Learning rules and network repair in spike-timing-based computation networks

J. J. Hopfield[†‡§] and Carlos D. Brody[‡§¶]

[†]Department of Molecular Biology, Princeton University, Princeton, NJ 08544-1014; and [¶]Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724

Contributed by J. J. Hopfield, October 1, 2003

Plasticity in connections between neurons allows learning and adaptation, but it also allows noise to degrade the function of a network. Ongoing network self-repair is thus necessary. We describe a method to derive spike-timing-dependent plasticity rules for self-repair, based on the firing patterns of a functioning network. These plasticity rules for self-repair also provide the basis for unsupervised learning of new tasks. The particular plasticity rule derived for a network depends on the network and task. Here, self-repair is illustrated for a model of the mammalian olfactory system in which the computational task is that of odor recognition. In this olfactory example, the derived rule has qualitative similarity with experimental results seen in spike-timing-dependent plasticity. Unsupervised learning of new tasks by using the derived self-repair rule is demonstrated by learning to recognize new odors.

Networks of neurons with modifiable synapses have an implicit self-repair problem. Synapses are constantly made and unmade, with a high turnover rate (ref. 1, but see also ref. 2). This high rate of turnover allows a network to learn and adapt to its environment, but it also allows noise to degrade a connectivity pattern that implements a useful behavioral competency. Suppose a connection that should be present for network operation disappears. How can an appropriate equivalent replacement connection be chosen on the basis of the activity of the network? We show here that time-dependent synaptic plasticity rules for self-repair of a network of spiking neurons may be derived from the firing times of a correctly working system.

In principle, plasticity rules for self-repair of a working system could be different from rules for learning of new functions. However, we show here that rules for self-repair also can be used as the basis of unsupervised learning of new tasks.

A Functioning Network for Studying the Repair Problem

As a test bed to explore the self-repair and *de novo* learning problems, we will use a network recently proposed as a model of the olfactory bulb (3). The approach is nevertheless a general one and can be directly applied elsewhere. For completeness, we will briefly describe the olfactory model itself, although the derivation and use of the plasticity rules, not being specific to the olfactory network, do not require an understanding of the overall design of the network.

The task performed by this network is that of stimulus recognition; the output neurons of the olfactory model are highly selective for specific odors. We will consider here a reduced network, adequate for the recognition of a single stimulus, that has a single postsynaptic cell and a multiplicity of presynaptic cells. Olfactory stimuli are encoded as a set of currents injected into the presynaptic cells, and a robust firing response from the postsynaptic cell signals recognition of its target stimulus. This recognition is highly selective, is robust to changes in odor concentration over a 50-fold range, and is robust to the presence of strong background odors. The principles behind these robustness features of the olfactory model are described elsewhere (3).

All presynaptic cells produce action potentials at comparable rates in the presence of any stimulus, and all could potentially be connected to the postsynaptic cell. To recognize a given odor, a particular subset of the presynaptic cells is chosen to make functional connections to the postsynaptic cell, with equal-strength synapses. It is the identity of the connected cells that defines the odor to be recognized. We thus focus on the presence or absence of functional connections and assume that all functional connections are of equal strength. We seek an automatic self-repair process. Based only on the pre- and postsynaptic cell firing patterns, this process should maintain, in the face of random synapse addition or deletion, connections from an appropriate set of presynaptic cells, such that the recognized odor remains fixed. The issue is illustrated in Fig. 1.

In general, the only information available to an automatic repair rule will be the spike times of presynaptic cells, both connected and unconnected, and the spike times of the postsynaptic cell. How can this information be used to derive a self-repair rule?

Given the nature of the information available, we will assume that the self-repair rule depends on the relative timing of pre- and postsynaptic spikes. When the system is sequentially exposed to a diversity of stimuli, nontarget stimuli produce no (or very few) spikes in the postsynaptic cell, so no repair of synapses to the postsynaptic cell will take place after a nontarget stimulus presentation. Repair will take place only when the target stimulus occurs, generating a significant number of postsynaptic spikes. Due to noise, some functional connections are lost between presentations of the target stimulus. We ask that a set of connections functionally equivalent to the original engineered set be relearned. That is, after a long time during which the system has been exposed many times to many stimuli, and most of the original connections have been lost and replaced, the postsynaptic cell should preserve its ability to recognize and discriminate the same stimulus that it initially recognized.

Another description of the idea would allow each exposure to the target stimulus to remodel all the synapses, making strong synapses of some of the previously silent ones (or vice versa) according to a repair rule. Again, the functionality of the postsynaptic cell should be preserved. This complete remodeling formulation leads to the same plasticity rule.

