nature neuroscience

The tempotron: a neuron that learns spike timing–based decisions

Robert Gütig[1–4] & Haim Sompolinsky[1,2,5]

The timing of action potentials in sensory neurons contains substantial information about the eliciting stimuli. Although the computational advantages of spike timing–based neuronal codes have long been recognized, it is unclear whether, and if so how, neurons can learn to read out such representations. We propose a new, biologically plausible supervised synaptic learning rule that enables neurons to efficiently learn a broad range of decision rules, even when information is embedded in the spatiotemporal structure of spike patterns rather than in mean firing rates. The number of categorizations of random spatiotemporal patterns that a neuron can implement is several times larger than the number of its synapses. The underlying nonlinear temporal computation allows neurons to access information beyond single-neuron statistics and to discriminate between inputs on the basis of multineuronal spike statistics. Our work demonstrates the high capacity of neural systems to learn to decode information embedded in distributed patterns of spike synchrony.

Uncovering the coding principles by which neurons in the CNS represent and process information has challenged brain researchers for over half a century. It is commonly assumed that neurons represent information through their mean rates of action potential firing. However, in visual, auditory, olfactory and somatosensory pathways[1–5], spikes elicited in response to a stimulus can be precisely timed relative to a stimulus event or to other action potentials of the same or other neurons. These data suggest that the temporal structure of spike trains serves as an important component of the neuronal representation of stimuli. For instance, in the human tactile system[5], the latencies of action potentials in individual tactile afferents carry information about the direction of the external force and the shape of the contacting surface. Spike latency codes have also been suggested for the olfactory system[6] and retinal ganglion cells[7]. The synchrony between spike trains of neuronal pairs in visual cortex has been implicated in the representation of stimulus 'gestalt' features[8]. Multineuronal synchronous events have been detected in the retina and have receptive fields different from those predicted by the firing rates of the individual neurons[1] (but see ref. 9). Indeed, some researchers have argued that, under certain circumstances, temporal neural codes offer significant computational advantages over rate codes[6,7,10–12]. Others have argued, however, that temporal codes are inferior because of the complexity of their decoding[9,13].

These experimental and computational issues highlight two important and long-standing questions. First, through what mechanisms of synaptic plasticity can neurons learn to discriminate between different spatiotemporal sequences of incoming spike patterns? That is, how do neurons learn to read out information carried by a temporal code? Second, what spatiotemporal features embedded within incoming spike patterns do neurons use to compute their response? Most existing computational models of supervised learning, such as the powerful perceptron[14,15] (**Supplementary Note** online), are formulated in a rate-based framework and do not apply to the learning of spike timing–based decision rules. This shortcoming has obstructed theoretical and experimental advances in understanding the role of spike-timing in neural information processing and learning.

To close this gap, we have devised a new, biologically plausible model of supervised learning, the tempotron, for decoding information embedded in spatiotemporal spike patterns. Using tempotron learning, we show that an integrate-and-fire neuron can learn to categorize a broad range of input classes. These include cases in which category information is not encoded in spike counts but, rather, in the latencies of single spikes or in pairwise and higher-order patterns of synchrony.

RESULTS

The tempotron learning rule

Our neuron model consists of a leaky integrate-and-fire neuron driven by exponentially decaying synaptic currents generated by its N synaptic afferents. The subthreshold membrane voltage is a weighted sum of postsynaptic potentials (PSPs) from all incoming spikes:

$$V(t) = \sum_i \omega_i \sum_{t_i} K(t - t_i) + V_{\text{rest}} \qquad (1)$$

Here t_i denotes the spike times of the ith afferent and $K(t - t_i)$ is the normalized PSP contributed by each incoming spike: that is, $K(t - t_i) = V_0 (\exp[-(t - t_i)/\tau] - \exp[-(t - t_i)/\tau_s])$. The parameters τ and τ_s denote decay time constants of membrane integration and synaptic currents, respectively (**Fig. 1**). The factor V_0 normalizes PSP kernels to 1 so that

[1]Racah Institute of Physics and [2]Interdisciplinary Center for Neural Computation, Hebrew University, 91904 Jerusalem, Israel. [3]Institute for Theoretical Biology, Humboldt University, 10115 Berlin, Germany. [4]Neuroscience Research Center, Charitè Medical Faculty of Berlin, 10117 Berlin, Germany. [5]Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA. Correspondence should be addressed to H.S. (haim@fiz.huji.ac.il) or R.G. (guetig@cc.huji.ac.il).

