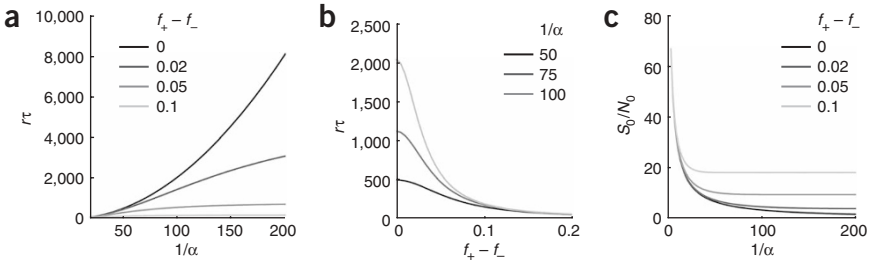

Figure 2 Memory performance with hard bounds. All lines are mean-field simulations for $n = 10{,}000$ synapses. (**a**) Memory lifetime (in units of $1/r$) versus $1/\alpha$ for different levels of imbalance, as indicated in the key. (**b**) Memory lifetime (in units of $1/r$) versus the degree of imbalance for different values of $1/\alpha$, as indicated in the key. (**c**) Initial signal-to-noise ratio versus $1/\alpha$ for different levels of imbalance, as indicated in the key.

We have described two effects of plasticity step size on memory lifetime depending on whether the plasticity is balanced or not, but what determines the degree of balance that separates these two behaviors? We can answer this question by computing the memory lifetime analytically. This is done by describing the ongoing plasticity that degrades the tracked memory as a Markov process. The detailed calculation is given in the **Supplementary Methods** online. Briefly, the decay of the memory trace is controlled by the subdominant eigenvalue λ_M of the Markov matrix describing ongoing synaptic modification, specifically $\tau = 1/(r(1 - \lambda_M))$. When the potentiation step size is small, the resulting lifetime is

$$\tau = \frac{1}{r\left(\sqrt{f_+} - \sqrt{f_-}\right)^2 + \alpha^2\pi^2 r\sqrt{f_+ f_-}} \qquad (4)$$

This illustrates the two behaviors we have been describing. When f_+ is close enough to f_- so that the second term in the denominator dominates over the first term, the memory lifetime is approximately

$$\tau \approx \frac{1}{\alpha^2\pi^2 r\sqrt{f_+ f_-}} \qquad (5)$$

which grows with decreasing α like $1/\alpha^2$. However, if the first term dominates, we have

$$\tau \approx \frac{1}{r\left(\sqrt{f_+} - \sqrt{f_-}\right)^2} \qquad (6)$$

which is independent of α. For any nonzero degree of imbalance ($f_+ - f_- \neq 0$), no matter how small, there is a point beyond which reducing α does not enhance memory lifetime quadratically. The requirement for approximately quadratic improvement is

$$\alpha > \frac{\left|\sqrt{f_+} - \sqrt{f_-}\right|}{\pi(f_+ f_-)^{\frac{1}{4}}} \qquad (7)$$

In summary, decreasing the plasticity step size leads to improved memory lifetimes only when synaptic potentiation and depression are balanced against each other, and for any degree of imbalance there is a minimum step size below which no further improvement in memory performance occurs.

Notice that in our analysis we decided to track a generic memory, which is preceded by a large number of events that have already generated memory traces. Hence the initial distribution of the synapses was chosen to be the equilibrium distribution. If we prepare the synapses in a special state before we store the memory that we intend to track, its memory lifetime can be improved by decreasing the plasticity step size even when potentiation and depression are not balanced. For example, if the synapses are all set to an intermediate strength, the decay of the memory trace becomes exponential and governed by the τ of equation (4) only after a time proportional to $1/\alpha$. Before this time, the synapses are dragged toward one of the two extremes of the synaptic range, but they do not feel the boundaries, so they behave like unbounded synapses. This improvement is illusory because the condition that all synapses start from a specific value can be imposed for only one memory, so the event that generates it becomes special. Such a dependence on the initial distribution has been exploited to build models of primacy and recency[28].

Soft bounds

The hard bounds considered in the previous section are a rather harsh way of limiting synaptic strength, and this raises the question of whether the requirement of balance for memory performance improvement might be due to this particular form of bounding. In this section we therefore consider a 'softer' way of limiting synaptic strength[29–31]. Soft bounds are introduced by allowing $q_+(w)$ and $q_-(w)$ to depend on synaptic strength and by requiring that they vanish at the boundaries. One way of doing this is to write

$$q_+(w) = \alpha(1 - w) \text{ and } q_-(w) = \alpha w \qquad (8)$$

In this case, the equilibrium distribution of synaptic strengths is approximately Gaussian, centered around a point $\bar{w}$ (**Fig. 3a**). Owing to ongoing plasticity, synapses increase in strength at an average rate of $q_+(\bar{w})f_+$, and they decrease in strength at an average rate of