We will make the assumption of pairwise additivity: we assume that the effects of all presynaptic–postsynaptic spike pairs that occur during a time window of relevance to a single-trial learning protocol (in our case ≈0.5 s) are additive. Deviations from additivity in spike-timing-dependent plasticity (STDP) have been reported, but, for the simple task we have chosen, the particular kind of deviation observed (4, 5) has little effect.

Deriving the Repair Rule

The problem the repair rule must solve is that of taking a pre- and a postsynaptic spike train and correctly classifying the

Abbreviation: STDP, spike-timing-dependent plasticity.

[‡]J.J.H. and C.D.B. contributed equally to this work.

[§]To whom correspondence may be addressed. E-mail: hopfield@princeton.edu or brody@cshl.edu.

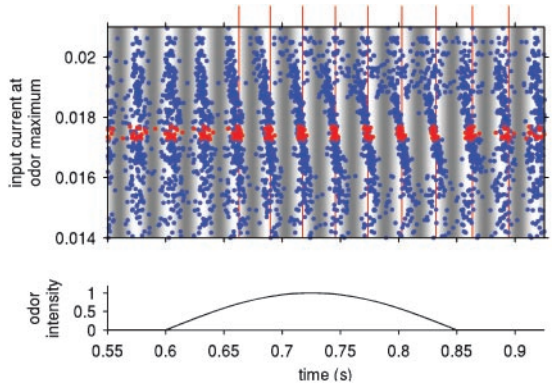

Fig. 1. Spiking network model of odor recognition. (*Upper*) Spike rasters of all presynaptic cells during a single trial (a sniff of the target odor). Spikes from cells that do and do not have a functional connection to the postsynaptic cell are shown in red and blue, respectively. Different presynaptic cells receive different peak currents during the odor sniff. Cells have been sorted vertically by the magnitude of that peak current. Vertical red lines indicate postsynaptic cell firing times. The essence of the self-repair problem is to automatically determine whether a presynaptic spike train belongs to the blue or the red raster set and, therefore, whether the presynaptic cell producing it should have a functional connection to the postsynaptic cell. A common underlying subthreshold oscillation promotes a systematic relationship between injected current and phase of firing of the presynaptic cells (indicated by the gray background). For any odor, there is a subset of presynaptic cells that, at the peak of the sniff, will fire at the same phase with respect to the oscillation, will therefore be synchronized, and can be used to drive an odor-selective postsynaptic cell (3). (*Lower*) The strength of the stimulus during the sniff is plotted.

presynaptic train as from a cell that would be "appropriate" (vs. "inappropriate") to connect functionally to the postsynaptic cell. Due to noise, this classification cannot be completely precise, but a substantial majority of such classifications need to be correct.

Consider a synapse in a case in which both the pre- and the postsynaptic cell produce one spike in response to a stimulus presentation. The decision as to whether this presynaptic cell is appropriate for connecting to the postsynaptic cell will be based on a function $W(\Delta t)$, where Δt is the time difference between the post- and the presynaptic spike. We will determine $W(\Delta t)$ from a large body of data on spike trains from presynaptic cells known to be either appropriate or inappropriate for connection.

In general, there may be more than one spike from the pre- or postsynaptic cell under consideration. There will then be several pre- and postspike pairings, with different values of postspike minus prespike time difference Δt_j, where the index j runs over the different pairs. We define a quantity M, on the basis of which the decision on appropriateness of connecting will be made. From the pairwise additivity assumption, M is given by

$$M = \sum_j W(\Delta t_j). \qquad [1]$$

We discretize the possible time intervals Δt of presynaptic–postsynaptic spike pairs into time bins of length δ indexed by k. The unknown function $W(\Delta t)$ can then be described in terms of a set of unknown parameters w_k, where $w_k = W(k\delta)$. We have used 75 bins with a width of 0.4 ms, with centers in the range of 15 ms before to 15 ms after the approximate time of the postsynaptic cell spike. [Synaptic currents were fast, and the postsynaptic cells in the olfactory model had a short membrane time constant, so a presynaptic spike >15 ms before a postsynaptic spike had little effect on the timing of that postsynaptic spike. Time differences of >15 ms are similarly presumed to have no effect on plasticity. If the postsynaptic cells have long time constants or synaptic currents are slow, the range used in determining $W(\Delta t)$ should be made correspondingly greater.] The pairings of a presynaptic cell's spikes and the postsynaptic cell's spikes can now be described by a set of integers n_k describing how many pairings occurred within each time bin k. In these terms, M is given by

$$M = \sum_k w_k n_k. \qquad [2]$$

For a given postsynaptic spike train, each presynaptic spike train will then produce a value of M. The parameters w_k must produce values of M that classify the spike patterns into two sets, those coming from presynaptic cells with appropriate connections and those coming from presynaptic cells with inappropriate connections.