Received 1 December 2005; accepted 17 January 2006; published online 12 February 2006; corrected 14 February 2006 (details online); doi:10.1038/nn1643

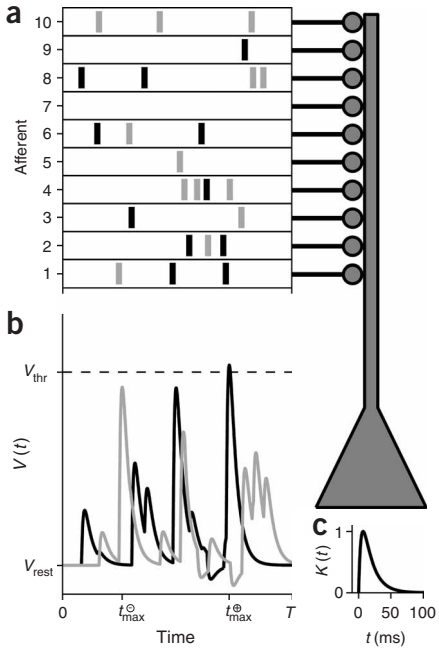

Figure 1 Tempotron classification. (**a**) Spike times (thick vertical lines) of ten afferents for two schematic input patterns, ⊕ (black) and ⊖ (gray). (**b**) Resulting postsynaptic voltage traces $V(t)$. Maximal voltages were reached at $t_{max}^{\oplus}$ and $t_{max}^{\ominus}$, respectively. Because $V(t_{max}^{\oplus}) > V_{thr} > V(t_{max}^{\ominus})$ with V_{thr} (dashed horizontal line) denoting the spike threshold, both patterns are classified correctly. Inputs arriving after a threshold crossing (black trace) are shunted. (**c**) Postsynaptic integration kernel $K(t)$ with membrane time constant $\tau = 15$ ms and synaptic time $\tau_s = \tau/4$. Here and in all other figures, the input arrival times are between 0 and $T = 500$ ms.

unitary PSP amplitudes are given by the synaptic efficacies ω_i. $K(t - t_i)$ is a causal filter, vanishing for $t_i > t$. When $V(t)$ crossed the firing threshold, the neuron is said to elicit a spike (referred to as an output spike), and the voltage undergoes a smooth reset to V_{rest} by shunting all incoming spikes that arrived after the output spike (Methods).

In the classification tasks considered here, each input pattern belongs to one of two classes (which are labeled ⊕ and ⊖). A pattern consists of N trains of input spikes, one train for each afferent, arriving between times 0 and T. The tempotron's task is to respond to a ⊕ pattern by firing at least one action potential and to remain quiescent when driven by a ⊖ pattern (**Fig. 1**). The tempotron learns this task by a rule that modifies the synaptic efficacies ω_i whenever an error occurs. In an iterative learning scheme, synapses should be depressed by an amount that reflects their contribution to an erroneous output spike on a ⊖ pattern or potentiated according to their responsibility for failure to spike on a ⊕ pattern. To resolve the temporal credit assignment problem of gauging the relative contribution of afferents to an erroneous decision, the tempotron uses the following rule: If no output spike was elicited in response to a ⊕ pattern, each synaptic efficacy ω_i is increased by the following amount:

$$\Delta\omega_i = \lambda \sum_{t_i < t_{max}} K(t_{max} - t_i) \qquad (2)$$

Here t_{max} denotes the time at which the postsynaptic potential $V(t)$ reaches its maximal value. The constant $\lambda > 0$ specifies the maximum size of the synaptic update per input spike. Conversely, if an output spike appears in response to a ⊖ pattern the synaptic efficacies are decreased by $\Delta\omega_i$ (**Fig. 2**).