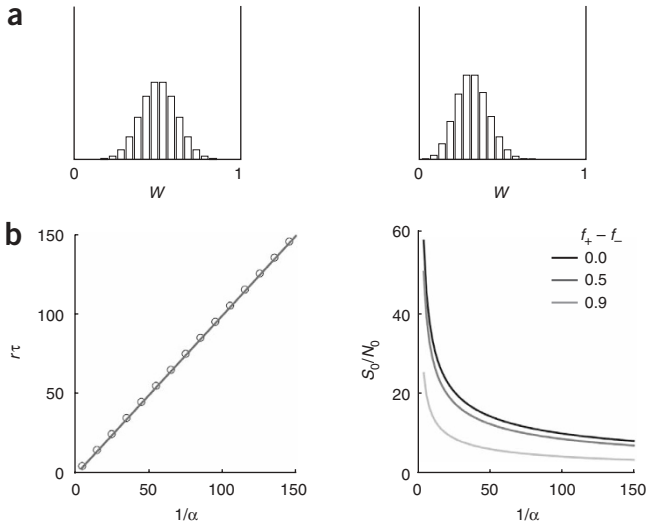

Figure 3 Soft bounds. (**a**) Histograms of equilibrium distributions of synaptic efficacies with soft bounds for balanced (left) and unbalanced (right) conditions, from mean-field simulations. (**b**) Memory performance. The level of imbalance does not affect the result in this case. Left, memory lifetime (in units of $1/r$) versus $1/\alpha$. Right, initial signal-to-noise ratio versus $1/\alpha$. The lines are mean-field simulations for $n = 10{,}000$ synapses and the points are the theoretical prediction based on equation (9).

$q_-(\bar{w})f_-$. The equilibrium point is where these two opposite synaptic drifts cancel each other. That is, $\bar{w}$ is determined by $q_+(\bar{w})f_+ = q_-(\bar{w})f_-$, which gives $\alpha f_+(1-\bar{w}) = \alpha f_- \bar{w}$. From this we find that $\bar{w} = f_+/(f_+ + f_-) = f_+$. The last equality follows from the fact that $f_+ + f_- = 1$.

The two panels in **Figure 3a** show distributions of synaptic strengths for the balanced ($f_+ = f_-$, left) and unbalanced ($f_+ > f_-$, right) cases. Note that these distributions are virtually identical except for a shift. This suggests that there may be no significant difference between the balanced and unbalanced cases when soft bounds are applied—which, as we will see, is the case.

A simple calculation can be done to estimate the memory lifetime to a remarkable degree of accuracy in the case of soft bounds. Recall that to calculate the memory lifetime we need to find the slowest exponential component in the convergence to the equilibrium distribution. The slowest component is determined by the location where the drift is minimal, which is in the neighborhood of the equilibrium point $\bar{w}$. Therefore, we focus on the motion of synapses in that region. We can think of the storage of the tracked memory as shifting the potentiated synapses starting at $\bar{w}$ by an amount $q_+(\bar{w})$. This moves them from the equilibrium point $\bar{w}$ to a value $\bar{w} + q_+(\bar{w})$. Shifting away from the equilibrium point induces a drift velocity that pushes synapses back toward equilibrium (in the presence of a drift, the effects of diffusion are small). The drift velocity is the difference between the rightward velocity at the shifted point, $rf_+q_+(\bar{w} + q_+(\bar{w}))$, caused by potentiation and the leftward velocity, $rf_-q_-(\bar{w} + q_+(\bar{w}))$, caused by depression. The resulting drift speed is then $v_{\text{drift}} = r|f_+q_+(\bar{w} + q_+(\bar{w})) - f_-q_-(\bar{w} + q_+(\bar{w}))|$. The drift of the distribution will undo the effects of memory storage at a time t when $v_{\text{drift}}t = q_+(\bar{w})$. This gives a memory lifetime of

$$\tau = \frac{q_+(\bar{w})}{r|f_+q_+(\bar{w} + q_+(\bar{w})) - f_-q_-(\bar{w} + q_+(\bar{w}))|} \tag{9}$$

Using the equations given above, we find, after a little algebra, that $\tau = 1/(\alpha r)$, which grows linearly with $1/\alpha$ and is independent of the state of balance or imbalance. This linear improvement is due to the fact that the drift velocity is proportional to α^2 and the drift distance is proportional to α. The value of τ given by equation (9) is compared with mean-field simulation results in the left panel of **Figure 3b**, and the linear improvement with $1/\alpha$ is readily apparent. In addition, the right panel of **Figure 3b** shows the initial signal-to-noise ratio, which is proportional to $\sqrt{\alpha}$ because the signal is proportional to α and the noise, given by the width of the equilibrium distribution, is proportional to $\sqrt{\alpha}$.

In conclusion, memory lifetime is proportional to $1/\alpha$ for synapses constrained by these soft bounds. This produces a smaller improvement in memory lifetime as α is decreased than the $1/\alpha^2$ dependence of the balanced case with hard bounds, but the improvement in this case is not destroyed by unbalancing. Thus, the effects of small α with soft bounds, though more modest, are more robust than those observed with hard bounds.