Formulated in this fashion, our problem is exactly the mathematical problem of pattern classification by a feed-forward "artificial neural network" having no "hidden" units. The n_k values are the inputs, the w_k values are the "weights," and the M value is the input to the output "unit." We have used a procedure (6) that trains an output unit to predict the probability that a presynaptic cell belongs to the class appropriate for connection. (Because of noise, it is possible that cells belonging to both appropriate and inappropriate classes can sometimes generate the same pattern of n_k values.) We later will use the value of M to prescribe which connections are made in learning.

The prediction made by the artificial neural network is taken to be the logistic function

$$P(\text{appropriate connection}) = 1/(1 + e^{-M}). \qquad [3]$$

The weights w_k are obtained through "learning" (described below) on a training set of preclassified spike trains. The training set was obtained from a working olfactory model in which connections were engineered to produce an odor-selective postsynaptic cell (3). We refer to this connection pattern as the engineered solution. The presynaptic cells (and their spike trains) engineered to have functional connections to the postsynaptic cell were labeled appropriate, and all other presynaptic cells and spike trains were labeled inappropriate. The postsynaptic and the full set of presynaptic spike trains produced in the engineered network in response to the postsynaptic cell's target odor were recorded. The resulting labeled spike trains then were used in an iterative weight change rule for the artificial neural network, as follows: For each presynaptic spike train example, change each weight w_j according to

$$\delta w_j \sim ([1 \text{ or } 0] - P) * n_j$$

$$(1 \text{ if an appropriate example, } 0 \text{ if inappropriate}). \qquad [4]$$

This procedure (6) minimizes the *K–L* distance (7) between the network-defined probability P and the actual probability distribution, without the necessity of explicitly defining the actual probability distribution. In this structure of a feed-forward network with the *K–L* measure of error, gradient descent in weight space determines the unique best w_k.

Training the artificial neural network is a mathematical procedure that allows us to derive the optimal plasticity rule. Biology is likely to find optimal rules through evolution.

Fig. 2*a* plots the optimal weights w_k derived from this procedure when applied to spike trains such as those in Fig. 1. In the olfactory example, the total number of spikes is very similar for both appropriate-connection and inappropriate-connection cells, so the strength of the connection to a bias unit can be traded against the addition of a constant to each of the weights w_k, with no change in the classification performance of the network. We chose w_b so that $W(\Delta t)$ goes to 0 for large Δt (i.e., so that pairings too far apart in time will be ineffective). With this