The tempotron learning rule implements a 'gradient-descent' dynamics, which minimizes a cost function that measures the amount by which the maximum voltage generated by erroneous patterns deviates from the firing threshold. For each ⊕ pattern that did not elicit a spike, the cost function is $V_{thr} - V(t_{max})$. For each ⊖ pattern that induced an output spike, the cost function is $V(t_{max}) - V_{thr}$. In gradient-based learning, the change in the synaptic efficacy is proportional to the negative of the derivative of the cost function with respect to ω_i. Differentiating the value of the voltage at the peak with respect to the efficacies (Methods) gives the tempotron learning rule, shown in equation (2). Using this rule, we studied the ability of a leaky integrate-and-fire neuron to learn the classification of spatiotemporal spike patterns that mimic scenarios of spike latency (**Fig. 3a**) and spike synchrony codes (**Fig. 3b**,**c**).

Learning to classify latency patterns

In our first test of the tempotron's ability to recognize spatiotemporal spike features, we assessed its performance in classifying populations of single spike latencies. The task consisted of p spike patterns, each of which was randomly assigned to one of the two classes ⊕ or ⊖. In each pattern, each afferent fired once at a fixed time chosen randomly from a uniform distribution between 0 and $T = 500$ ms (**Fig. 3a**). The ability of the tempotron to classify such random patterns correctly depends crucially on the richness of the filtering of its spatio-temporal synaptic integration, which in turn is a function of the number of synapses. Therefore, a useful measure of the load on the system is the number of patterns p relative to the number of synapses, denoted by $\alpha = p/N$. An important characteristic of the neuron's capacity is the maximal load that it can learn. The tempotron can learn to perform the task as long as α is less than a critical value, approximately 3 (**Fig. 4a**). The similarity of the dependence of the learning time on α for different values of N (**Fig. 4a**) indicates that the maximum number of patterns that the tempotron can learn grows linearly with the number of afferents. Notably, this capacity exceeds the known capacity of $\alpha = 2$ of a single-layer perceptron[14–16] (**Supplementary Note** and **Supplementary Fig. 1** online), reflecting the higher complexity of the computation performed by the integrate-and-fire neuron across time. In the classification task studied above, the spike count was the same for all inputs in all patterns; hence, the pattern's class membership was specified entirely by the spike latencies. By contrast, an example of rate coding could be implemented through spike patterns in each of which a randomly chosen half of the afferents do not spike while the other half fire synchronously[17] (Methods). In this case, the tempotron learning rule reduces to the perceptron learning rule (see also **Fig. 4a**, left curve). Thus, notably, the tempotron can read both spike timing–based and rate-based codes.

The computation performed by the neuron depends on the PSP integration window relative to the stimulus duration. For concreteness, we varied τ and τ_s while holding the ratio between them fixed ($\tau = 4\tau_s$). For τ much shorter than the average time between input spikes T/N, no temporal summation of separated inputs takes place. In contrast, slow kinetics renders all spikes effectively simultaneous. Hence, in both limits, the extraction of temporal features is impaired and differentiation between ⊕ and ⊖ patterns on basis of spike timing becomes impossible. Correspondingly, the tempotron solves the random latency classification task within an intermediate range of PSP integration times (**Fig. 4b**). This range and the optimal τ, for which learning was fastest, depend on the number of synapses N. Increasing N substantially reduces the lower limit of τ (data not shown). Additionally, whereas the optimal τ is ~10 ms for $N = 500$, it reduces to ~4 ms