Generalized soft bounds

A more general form for $q_+(w)$ and $q_-(w)$ that still imposes soft bounds is[31]

$$q_+(w) = \alpha(1-w)^\gamma \text{ and } q_-(w) = \alpha w^\gamma \tag{10}$$

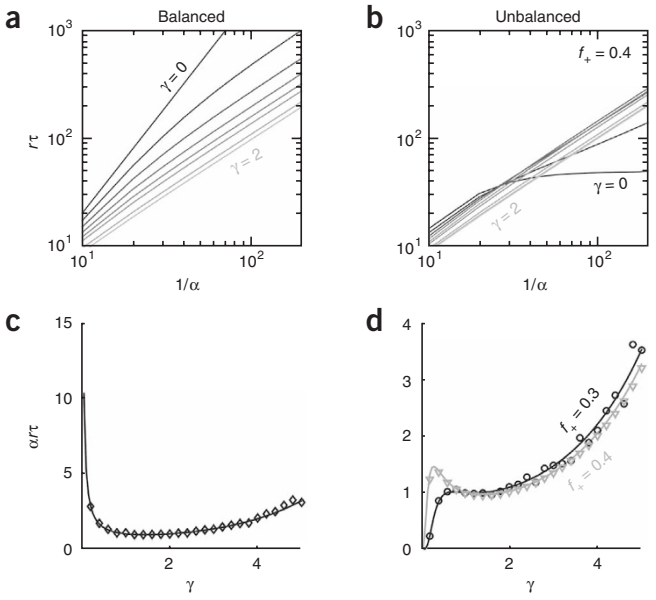

Figure 4 Memory performance with generalized soft boundaries. (**a**,**b**) Memory lifetime versus the inverse step size α when potentiation and depression are balanced ($f_+ = f_- = 0.5$; **a**) and unbalanced ($f_+ = 0.4$ and $f_- = 0.6$; **b**). The memory lifetimes are computed using a mean-field approach. Different curves correspond to different values of the power γ. For $\gamma > 0$, the memory lifetime always scales linearly (slope of 1 on the log-log plot) with $1/\alpha$, for small α, in both the balanced and the unbalanced cases. (**c**,**d**) The slope of $r\tau$ versus $1/\alpha$ estimated according to equation (13) as a function of γ for the balanced case (**c**), and for two unbalanced cases (**d**): $f_+ = 0.4$ and $f_- = 0.6$, as in (**b**), and $f_+ = 0.3$ and $f_- = 0.7$. The theoretical estimates (solid lines) are compared to the results of the mean-field simulations (diamonds, which correspond to the balanced case (**c**), circles to $f_+ = 0.4$ (**d**) and triangles to $f_+ = 0.3$ (**d**) (with $f_- = 1 - f_+$). For all plots the number of synapses n is 10^4.

In this case, the average synaptic strength at equilibrium is

$$\bar{w} = \frac{1}{1 + (f_-/f_+)^{\frac{1}{\gamma}}} \tag{11}$$

We can compute the lifetime in this case from equation (9). The calculation simplifies if we assume that α is small, so that we can write

$$q_\pm(\bar{w} + q_+(\bar{w})) \approx q_\pm(\bar{w}) + q'_\pm(\bar{w})q_+(\bar{w}) \tag{12}$$

where the prime symbol denotes differentiation. After some algebra, we find that

$$\tau = \frac{1}{\alpha\gamma r(f_+(1-\bar{w})^{\gamma-1} + f_-\bar{w}^{\gamma-1})} \tag{13}$$

with $\bar{w}$ given by equation (11). As in the case of $\gamma = 1$, considered in the previous section, this lifetime scales as $1/\alpha$ whether potentiation and depression are balanced or not, as long as $\gamma > 0$. The memory lifetime is plotted against $1/\alpha$ in **Figure 4a**,**b** for the balanced and the unbalanced cases, respectively. Different curves correspond to different values of γ. For $\gamma = 0$, we recover the results for the hard bounds: in the balanced case, the memory lifetime grows quadratically with $1/\alpha$, whereas when potentiation and depression are unbalanced, the memory lifetime saturates. For $\gamma > 0$, the memory lifetime τ always scales linearly with the inverse step size $1/\alpha$. In all these cases, τ is proportional to $1/\alpha$ and the multiplying factor can be estimated by using the approximate expression of equation (13). The theoretical estimate is plotted for the balanced and unbalanced cases in **Figure 4c**,**d**, respectively, and compared to the value obtained by running mean-field simulations (diamonds, triangles and circles). The agreement between the theoretical estimate and the simulations is excellent and the simple estimate of equation (13) also captures the nonmonotonicity of the curve corresponding to $f_+ = 0.3$. Models with large γ seem to perform better in terms of memory lifetime. However, the price to be paid is a reduced initial signal-to-noise ratio that actually decreases rapidly with γ (as a power law).

Optimally adjusted bounds

The soft bound cases considered in the previous two sections have memory lifetimes proportional to $1/\alpha$ owing to their nonzero drift velocities away from the equilibrium point. These models outperform those with hard bounds for small α, except when potentiation and depression are balanced, in which case hard bounds give a lifetime proportional to $1/\alpha^2$ as opposed to the $1/\alpha$ dependence that occurs with soft bounds. We now ask whether it is possible to construct a model with soft bounds that retains the $1/\alpha$ dependence in the unbalanced case, but matches the $1/\alpha^2$ performance of hard bounds when potentiation and depression are balanced.