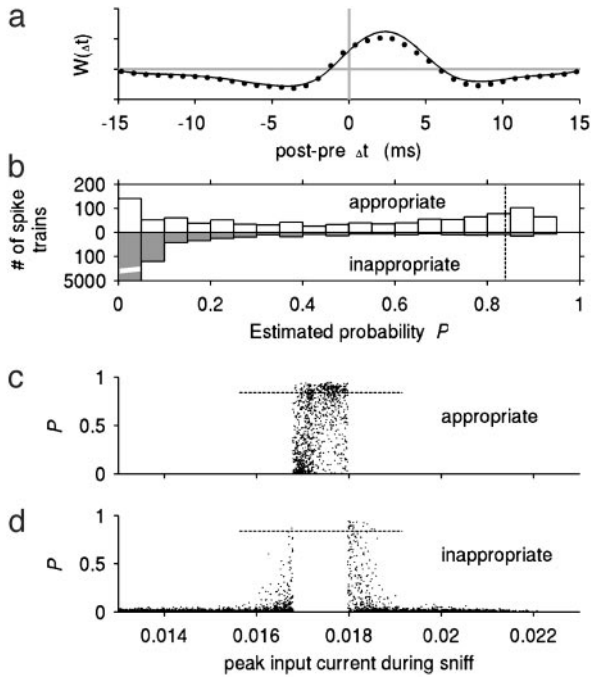

Fig. 2. (*a*) The weights w_k after training on the classification problem, plotted against the time difference Δt of the post- and presynaptic spike pairs. Dots indicate an evaluation of the weights based on the spike trains from a single stimulus presentation. The solid line is a spline fit through w_k points averaged over 16 runs, and it should closely approximate the function $W(\Delta t)$. In the training set used, there were ≈1,000 appropriate presynaptic spike train examples and 4,600 inappropriate spike train examples. (*b*) Histogram of the number of training set examples having estimated probability P of belonging to the appropriate class (determined by using the w_k shown in *a* and Eqs. **2** and **3**). The known appropriate and inappropriate spike trains are shown. The vertical dashed line indicates the cutoff threshold for P, defining the top-ranked presynaptic cells for making functional connections. (*c* and *d*) For the olfactory network, selective odor recognition corresponds to choosing connections from cells with a narrow range of peak input currents, as compared with the total possible range. We plot the estimated probability P as a function of the peak input current of the presynaptic cells. Choosing cells with an estimated P above the threshold setting (indicated by the dashed line in *c* and *d*) also chooses cells with peak currents within a narrow range. In general, such an underlying principle allowing visualization in plots such as those in *c* and *d* might not be known.

choice, it was also the case that $\Sigma w_k \approx 0$. Thus, the mere existence of a presynaptic–postsynaptic pairing at some point in the interval contained no evidence about the probability of the presynaptic cell belonging to the appropriate class. It is only the value of Δt that contained evidence of to which of the two classes a cell belonged, and all decision information was in the timing domain, not in the number of spike pairs *per se*.

The qualitative nature of the shape of the w_k vs. Δt plot in Fig. 2*a* could be anticipated from the data shown in Fig. 1 or in related earlier work (3). The neurons in the designed solution on average fire in synchrony when the odor is present and will induce the postsynaptic cell to fire after the integration of the excitatory synaptic currents they generate. A positive peak is thus expected near the peak of the integrated synaptic current, which for this system occurs at 3 ms. Presynaptic neurons that systematically fire after the postsynaptic cell cannot contribute to the functionality of the network and should be discriminated against. This fact leads one to expect a negative peak within the postsynaptic minus presynaptic Δt region.

The data shown in Fig. 2 *b–d* display the classifications of 5,600 presynaptic cells. On combination with the spikes of the postsynaptic cell, the spikes of each presynaptic cell generate a spikes-pair pattern vector n_k. On the whole, appropriate examples are assigned much higher estimated probabilities of being appropriate examples than are inappropriate examples (Fig. 2*b*). Very few inappropriate examples are assigned high probability.

Whereas the shape of this synapse-choice function has qualitative similarity to the shape of the synapse-change timing relationship seen in STDP (8), there are some important quantitative differences. Both rules favor making connections when postsynaptic spikes occur after pre- and suppressing connections when postsynaptic spikes occur before pre-. The optimal derived rule, however, contains a slight positive spillover into the negative Δt region that has important implications for the long-term stability of the system (see *Long-Term Stability*). The optimal shape of the timing rule, as derived here, depends on the task being performed, the synaptic and cell time constants, and the level of noise present. We have found that increasing the noise-induced uncertainty in the firing time of the presynaptic cells broadens the positive peak in Fig. 2*a* and that decreasing the width of the band of positive examples in Fig. 2 *c* and *d* sharpens this peak. The biological system has delay from the back-propagating action potential and synaptic delay, additional issues that must be taken into account at the millisecond level when comparing experiment with theory.

During each iteration of the self-repair task, the value of M for each presynaptic spike train is used to rank the probability of its corresponding to a cell appropriate for connection. The choice of making a functional connection is based on this ranking. In the particular case of the olfactory model, the top 200 cells were chosen for functional connections. There is thus a value of M that acts as a threshold. Information about the degree to which a cell's M value is above or below the threshold is discarded. Thus, although $W(\Delta t)$ is a well defined function to use in estimating a probability, in the replacement task most of the detailed structure of the curve of $W(\Delta t)$ is not relevant. To see why, consider the case where pre- and postsynaptic cells fire only one spike. The relative timing of a presynaptic spike with respect to the postsynaptic spike, Δt, then precisely defines the M value for a cell, which will be $W(\Delta t)$ (see Fig. 2*a*). The choice of which cells to include for connections would depend only on whether $W(\Delta t)$ for a cell was above or below the threshold. In such a case, the shape of $W(\Delta t)$ above or below the threshold is irrelevant. When many correlated action potential pairs are involved, the same effect may continue to be a dominating influence. For this reason, curves as different as the rules of Figs. 2*a* and 4*c* can be similarly successful in a replacement task.