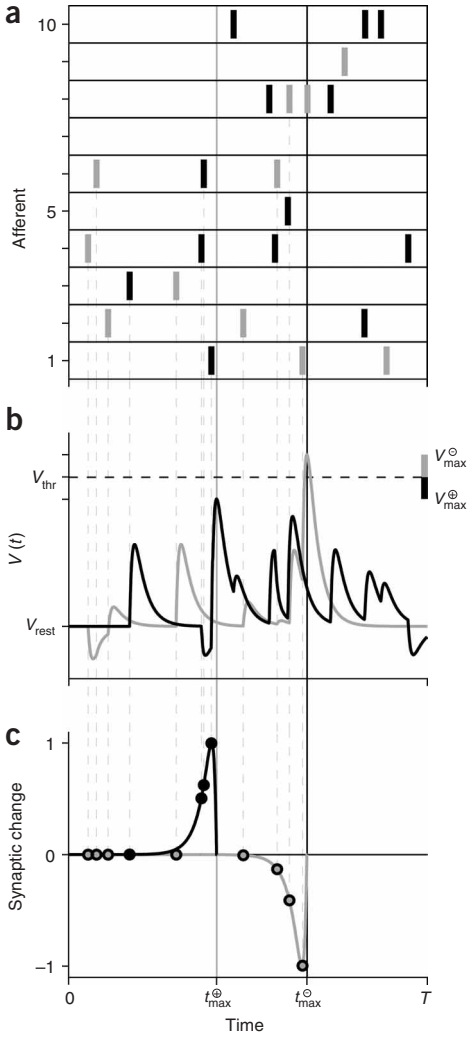

Figure 2 Tempotron learning rule. (**a**) Spike times (thick vertical lines) of ten afferents in two schematic input patterns, ⊕ (black) and ⊖ (gray). (**b**) Resulting postsynaptic voltage traces $V(t)$. Maximal voltages are reached at $t_{max}^{\oplus}$ and $t_{max}^{\ominus}$ (thin solid vertical lines), respectively. Because $V(t_{max}^{\oplus}) < V_{thr} < V(t_{max}^{\ominus})$ with V_{thr} (dashed horizontal line) denoting the spike threshold, both patterns generate an error. Inputs arriving after a threshold crossing (gray trace) are shunted. Black and gray thick vertical lines indicate the cost terms $V_{max} - V_{thr}$ and $V_{thr} - V_{max}$ associated with the ⊕ and ⊖ patterns, respectively. (**c**) Resulting synaptic changes (in units of maximal change, λ) depend on presynaptic spike times (circles) relative to the corresponding voltage maximum. Thin dashed vertical lines in (**a**–**c**) mark presynaptic spike times.

for $N = 1{,}000$ (see also **Fig. 4a**). The range of permissible values of τ shrinks with increasing load α (**Fig. 4b**); however, for intermediate load, the range of integration time constants for which the tempotron solves the task is broad and covers sparse and dense modes of decision making. These modes are distinguished by the size of the input populations that contributed to a given decision to fire, N_{dec} (**Fig. 4c**, Methods). The mean of N_{dec} was 4.7 for $\tau = 2$ ms and 97 for $\tau = 75$ ms, confirming that the tempotron can operate in both sparse and dense regimes. The two cases differ also in the distributions of learned synaptic efficacies. Decisions based on integration of a small number of inputs require a substantial fraction of strong synaptic efficacies (**Fig. 4d**).

A characteristic feature of the tempotron is that the time of the output spike is not constrained by the classification task, allowing the neuron to distribute its decision times to optimize its performance. In the present task, the output spike times are broadly distributed: early output spikes appear within τ of the first input spikes and late output spikes appear at the end of the input patterns (**Fig. 5a**). This temporal homogeneity is also reflected in the mean subthreshold voltage traces, averaged separately over all ⊕ and ⊖ patterns after learning converged to zero error. After an onset ramp of duration τ, the mean voltage traces maintain a steady plateau value across the entire input time interval (**Fig. 5b**).

To assess the robustness of tempotron learning in the presence of temporal noise, we trained the neuron with jittered versions of p templates of latency patterns. In each presentation, a template was jittered by adding a Gaussian noise to all spike times. The tempotron's generalization error—that is, the probability of misclassifying a jittered spike pattern after the learning converged—grows with increasing jitter and higher learning load α (**Fig. 6a**). For a broad range of postsynaptic integration times τ, from roughly 2 ms to 75 ms, the tempotron can tolerate a jitter of roughly 1% of τ for $\alpha = 2$ and 10% for $\alpha = 1$ (data shown for $\tau = 15$ ms). Hence, with PSP kinetics on the order of tens of milliseconds and spike precisions in the millisecond range[5,11], the categorization of 1–2 templates per synapse is feasible. Notably, when discriminating between only two templates, the tempotron can overcome jitter of roughly 10τ. This demonstrates that for small load, the tempotron is able to learn temporal features that extend well beyond the timescale of its intrinsic dynamics. Another feature of the tempotron's robustness is the relative stability of the timing of output spikes to the temporal jitter of a ⊕ template. A small jitter in input spikes generates a distribution of output spikes for each template that has a prominent narrow peak around the unperturbed spike time. Input jitter causes a large change in the output spikes in only a few percent of the trials (**Fig. 6b**,**c**). The tempotron is also robust to the deletion or insertion of a small number of spikes. For $\alpha = 1$, its generalization error roughly equals the percentage of the perturbed spikes (data not shown).