Such a model does indeed exist. Consider

$$q_+(w) = \frac{\alpha}{2}(1 - (2w-1)^\gamma) \quad \text{and} \quad q_-(w) = \frac{\alpha}{2}(1 + (2w-1)^\gamma) \tag{14}$$

for γ equal to an odd positive integer. The memory lifetime for this model can be derived from equations (9) and (12), yielding

$$\tau = \frac{1}{\gamma\alpha r(2\bar{w}-1)^{\gamma-1}}$$

This shows that the model matches the $1/\alpha$ dependence of other soft-bound models. However, consider what happens in the balanced situation when $f_+ = f_-$. In this case, $\bar{w} = 1/2$ (as can be derived by setting $q_+(\bar{w}) = q_-(\bar{w})$), and the denominator in the above lifetime goes to zero. Clearly, something has gone wrong with the calculation in this case. Looking back at the expansion of equation (12), we see what the problem is. The first derivatives $q'_\pm(\bar{w}) = q'_\pm(1/2)$ are zero. As a result, we must extend the expansion to second derivatives, but at this point the diffusion term becomes as important as the drift term (see the discussion below). Thus, we expect a quadratic dependence on $1/\alpha$, as in the case of hard boundaries. This is confirmed by the mean-field results shown in **Figure 5**.

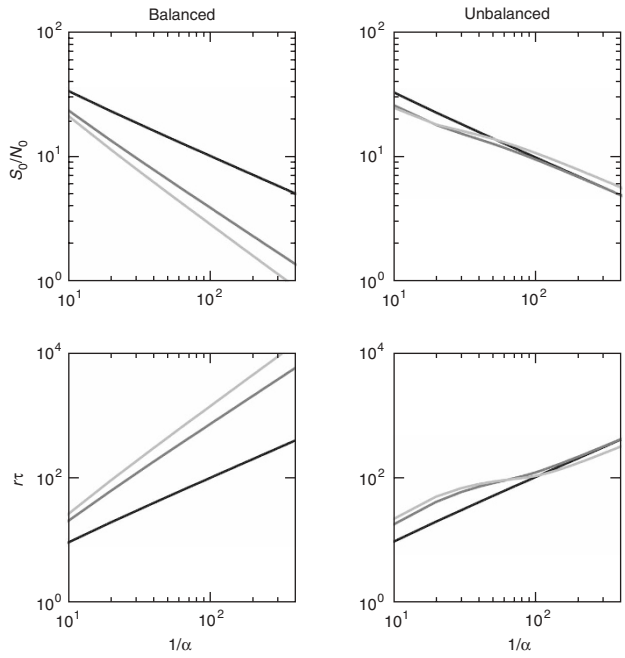

Figure 5 Optimal soft-boundaries. Mean-field analysis of initial signal-to-noise ratio (top) and memory lifetime (bottom) versus $1/\alpha$ with $n = 10^4$ for the balanced (left) and the unbalanced case (right) with $f_+ = 0.4$ and $f_- = 0.6$. Different curves correspond to different values of γ ($\gamma = 1$, 3, 5, from dark to light). The black curve corresponds to the soft bound case ($\gamma = 1$), and it does not depend on the imbalance between potentiation and depression. The other two curves show a memory lifetime which has an almost quadratic dependence on $1/\alpha$ (slope of 2 on the log-log plot) in the balanced case, and a linear dependence (slope of 1 on the log-log plot) in the unbalanced case.

Table 1 Dependences of memory lifetime and initial signal-to-noise ratio on α

Model	$f_+ = f_-$	τ	S_0/N_0
Hard bounds (equation 3)	Balanced	α^{-2}	α
Hard bounds (equation 3)	Unbalanced	α^{0}	α^{0}
Soft bounds (equation 8)	Balanced	α^{-1}	$\sqrt{\alpha}$
Soft bounds (equation 8)	Unbalanced	α^{-1}	$\sqrt{\alpha}$
Generalized soft bounds (equation 10)	Balanced	α^{-1}	$\sqrt{\alpha}$
Generalized soft bounds (equation 10)	Unbalanced	α^{-1}	$\sqrt{\alpha}$
Special bounds (equation 14)	Balanced	$\alpha^{1/\gamma-2}$	$\alpha^{1-1/(2\gamma)}$
Special bounds (equation 14)	Unbalanced	α^{-1}	$\sqrt{\alpha}$

Shown are the dependences of memory lifetime and initial signal-to-noise ratio on the plasticity step size α for the models considered in the text in the balanced ($f_+ = f_-$) and unbalanced ($f_+ \neq f_-$) cases. The equations refer to the definitions of $q_+(w)$ and $q_-(w)$ in the different models.

The above calculation might suggest that we can improve performance even more by increasing the value of γ. However, this does not actually improve performance because diffusion, not drift, dominates the dynamics for larger γ, producing a $1/\alpha^2$ dependence (**Fig. 5**). For $\gamma = 5$, the diffusion process again dominates, and the dependence on $1/\alpha$ is still quadratic in the balanced case and linear in the unbalanced case. Notice that the estimate of τ of equation (12) is based on the assumption that the dynamics are entirely dominated by drift. In the case in which diffusion also plays a role, the estimate is no longer valid.