Applying the Repair Rule: Functional Properties of a Network Composed of Fully Replaced Synapses

We will examine stability and self-correction abilities under a protocol in which all synapses are remodeled after presentation of a stimulus. We use the olfactory model as a case study and test bed. Consider a system repeatedly exposed to odors randomly chosen from a set of odors a, b, c, d, etc. Begin with an engineered set of synapses designed to recognize odor c. In our system, this set has 200 functional synapses and 5,400 potential but silent synapses. The postsynaptic cell recognizing odor c will not respond to odors a, b, d, etc. The connections to this cell will change only when that cell fires and thus (at least initially) only when odor c is present. When odor c occurs, the postsynaptic cell fires and we then eliminate all synaptic connections to it. We use the prediction algorithm to rank the probability that each of the 5,600 presynaptic cells is appropriate for connecting to the postsynaptic cell, and we then make functional connections from the 200 most highly probable presynaptic cells.

Fig. 3*a* illustrates the selectivity produced by the original connections: of 500 random odors, the postsynaptic cell produced spikes only for one, the target odor. An iteration of the protocol in which all functional connections are replaced by the

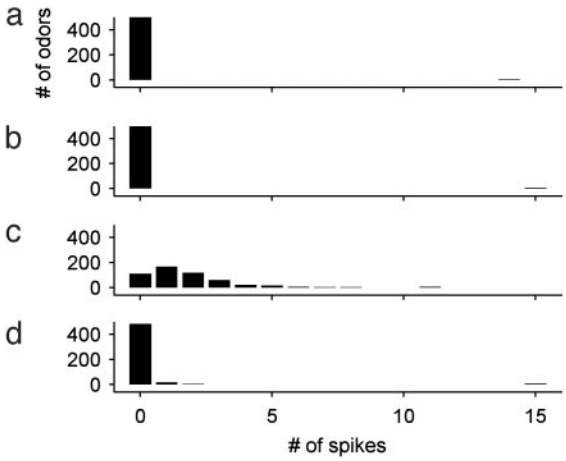

Fig. 3. Histograms of the number of odors producing *n* spikes in the postsynaptic cell; 499 random odors plus the postsynaptic cell's target odor were used. (*a*) Responses using the original, engineered connections are shown (3). (*b*) Responses using the connections produced after one iteration of complete functional connection replacement are shown. (*c*) Original response of a broadly tuned postsynaptic cell connected to five presynaptic cells to the presentation of 500 random odors. It responded with 11 spikes to one of the odors. (*d*) With learning (one iteration) turned on, when some particular odor drove the broadly tuned cell to produce 11 spikes, the synapses were remodeled. This remodeling led to the odor selectivity shown. The odor that triggered the synaptic change now produces many spikes, but all other odors produce very little response. The triggering odor has become the target odor of a highly selective cell.

200 top-ranked cells, according to the repair rule (Fig. 2 and Eq. **3**), produces a new set of functional connections, of which $\approx 1/3$ are also members of the set of original connections. Fig. 3*b* shows that, with these new functional connections, the odor to which the postsynaptic cell is responsive is still the target odor, and responses to other odors remain negligible. The principles of operation of the network indicate that the actual selectivity will remain very high, >1 in 10^6. Concentration invariant recognition and robustness to background odors also remain (data not shown). Although the replacement connections are not identical to the initial pattern, they are functionally equivalent. The system has a superfluous number of acceptable, equally functional presynaptic cells available, and the iteration chooses a similar, overlapping subset of these.

If a small fraction of the connections is lost, the remaining connections will be sufficient to drive the postsynaptic cell to spike almost exactly as it would have done with all of the connections. This spiking pattern then can be used to appropriately select replacement synapses for those that were lost. Self-repair in this framework requires a reoccurrence of the target stimulus, because there is no other location in the system that contains the fundamental knowledge of what is lost when a connection fails.

Complete replacement of functional synapses may be iterated, occurring each time the target odor is presented. Under such multiple iterations, there are technical issues concerning long-term stability that apply both to our derived plasticity rule (Fig. 2*a*) and to any other plasticity rule. In a later section we return to a consideration of these issues.

An alternative repair procedure would be to make connections to those neurons for which M (see Eq. **2**) surpasses a fixed threshold, rather than to all those necessary to generate a given number of total connections. However, stability is most robust when the total synapse strength driving the cell is maintained near a target value, by a homeostatic mechanism, such as keeping the average activity of the cell stable as synapse number changes (9).

Single-Trial Unsupervised Learning

Can the self-repair timing rule (Fig. 2*a*) be used for *de novo* unsupervised learning? Unsupervised learning is a learning paradigm in which no recognition cell is externally "instructed" when or what to learn. We now show that postsynaptic cells that are not sharply tuned to a specific stimulus can become highly selective for a novel stimulus after it is presented, through the use of our spike-timing plasticity rule.