Learning to classify patterns of synchrony

What information about the timing of spikes in ⊕ and ⊖ patterns does the tempotron imprint on its learned synaptic efficacies? Differences between the voltage traces averaged separately over the ⊕ and ⊖ random latency patterns are small relative to the fluctuations in individual traces (**Fig. 5b**), implying that the tempotron also relies on higher-order spike time statistics to separate the two input classes. To directly test the sensitivity of the tempotron to multineural firing statistics, we trained it to discriminate between classes of input patterns that cannot be distinguished on the basis of single neuron spike statistics. Specifically, we examined the simple scenario in which each input pattern consists of pairwise-synchronous events. All N afferents are grouped into $N/2$ pairs. Afferents in a given pair fire synchronously, each a single spike. Notably, this grouping is fixed for each of the two categories (⊕ and ⊖) but differs between them (**Figs. 3b** and **7a**; Methods). Synchronous events occur at random, uniformly distributed times in both pattern categories, so that class information is embedded solely in the patterns of synchrony; neither spike counts nor spike timing of individual neurons carry any information relevant for the classification task. This task mimicked spike synchrony–based sensory processing[4,8,18–20]. When trained on these patterns with a jitter of up to 2–3 ms, the tempotron quickly achieved practically zero generalization error. Learning splits the synapses into two groups, roughly equal in size, of excitatory ($\omega > 0$) and inhibitory ($\omega < 0$) efficacies (**Fig. 7b**).

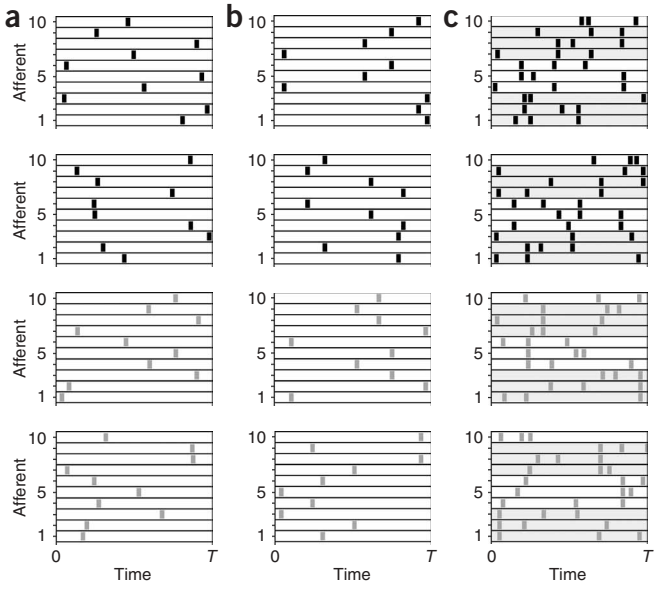

Figure 3 Input patterns (see Methods for details). Spike times (thick vertical lines) of ten afferents in input patterns representative of the three classification tasks considered in this work. The two upper rows depict two examples of ⊕ patterns (black rasters) and the two lower rows show two ⊖ patterns (gray rasters). (**a**) Random latencies. Each pattern consists of one spike per afferent, the time of which is drawn from a uniform distribution. In this classification task, class membership is assigned at random to the latency patterns. (**b**) Pairwise synchrony. Afferents in each input class, ⊕ and ⊖, are divided into $N/2$ synchronous pairs: for example, (1, 3) and (2, 10) in ⊕ patterns and (1, 6) and (2, 7) in ⊖ patterns. The identity of the synchronous pairs is conserved across all ⊕ and across all ⊖ patterns. In each realization of inputs, the arrival times of synchronous spikes are drawn from a uniform distribution. (**c**) Third-order interaction. In both input classes afferents are divided into the same groups: for example, (1, 2, 3), (4, 5, 6), (7, 8, 9), (alternating background shading). Each afferent fires three times during the input interval. In ⊕ patterns, each afferent participates in two synchronous spike pairs with its interacting neighbors and contributes one additional asynchronous spike. In contrast, in ⊖ patterns, each afferent participates in one synchronous spike triplet and contributes two additional single asynchronous spikes. In each realization, the times of all single, pair and triple events are random.