DISCUSSION

A summary of our results (**Table 1**) indicates that an improvement with small plasticity step size proportional to $1/\alpha^2$ cannot be achieved without some form of fine-tuning that maintains a balance between potentiation and depression. It is more reasonable to expect that memories might be retrievable for a time of order $\ln(n)/(\alpha r)$ (recalling the logarithmic dependence on the number of synapses mentioned in the introduction to this article). Although we discuss primarily memory lifetimes, we have included results for the initial signal-to-noise ratios of the memory trace in **Table 1**. This is because it is often possible to extend memory lifetimes at the expense of the initial signal-to-noise ratio, so both must be considered. Memory storage involves a compromise between plasticity, which improves the initial signal-to-noise ratio, and rigidity, which maintains memories.

The factor $1/\alpha$ is obtained in the unbalanced case by introducing soft bounds that prevent the weights from moving too far away from their equilibrium value. The equilibrium point $\bar{w}$ is where the two synaptic drifts cancel. So far we have restricted our analysis to the case of a single equilibrium point, but we also investigated whether the existence of multiple equilibrium points could significantly improve memory performance. To answer this question, we have analyzed the periodic case ($q_+(w) = q_+(w + \beta)$, $q_-(w) = q_-(w + \beta)$), with $m = 1/\beta$ different equilibria. To simplify the analysis, we assumed that $q_+(w) = q_-(w) = q(w)$. We found that the dependence of memory lifetime on m is similar to the dependence on $1/\alpha$ in the case of hard bounds: in the balanced case the memory lifetime grows quadratically with the number of minima, whereas in the unbalanced case it does not depend on m. The memory lifetime is also inversely proportional to the rate at which the transitions from one equilibrium to the neighboring one occurs. This rate is always larger than the minimum of $q(w)$ over one period[32]. By reducing this rate, the memory lifetime can be extended without limitations. However, this happens at the price of reducing the initial signal-to-noise ratio by the same factor. If the initial signal-to-noise ratio becomes too small, memories cannot be retrieved, even immediately after they have been stored. To conclude, even in the more complex case of multiple equilibrium points, the memory lifetime is of order $\ln(n)/(\alpha r)$, where $1/\alpha$ can be regarded either as the number of synaptic states or the number of equilibria.

Using $n = 10^{12}$ synapses and taking $\alpha = 0.01$ and $1/r = 1$ min allows memories to be recovered over a period of around 2 d, which is still quite modest. We think it is more likely that α is not much smaller than 1 (refs. 33–36), and that memory storage through synapses like the ones we have been describing lasts only for several minutes. It is interesting to note that this is similar to the retention times for explicit memories in people with medial temporal lobe lesions[37]. Perhaps the residual memory seen in such cases is due to storage in the type of synapses we have been studying. If so, it is also interesting to note that there is no need for the potentiation and depression responsible for such memory storage to persist for more than the several minutes that memories last before ongoing plasticity erases them. In other words, there is no need for the synapses to be more persistent than the memories they store.

We can extend our analysis to consider the possibility of synapses being created and destroyed[38] by introducing an additional state that corresponds to the absence of the synapse. This can then be treated in the same way as the other synaptic states. In this case, the memory lifetime is still of the order of $\ln(n)/(\alpha r)$ with roughly the same α, but n can now be regarded as the number of potential synapses, which can be as large as 10^{20}. Even if we consider all these synapses, we still get a modest memory lifetime because the increase in the logarithm is a factor less than 2. If the transitions to and from the special state are slow, the situation is similar to the case of multiple equilibria discussed previously: the memory lifetime is controlled by the lowest transition rate and can be very long, but at the price of a reduced initial signal-to-noise ratio.

In our study, we considered a mnemonic trace (the signal) that is read out by comparing the average synaptic weight of synapses potentiated by the tracked memory to the average of synapses that have been depressed. One might wonder whether the variance of the synapses or some other higher-order statistics of the distribution might contain a mnemonic trace that lasts longer than such averages (a first-order statistics). Although we cannot answer this question in the most general case, we believe that it is not possible. An analysis of second-order statistics shows that the subdominant eigenvalue of the Markov matrix controlling the convergence to equilibrium of the variance of the synaptic weights is the same as the subdominant eigenvalue that controls the decay of the mean, even if we consider correlations between different synapses on the same dendritic tree.

How, then, are memories stored for longer periods of time? One possibility is that long-lasting memories are protected from overwriting by some mechanism[39,40], or that synapses are modified in a clever way by exploiting the feedback of an internal or external supervisor that knows which synapses have to be modified to store a new memory without compromising the old ones (for example, the mechanism of learning of the perceptron[41–43]). A second possibility is that the amount of information acquired with each stored memory is small, to the point that previous memories are essentially left untouched. This can be achieved either by reducing the amount of information contained in each memory (for example, in the case of sparse patterns of neural activity[6,44]) or by reducing the rate at which synapses are modified (for example, by randomly selecting only a small fraction of synapses whose modifications are consolidated[5,6,45]). In the latter case, it is possible to store a large number of memories, provided that the events that generate them are repeated a large number of times[26]. This is yet another expression of the tradeoff between memory lifetime and signal-to-noise ratio discussed earlier: slowing down learning can extend memory lifetimes, but this happens at the price of reducing the amount of information stored after every synaptic modification.