To use a plasticity rule that depends on pre- and postsynaptic spike timing, there must be postsynaptic spikes occurring at appropriate times. We therefore commenced learning with a postsynaptic cell that was very broadly "tuned," responding to many different stimuli (Fig. 3*c*). This quality was engineered by connecting a few presynaptic cells chosen at random (usually five) to the postsynaptic cell and increasing the strength of each connection so that the total strength of connections to the postsynaptic cell was about the same as in the 200-connection engineered case. Such a postsynaptic cell "cares" about only five components of the odor. It is far easier to find a good random match to five components than to 200, so the cell is much more broadly tuned in the space of odors than would be the case for a cell receiving 200 different inputs. A typical such postsynaptic cell responds to $\approx 1/2{,}000$ of all random odors by producing ≥ 11 spikes during the sniff. Fig. 3*c* shows the number of spikes fired by such a broadly tuned cell in response to 500 different random odors. We assume that a synapse modification process is constantly present and is activated whenever the postsynaptic cell fires more than a threshold number of spikes during a stimulus presentation (i.e., a sniff in the olfactory case). In our simulations, this threshold was set to 11 spikes over a 0.5-s sniff. (Different networks and computations will have different appropriate decision levels.) Integrating the number of spikes over such a timescale can be achieved easily by known mechanisms present in cell biology. Once activated, the synapse modification process is the same as that used in self-repair: namely, all existing functional connections to the postsynaptic cell are deleted, and the 200 top-ranked presynaptic cells (according to the self-repair rule; Fig. 2*a*) are given functional synapses equal in strength to those of the postsynaptic cell.

Fig. 3*d* shows the response of the postsynaptic cell after learning. Where previously it was broadly tuned, it is now highly selective. It fires robustly in response to the particular odor that triggered the synapse change procedure, responding very weakly or not at all to the other odors. As was the case with self-repair, the design of this particular olfactory model (3) leads to the selectivity of this postsynaptic cell being invariant to the concentration of its new target odor, as well as robust to background odors. To be able to learn any new stimulus, a system based on these ideas should have an ensemble of broadly tuned postsynaptic cells, which together cover the space or possible stimuli.

Some applications would be better served by a slower remodeling of the synapses, changing fewer of them or changing their weights by a small amount. A more gradual approach to selectivity results, averaging over the set of stimuli that drive the postsynaptic cell to obtain an averaged target template.

Single-trial *de novo* learning as implemented here is closely related to self-repair. The presynaptic action potentials are the same in both cases: the two cases differ only in the postsynaptic action potentials. In one case (*de novo* learning), the postsynaptic spikes are generated by a broadly tuned cell; in the other case (self-repair), they are generated by a highly selective cell that might be missing a few connections or have a few erroneous functional connections. When the broadly tuned cell is well driven, these two postsynaptic spike trains are very similar to each other, resulting in a selection of new synapses that is very similar in the two cases.

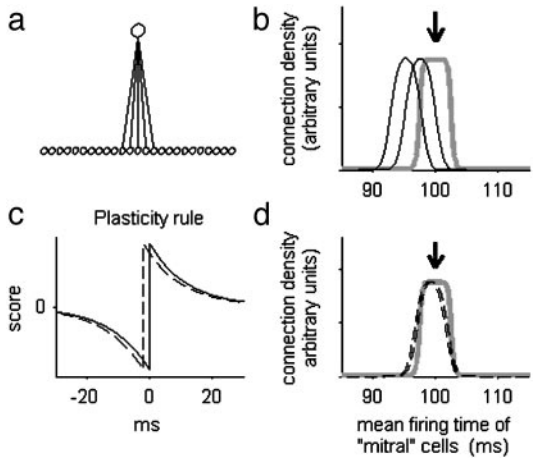

Fig. 4. The problem of drift with STDP rules. (*a*) Schematic of a linear array of neurons, indexed by k, that each fire at a time $t = k + \eta$, where η is jitter due to noise. All of these neurons could potentially make connections to the postsynaptic neuron shown above them, but initially only a small subset, with almost synchronous firing times, makes a functional connection, as shown. (*b*) The gray line indicates the initial density of connections from presynaptic cells to the postsynaptic cell. Presynaptic cells are labeled by their mean firing time, in ms. The arrow indicates the firing time of the postsynaptic cell. Black lines show the connection densities resulting from two successive iterations of the self-repair procedure, using the rule shown as a solid line in *c*. (*c*) The solid line shows a purely causal (all positive in the $\Delta t > 0$ region, all negative in the $\Delta t < 0$ region) plasticity rule. (This plasticity rule is used in *b*, where it is shown that it leads to strong drift.) The dotted line is the same learning rule but shifted 2 ms toward the negative Δt region. (*d*) Same format as in *c*, showing initial connection density and two iterations of self-repair, but now using the plasticity rule shown as a dashed line in panel *c*. Drift is sharply reduced.