Pairs of synapses that are synchronized in ⊕ patterns have the same sign whereas pairs coincident in the ⊖ patterns have opposite signs. As a result, coincident inputs cancel on the ⊖ patterns but add in the ⊕ patterns. The nature of this solution reveals that the tempotron does not rely on a single synchronous pair but on a large number of pairs, using this redundancy to overcome ambiguities generated by the input jitter.

The ability of the tempotron to extract spike correlations is not limited to pairwise statistics. Motivated by indications of the potential importance of higher-order multineural spike statistics[21,22], we used synchronous spike pairs versus spike triplets to generate input classes that have the same first- and second-order statistics but differ in their third-order interactions. We did this by dividing all N afferents into $N/3$ groups of three. This grouping is conserved across all patterns and both classes. Depending on a pattern's class membership, the spike times within each such group of three afferents contain either three pairwise coincident events or a single coincident spike triplet (**Fig. 3c**; Methods). The tempotron can learn to discriminate these classes either by selectively firing in response to patterns with spike triplets (data not shown) or, even when required to detect the 'missing spikes', by firing exclusively in response to patterns with only pairwise coincidences (**Fig. 7c**). In contrast to the previous task (**Fig. 7b**), here the tempotron weakens almost all synapses and, by spontaneous symmetry breaking, selects a single group of three interacting afferents with which to compute its decision (**Fig. 7d**). Two of these weights are positive and one is negative. The magnitudes of the positive weights suffice to elicit an output spike when activated in synchrony in a ⊕ pattern but cannot overcome the negative contribution of the third coincident spike in the triplet in the ⊖ patterns. The strategy of computing the decision using a single group of interacting afferents minimizes the probability of errors due to random spike coincidences from other triplets. The ability to separate patterns according to higher-order spike time statistics reflects the underlying nonlinear computation across time performed by the tempotron, which goes beyond extensions of perceptron learning to the time domain[23].

Implementation by voltage convolution

One challenge for the biological implementation of our model is that the tempotron's solution to the temporal credit assignment problem requires computation of the time of the maximum total PSP. Notably,

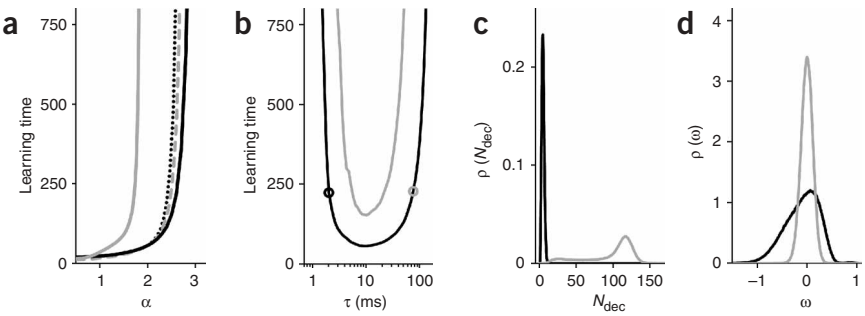

Figure 4 Tempotron performance. (**a**) Learning time—that is, mean number of cycles of pattern presentations required for error-free classification—versus load α. The load is defined as the ratio of the number of patterns over the number of afferents. Results for random latency patterns with $N = 500$ afferents (black solid line) were obtained with a synaptic integration time of $\tau = 10$ ms. For $N = 1{,}000$ afferents (gray dashed line) and $N = 1{,}500$ afferents (black dotted line) τ was set to 4 ms and 3 ms, respectively. These values of τ are close to the respective optimal τ. For comparison, the learning times for $N = 1{,}000$ and $N = 1{,}500$ were scaled to match the learning time for $N = 500$ at $\alpha = 2$. The gray solid line depicts learning times for perceptron-like input patterns with $N = 500$ afferents (Methods). (**b**–**d**) In **b**, learning time for $N = 500$ afferents versus PSP time constant τ for $\alpha = 2$ (black line) and $\alpha = 2.5$ (gray line). Open circles mark $\tau = 2$ ms (black) and $\tau = 75$ ms (gray) used in **c**–**d**. In **c**, distributions of effective number of contributing synapses N_{dec} (Methods) after learning ($\alpha = 2$) for $\tau = 2$ ms (black) and $\tau = 75$ ms (gray). In **d**, distributions of learned synaptic efficacies ω (in units of the spike threshold V_{thr}) for parameters as in **c** (same colors). Data are averaged over 1,000 realizations (Methods).