Our results indicate that good memory performance requires multistate synapses, but that the multiple states should not differ simply in their synaptic efficacy. We have suggested elsewhere that longer-term memory storage is possible using synapses the combine plasticity with metaplasticity[46–48] in a cascade of states[25]. The experimental implication of our results is that connecting synaptic plasticity to memory requires more than simply accumulating evidence about long-lasting modifications of synaptic efficacy. Rather, we must map out the full synaptic state space and focus on transitions not merely between states with different strengths but, more importantly, between states that are subject to different degrees and forms of plasticity.

METHODS

The model. We study the forgetting process by considering a scenario in which synapses are exposed to a continuous stream of events, each generating a memory. We select a particular memory and track its mnemonic trace while other memories are being stored. In particular we consider the synapses that are modified when the tracked memory is stored. Some of these are potentiated (group 'pot', which is a fraction f_+ of all synapses) and others are depressed (group 'dep', which is a fraction f_-):

$$w_i \to w_i + q_+(w_i),\ i \in \text{pot}$$

$$w_i \to w_i - q_-(w_i),\ i \in \text{dep}$$

where w_i is the synaptic strength of synapse i and q_+, q_- are the amounts by which the synaptic strength is modified. In general, these factors depend on the synaptic strength preceding the synaptic update. When other memories are stored, we assume that the synaptic modifications are random and uncorrelated with the modifications that generated the tracked memory. In particular, every synapse is modified at rate r (that is, in a time interval t each synapse is updated rt times on average). Every time the synaptic strength w is modified, it is either potentiated with probability f_+ or depressed with probability f_-:

$$w \to w + q_+(w) \text{ with probability } f_+$$

$$w \to w - q_-(w) \text{ with probability } f_-$$

Both synapses in group pot and in group dep are updated in the same way, other than when the tracked memory is stored.

Signal and noise. The memory trace reflects the difference between the synapses that have been potentiated and the synapses that have been depressed by the tracked memory. The memory signal is a measure of this difference defined as the average distance S_{pot} separating the potentiated synapses from their equilibrium value minus the average distance S_{dep} between the depressed synapses and their equilibrium values:

$$S(t) = S_{\text{pot}}(t) - S_{\text{dep}}(t) \tag{15}$$

where:

$$S_{\text{pot/dep}}(t) = \frac{1}{n}\sum_{i=\text{pot/dep}} (w_i(t) - \overline{w}) \tag{16}$$

The noise is defined as the s.d. of the signal:

$$N(t) = \sqrt{\frac{1}{n}\sum_{i=pot} (w_i(t) - \overline{w})^2 - S_{pot}^2 + \frac{1}{n}\sum_{i=dep} (w_i(t) - \overline{w})^2 - S_{dep}^2} \tag{17}$$

Mean-field analysis. The mean-field analysis was performed on a discrete version of the synaptic model that is equivalent to the continuous model described in the text. We discretized the synaptic weights by introducing $1/\alpha$ states with synaptic strengths 0, α, 2α, …1. Every time the synapses are updated, transitions from one state to one of the neighboring states occur stochastically:

$$w \to w + \alpha \text{ with probability } f_+ q_+(w)/\alpha$$

$$w \to w - \alpha \text{ with probability } f_- q_-(w)/\alpha$$

This model is equivalent to the continuous one if α is small enough. We now introduce the occupancy F_k ($k = 1,\ldots, 1/\alpha$), defined as the fraction of synapses in the k^{th} synaptic state. The occupancies for the synapses that have been potentiated by the event generating the tracked memory are denoted by F_k^+. Analogously, F_k^- are the occupancies for the synapses that have been depressed. When the subscript index k is dropped, $F^{\pm}$ denotes a vector. The signal is estimated by

$$S(t) = n[f_+ W^T(F^+(t) - F^\infty) - f_- W^T(F^-(t) - F^\infty)]$$

where W is a vector containing the strengths of the synaptic states ($W = 0$, α, 2α,…, 1) and F^∞ is the set of occupancies for the equilibrium distribution.

To compute the occupancies at time t we proceed as described in refs. 6 and 7: every time a new memory is stored, the distribution of synaptic weights is updated as a Markov process. Because the mean-field dynamics is the same for F^+ and F^-, we drop the superscript index $\pm$ to simplify the notation. Whenever a new memory is stored, the vector F of occupancies of the synaptic states undergoes the transformation

$$F^T \to F^T M$$

where M is the matrix of transition probabilities and is 0 everywhere, except on the diagonal where $M_{kk} = 1 - f_+q_+(W_k) - f_-q_-(W_k)$, and around the diagonal where $M_{k,k+1} = f_+q_+(W_k)$, $M_{k,k-1} = f_-q_-(W_k)$. After the storage of rt memories,

$$F^T \to F^T M^{rt}$$

In order to compute M^{rt} we use its spectral decomposition

$$M = \sum_{k=1}^{m} \lambda_k u_k v_k^T$$

where u and v are, respectively, the left and the right eigenvectors of M

$$Mv_k = \lambda v_k,\ u_k^T M = \lambda u_k^T,\ v_k^T u_l = \delta_{kl}$$

Using this,

$$M^{rt} = \sum_{k=1}^{m} \lambda_k^{rt} u_k v_k^T$$

If the Markov process is irreducible[6,7], there is a single eigenvalue equal to 1, whose right eigenvector corresponds to the equilibrium distribution F^∞. All the other eigenvalues are smaller. The largest of these, called the subdominant eigenvalue, determines the speed of convergence to the equilibrium distribution, and hence the memory lifetime. In other words, for rt large enough, we find that