Long-Term Stability

When many iterations of self-repair (as described above) were carried out, the selectivity for the target stimulus remained high, but the timing of postsynaptic cell firing often underwent a slow but consistent drift toward earlier times. In the olfactory model, this drift corresponds to drift in the concentration of odor to which the system best responds. In the very long term, this drift can go beyond the dynamic range of the system, causing distortions such that the postsynaptic cell no longer responds to the target odor at any concentration.

The problem of slow drift is relevant to all timing-dependent plasticity rules, not only the particular rule we have derived (10). We can capture the essence of the general drift problem in a highly simplified abstract system where the origin of the drift is apparent. Fig. 4*a* shows the "anatomy" of the system. Neurons k, arranged in sequence along a line, are available for connections to a postsynaptic cell. Each of the neurons k fires a single action potential at time $t = k + \eta$, where η is jitter due to noise. A set of connections from a contiguous clump of these cells is chosen so that an almost synchronous packet of action potentials coming from the clump generates an action potential in the postsynaptic cell at a time labeled $t = 100$ ms. The synapses are iteratively replaced according to the self-repair prescription discussed above (see *Applying the Repair Rule*). An ideal synaptic plasticity rule would keep the set of connections functionally invariant over multiple iterations of synapse replacement, leaving the clump in its original position and with its original width. However, if the width of the clump remains narrow (the equivalent of keeping high odor selectivity and responsiveness in the olfactory system) but the repair process moves the center of gravity of the clump, the postsynaptic firing time will move by the same amount, and long-term drift will occur.

The abstraction of Fig. 4 is closely connected to the situation in neurobiology. A postsynaptic cell has connections or potential connections to many cells. If we consider the case of a particular postsynaptic spike, only the presynaptic spikes near it in time are relevant, and only a single spike in each presynapatic cell is of any interest. The relevant presynaptic cells with functional connections to the postsynaptic cell fired at times close to each other, thus providing the "packet" of incoming spikes and synaptic current that led the postsynaptic cell to fire. If the presynaptic cells then are arranged in a line according to when they spiked, one obtains the idealization sketched in Fig. 4*a*.

We examine the drift problem for different timing-dependent plasticity rules [i.e., different functions $W(\Delta t)$] in Eq. 1. Because the postsynaptic cell fires in response to its presynaptic inputs, a purely "causal" rule relating pre- and postsynaptic spikes, such as that of Bi and Poo (8) (Fig. 4*c*, solid line), might appear adequate. However, it is impossible in practice to guarantee that the postsynaptic cell will fire after the end of the entire presynaptic packet: timing jitter in the pre- or postsynaptic spikes, variations in the strength of the connections, or having more connections than minimally necessary will, in general, cause some presynaptic spikes, from appropriately connected cells, to occur after the postsynaptic spike. Any strictly causal rule systematically rejects such connections and thus leads to a systematic drift of the connection packet along the line of neurons, as shown in Fig. 4*b*. The postsynaptic cell initially fires on average at time $t = 100$ ms with a small noise jitter. With synapse replacement, the time at which the postsynaptic cell fires drifts toward earlier times.

This effect is due to the location of the center of mass of the replacement synapses. If a given spike-timing plasticity rule is shifted to the left or right, it will increase or decrease the rate of drift. Indeed, by merely shifting the spike-timing plasticity rule a bit, the drift can be eliminated. Fig. 4*d* shows the result of the same iterations shown in Fig. 4*b* for the shifted plasticity rule appearing as a dotted line in Fig. 4*c*. The spike-timing rule that we derived (Fig. 2*a*) also showed little drift, because its center of gravity was well located: in deriving the rule, connections from presynaptic cells that tended to fire after the postsynaptic cell but were nevertheless appropriate were included in the appropriate class of the training set. However, we can see from the data of Fig. 4*c* that the center of gravity of the chosen synapses will depend on the level of P at which the decision boundary is placed. A single functional form of $W(\Delta t)$ does not universally solve the problem of drift.

For a given decision threshold and a given shape of $W(\Delta t)$, there will be a precise time-shift location of the replacement rule that leads to no average drift. However, even with this time-shift, there will be nonsystematic random walk-diffusive drift. Thus, truly long-term stability requires further mechanisms to anchor the center of mass of the connections. Possible examples include making a subset of the original connections fixed and nonreplacing (1). Another possibility depends on the fact that the speed of the drift is also a function of the density of cells on the line of presynaptic neurons.