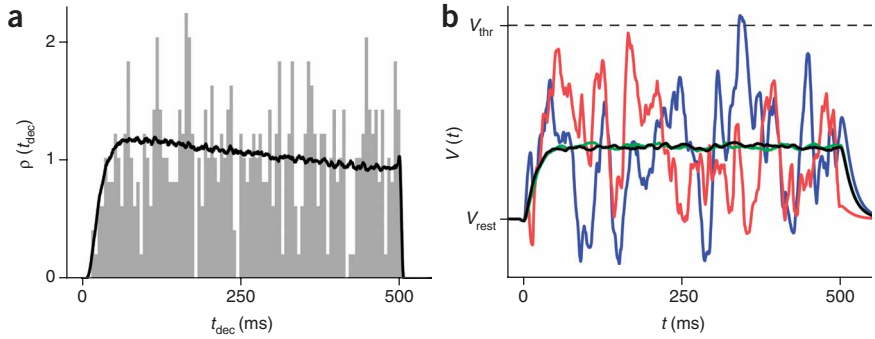

Figure 5 Tempotron decision times. (**a**) Distribution of output spike times on ⊕ patterns after learning ($\alpha = 2$, $N = 500$, $\tau = 15$ ms) for a single batch of $p = 1{,}000$ patterns (gray histogram) and average over 1,000 realizations (black line). (**b**) Typical voltage traces for individual ⊕ (blue) and ⊖ (red) patterns after learning and average voltages over all ⊕ (green) and ⊖ patterns (black) of one batch. Parameters as in **a**. The dashed horizontal line denotes the spike threshold V_{thr}. For the purpose of this comparison, we show the total PSP traces for the ⊕ patterns, ignoring the shunting effect of the output spike.

this temporally global operation can be approximated by temporally local computations that are based on the postsynaptic voltage traces following input spikes. Several functional forms that capture the temporal correlation between presynaptic spiking and postsynaptic voltage trace are possible. In one form of such a rule, the changes in synaptic efficacies following error trials are based on the convolution of the PSP kernel with the postsynaptic potential:

$$\upsilon_i = \int_{t_i}^{\infty} \mathrm{d}t\, V(t)K(t - t_i) \qquad (3)$$

The voltage maximum operation is approximated by thresholding υ_i: plasticity is induced at synapses for which υ_i is greater than a plasticity induction threshold κ. Efficacies are incremented by an amount λ for ⊕ patterns and by $-\lambda$ for ⊖ patterns (Methods). This rule ensures that synapses are changed only if the voltage within a postsynaptic integration time after their activation is sufficiently high. We found that, on all tasks considered here, this rule performs reasonably well compared to the tempotron rule. Although maximum capacity is somewhat reduced, by less than a factor of 2 (data not shown), for intermediate loads the learning time of the voltage convolution–based tempotron is only moderately increased (**Fig. 8**, inset). To investigate the relationship between this rule and the original tempotron rule, we computed the changes of all synapses that were activated at a given temporal offset from the time of maximal V, averaged separately for errors following ⊕ and ⊖ patterns. The resultant mean plasticity temporal profile (**Fig. 8**, black solid curve) is similar to that specified by the tempotron plasticity rule (**Fig. 8**, black dotted curve). This suggests that a temporally local mechanism that extracts the contiguity between the elevation of postsynaptic membrane potential and the timing of incoming spikes suffices to roughly mimic the tempotron's global synaptic credit assignment. This was also supported by our finding that voltage convolution learning converges to the same type of solution as that of tempotron learning in tasks involving multineuronal spike statistics (data not shown).

Comparison with reinforcement learning

Another approach to the credit assignment problem is to correlate the synaptic changes with neuronal noise, a method used by reinforcement learning. We implemented a synaptic reinforcement learning scheme[24] in which synaptic changes are governed by correlations between random synaptic noise, which was added to the synaptic efficacies, and a global error signal (Methods). Due to its inherent stochasticity, learning in this reinforcement scheme is dramatically slower than that in the tempotron (roughly four orders of magnitude in the example of **Fig. 8**, inset) and does not converge to zero error. We computed the mean synaptic changes induced by this learning rule as a function of the time of synaptic activation relative to the maximum $V(t)$. During the early phase of learning, the resultant temporal profile is substantially different from that of the tempotron (data not shown); in contrast, in the late phase both profiles are remarkably similar (**Fig. 8**, gray solid line) even though plasticity in the reinforcement learning scheme does not contain explicit reference to the postsynaptic voltage. This suggests that the computational principle underlying tempotron learning has a fundamental role in spike-based processing.