$$(F(t))^T \approx (F^\infty)^T + (F(0))^T \lambda_M^{rt} \mu_M v_M^T$$

where $F(0)$ is the distribution after the tracked memory is stored and λ_M is the subdominant eigenvalue. We can then rewrite the signal as

$$S(t) \approx S(0)\lambda_M^{rt} \approx S(0)\exp(-rt(1-\lambda_M)) = S(0)\exp(-t/\tau)$$

where $\tau = 1/(r(1-\lambda_M))$ is the memory lifetime (here we have assumed that λ_M is close to 1). The expected value of the noise is

$$N(t) = \sqrt{nf_+[(W^2)^T F^+(t) - (W^T F^+(t))^2] + nf_-[(W^2)^T F^-(t) - (W^T F^-(t))^2]}$$

where $(W^2) = 0$, α^2, $(2\alpha)^2$, $(3\alpha)^2$,…1. The noise depends weakly on t because F^∞ does not cancel out as it does in the case of the signal[6]. Hence, for large t we can rewrite the noise as

$$N = \sqrt{n[(W^2)^T F^\infty - (W^T F^\infty)^2]}$$

The signal-to-noise ratio scales as

$$\frac{S(t)}{N(t)} \approx \sqrt{n}\exp(-t/\tau)$$

Such a quantity decreases exponentially with t and it remains above some threshold θ as long as

$$t < \tau\left(\frac{1}{2}\log n - \log\theta\right)$$

The total time for which the signal-to-noise ratio is larger than θ is proportional to τ and to log n (which reproduces the result stated above).

Mean-field simulations. In the mean-field simulations we computed numerically $F^{\pm}(t)$ for all times in order to compute the signal and the noise. We started from $F^{\pm} = F^\infty$. Then, for the synapses which have to be potentiated, $F^+(0)$ is given by

$$F^+(0)^T = (F^\infty)^T M_+$$

where M_+ is M with $f_+ = 1$ and $f_- = 0$. Analogously for M_-, we impose $f_+ = 0$ and $f_- = 1$. For the next time steps we used the equation