Conclusion

We have shown that a STDP rule that identifies appropriate connections for self-repair can be derived from the timing patterns of a functioning network. The idea of using the pattern of action potentials in a functioning network to derive a plasticity rule appropriate for self-repair is general; the task and network structure will determine what the learning rule is. The particular task and network structure used as a test bed here consists of a model network in which function is largely defined by the identity of the presynaptic cells that make nonsilent connections to a postsynaptic cell. The computation defined by these connections makes use of the relative timing of the connected

presynaptic cells' spikes. Self-repair in such a network requires the use of timing information to specify the identity of appropriate functional connections. The form of the derived rule has qualitative resemblance to experimentally described STDP rules (8, 11).

The direct application of these ideas to experimental data also may be possible. The procedure to derive the self-repair rule required a database of postsynaptic spike trains and presynaptic spike trains, which may be obtained experimentally. It may thus be possible to predict, for specific biological networks, plasticity rules that are optimal for the self-repair of each network and to compare the predicted rules to experimentally measured rules.

The spike-timing rule was derived from self-repair in a task involving recognition of previously known patterns. The same timing rule can be used successfully to learn to recognize a hitherto unknown pattern, in a single exposure or learning trial. The application of the self-repair rule requires knowledge of the spike times of the postsynaptic cell and those of an array of presynaptic cells (connections from a subset of which would produce the observed postsynaptic cell spikes). In the *de novo* learning situation, the presynaptic spikes are the same as in the self-repair situation. What is thus required to apply the rule is a matching set of appropriately timed postsynaptic spikes. A broadly tuned postsynaptic cell can provide such spikes, in essence functioning as a cell that is in bad need of repair. When a broadly tuned cell responds well to a stimulus pattern, the self-repair rule thus can be used as is for *de novo* learning. No explicit instruction to learn is needed: when a broadly tuned cell happens to respond strongly to a pattern, that pattern will be learned, in the sense that the cell now will become narrowly tuned for it. Selectivity for patterns learned in this fashion is in every way equivalent to the properties of a network with connections designed to recognize that pattern.

Purely causal plasticity rules neglect the fact that jitter occasionally will make appropriately connected presynaptic cells fire after the postsynaptic cell. Causal learning rules thus induce a systematic bias that manifests itself as a drift in the firing time of the postsynaptic cell and of the presynaptic cells chosen for connections. Drift due to such a mechanism has been made a useful feature in models in which cells learn to predict (12–14). In the present case, however, such drift is deleterious, as it eventually causes the system to run out of dynamic range.

The shape of the $W(\Delta t)$ was derived from consideration of what synapses, now silent, should be made functional. The interpretation of the $W(\Delta t)$ differs conceptually from the magnitude modification interpretation usually given to similarly shaped curves in experimental STDP studies. Nevertheless, when averaged over many synapses in various states of facilitation, the two are closely related.

This work was supported in part by National Institutes of Health Grant R01 DC06104-01.

1. Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., Sanes, J. R., Welker, E. & Svoboda, K. (2002) *Nature* **420,** 788–794.
2. Grutzendler, J., Narayanan, K. & Gan, W.-B. (2002) *Nature* **420,** 812–816.
3. Brody, C. D. & Hopfield, J. J. (2003) *Neuron* **37,** 843–852.
4. Sjostrom, P. J., Turrigiano, G. G. & Nelson, S. B. (2001) *Neuron* **32,** 1149–1164.
5. Froemke, R. C. & Dan, Y. (2002) *Nature* **416,** 433–438.
6. Hopfield, J. J. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8429–8433.
7. Cover, T. M. & Thomas, J. A. (1991) *Elements of Information Theory* (Wiley, New York), pp. 18–19.
8. Bi, G. Q. & Poo, M. M. (1998) *J. Neurosci.* **18,** 10464–10472.
9. Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C. & Nelson, S. B. (1998) *Nature* **391,** 892–896.
10. Blum, K. I. & Abbott, L. F. (1996) *Neural Comput.* **8,** 85–93.
11. Markram, H., Lubke, J., Frotscher, M. & Sakmann, B. (1997) *Science* **275,** 213–215.
12. Mehta, M. R. & Wilson, M. A. (2000) *Neurocomputing* **32,** 905–911.
13. Rao, R. P. N. & Sejnowski, T. J. (2001) *Neural Comput.* **13,** 2221–2237.
14. Mehta, M. R., Lee, A. K. & Wilson, M. A. (2002) *Nature* **417,** 741–746.