$$F^{\pm}(t)^T = (F^{\pm}(0))^T M^{rt} = (F^{\infty})^T M_{\pm} M^{rt}$$

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We are grateful to M. Mattia for useful discussions about Brownian particles in periodic potentials. This research was supported by US National Institute of Mental Health grant 58754 and by a US National Institutes of Health Director's Pioneer Award, part of the NIH Roadmap for Medical Research, through grant number 5-DP1-OD114-02.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Bliss, T.V. & Collingridge, G.L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
2. Bredt, D.S. & Nicoll, R.A. AMPA receptor trafficking at excitatory synapses. *Neuron* **40**, 361–379 (2003).
3. Amit, D.J. *Modeling Brain Function* (Cambridge University Press, New York, 1989).
4. Hertz, J., Krogh, A. & Palmer, R.G. *Introduction to the Theory of Neural Computation* (Addison Wesley Longman, Boston, 1991).
5. Amit, D.J. & Fusi, S. Constraints on learning in dynamic synapses. *Network* **3**, 443–464 (1992).
6. Amit, D.J. & Fusi, S. Learning in neural networks with material synapses. *Neural Comput.* **6**, 957–982 (1994).
7. Fusi, S. Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. *Biol. Cybern.* **87**, 459–470 (2002).
8. Staubli, U. & Lynch, G. Stable depression of potentiated synaptic responses in the hippocampus with 1–5 Hz stimulation. *Brain Res.* **513**, 113–118 (1990).
9. Larson, J., Xiao, P. & Lynch, G. Reversal of LTP by theta frequency stimulation. *Brain Res.* **600**, 97–102 (1993).
10. O'Dell, T.J. & Kandel, E.R. Low-frequency stimulation erases LTP through an NMDA receptor–mediated activation of protein phosphatases. *Learn. Mem.* **1**, 129–139 (1994).
11. Xiao, M.Y., Niu, Y.P. & Wigstrom, H. Activity-dependent decay of early LTP revealed by dual EPSP recording in hippocampal slices from young rats. *Eur. J. Neurosci.* **8**, 1916–1923 (1996).
12. Zhou, Q., Tao, H.W. & Poo, M-m. Reversal and stabilization of synaptic modifications in a developing visual system. *Science* **300**, 1953–1957 (2003).
13. Barnes, C.A. Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat. *J. Comp. Physiol. Psychol.* **93**, 74–104 (1979).
14. Ahissar, E. *et al.* Dependence of cortical plasticity on correlated activity of single neurons and on behavioral context. *Science* **257**, 1412–1415 (1992).
15. Manahan-Vaughan, D. & Braunewell, K.H. Novelty acquisition is associated with induction of hippocampal long-term depression. *Proc. Natl. Acad. Sci. USA* **96**, 8739–8744 (1999).
16. Fu, Y.-X. *et al.* Temporal specificity in the cortical plasticity of visual space representation. *Science* **296**, 1999–2003 (2002).
17. Xu, L., Anwyl, R. & Rowan, M.J. Spatial exploration induces a persistent reversal of long-term potentiation in rat hippocampus. *Nature* **394**, 891–894 (1998).
18. Abraham, W.C., Logan, B., Greenwood, J.M. & Dragunow, M. Induction and experience-dependent consolidation of stable long-term potentiation lasting months in the hippocampus. *J. Neurosci.* **22**, 9626–9634 (2002).
19. Villarreal, D.M., Do, V., Haddad, E. & Derrick, B.E. NMDA receptor antagonists sustain LTP and spatial memory: active processes mediate LTP decay. *Nat. Neurosci.* **5**, 48–52 (2002).
20. Jenkins, J. & Dallenbach, K. Oblivescence during sleep and waking period. *Am. J. Psychol.* **35**, 605–612 (1924).
21. Brown, M.W. & Xiang, J.Z. Recognition memory: neuronal substrates of the judgement of prior occurrence. *Prog. Neurobiol.* **55**, 149–189 (1998).
22. Wixted, J.T. & Ebbesen, E.B. Genuine power curves in forgetting: a quantitative analysis of individual subject forgetting functions. *Mem. Cognit.* **25**, 731–739 (1997).
23. Bi, G.-Q. & Poo, M.-M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, ad postsynaptic cell type. *J. Neurosci.* **18**, 10464–10472 (1998).
24. Parisi, G. A memory which forgets. *J. Phys. A.* **19**, L617–L620 (1986).
25. Fusi, S., Drew, P.J. & Abbott, L.F. Cascade models of synaptically stored memories. *Neuron* **45**, 599–611 (2005).
26. Brunel, N., Carusi, F. & Fusi, S. Slow stochastic Hebbian learning of classes of stimuli in a recurrent neural network. *Network* **9**, 123–152 (1998).
27. Fusi, S. & Senn, W. Eluding oblivion with smart synaptic updates. *Chaos* **16**, 026112 (2006).
28. Kahn, P.E., Wong, K.Y.M. & Sherrington, D. A memory model with novel behaviour in sequential learning. *Network. Comput. Neural Sys.* **6**, 415–427 (1995).
29. van Rossum, M.C., Bi, G.Q. & Turrigiano, G.G. Stable Hebbian learning from spike timing-dependent plasticity. *J. Neurosci.* **20**, 8812–8821 (2000).
30. Rubin, J.E. Steady states in an iterative model for multiplicative spike-timing dependent plasticity. *Network* **12**, 131–140 (2001).
31. Gutig, R., Aharonov, R., Rotter, S. & Sompolinsky, H. Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *J. Neurosci.* **23**, 3697–3714 (2003).
32. Festa, R. & Galleani D'Agliano, E. Diffusion coefficient for a brownian particle in a periodic field of force. *Physica A* **90A**, 229–244 (1978).
33. Petersen, C.C., Malenka, R.C., Nicoll, R.A. & Hopfield, J.J. All-or-none potentiation at CA3–CA1 synapses. *Proc. Natl. Acad. Sci. USA* **95**, 4732–4737 (1998).
34. O'Connor, D.H., Wittenberg, G.M. & Wang, S.S.-H. Graded bidirectional synaptic plasticity is composed of switch-like unitary events. *Proc. Natl. Acad. Sci. USA* **102**, 9679–9684 (2005).
35. O'Connor, D.H., Wittenberg, G.M. & Wang, S.S.-H. Dissection of bidirectional synaptic plasticity into saturable unidirectional processes. *J. Neurophysiol.* **94**, 1565–1573 (2005).
36. Montgomery, J.M. & Madison, D.V. Discrete synaptic states define a major mechanism of synapse plasticity. *Trends Neurosci.* **27**, 744–750 (2004).
37. Scoville, W.B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).
38. Chklovskii, D.B., Mel, B.W. & Svoboda, K. Cortical rewiring and information storage. *Nature* **431**, 782–788 (2004).
39. Willshaw, D.J., Buneman, O.P. & Longuet-Higgins, H.C. Non-holographic associative memory. *Nature* **222**, 960–962 (1969).
40. Grossberg, S. Processing of expected and unexpected events during conditioning and attention: a psychophysiological theory. *Psychol. Rev.* **89**, 529–572 (1982).
41. Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386–408 (1958).
42. Block, H. The perceptron: a model for brain functioning. I. *Rev. Mod. Phys.* **34**, 123–135 (1962).
43. Minsky, M.L. & Papert, S.A. *Perceptrons* (MIT Press, Cambridge, Massachusetts, 1969; expanded edition, 1988).
44. Tsodyks, M.V. & Feigelman, M.V. The enhanced storage capacity in neural networks with low activity level. *Europhys. Lett.* **6**, 101–105 (1988).
45. Tsodyks, M.V. Associative memory in neural networks with binary synapses. *Mod. Phys. Lett. B* **B4**, 713–716 (1990).
46. Abraham, W.C. & Bear, M.F. Metaplasticity: the plasticity of synaptic plasticity. *Trends Neurosci.* **19**, 126–130 (1996).
47. Fischer, T.M., Blazis, D.E., Priver, N.A. & Carew, T.J. Metaplasticity at identified inhibitory synapses in Aplysia. *Nature* **389**, 860–865 (1997).
48. Montgomery, J.M. & Madison, D.V. State-dependent heterogeneity in synaptic depression between pyramidal cell pairs. *Neuron* **33**, 765–777 (2002).