DISCUSSION

The tempotron model posits that synaptic changes depend on the value of the postsynaptic voltage. A crucial role of the postsynaptic voltage in the induction of long-term plasticity has been demonstrated in several

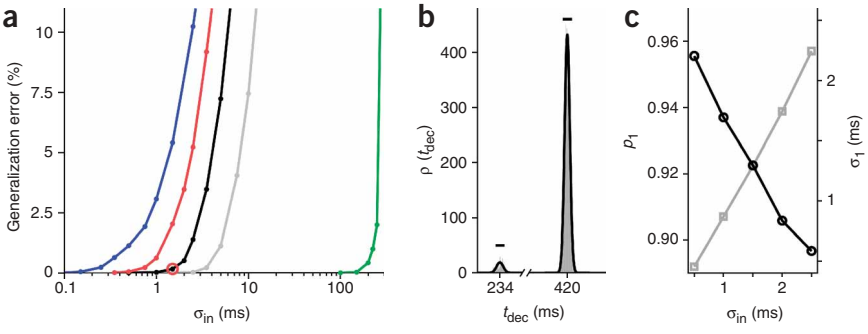

Figure 6 Tempotron robustness. (**a**,**b**) In **a**, generalization error—that is, the probability of misclassifying a jittered spike pattern after the learning converged—versus input jitter standard deviation σ_{in} for $\alpha = 2$ (blue), 1.5 (red), 1 (black), 0.5 (gray) and 1/250 (green). Results were obtained for $N = 500$ afferents and a synaptic time constant of $\tau = 15$ ms. Red open circle on $\alpha = 1$ curve (black) marks the σ_{in} used in **b**. In **b**, distribution of output times (gray histogram) originating from a single ⊕ pattern template with jitter $\sigma_{in} = 1.5$ ms after learning ($\alpha = 1$). The black line depicts the estimated density of a two-component Gaussian-mixture model (**Supplementary Methods**). Short horizontal scale bars centered above the mean of each component extend over $\pm\sigma_{in}$. (**c**) Mean weight p_1 (black circles, left y-axis) and mean standard deviation σ_1 (gray squares, right y-axis) of the dominant component of the distribution of output times versus σ_{in}. The quantities were computed by averaging the parameters of Gaussian-mixture fits across all ⊕ patterns. Data in **a** and **c** were averaged over five realizations.

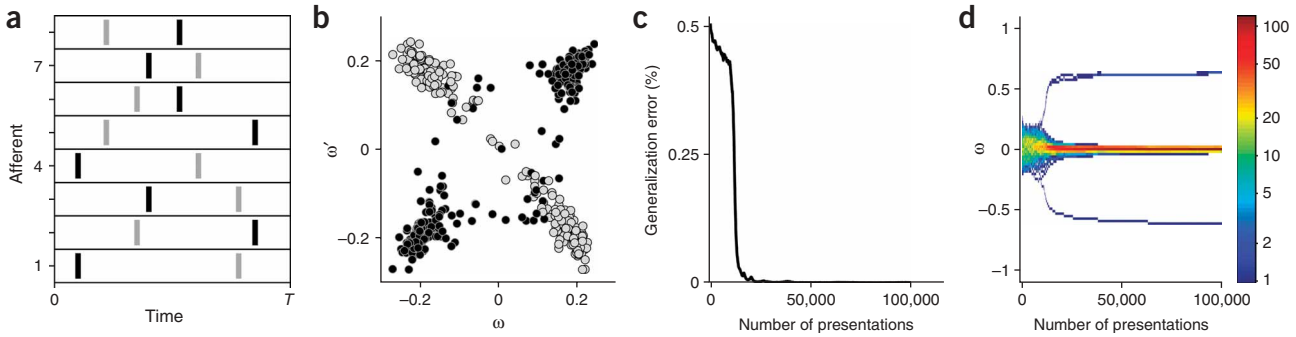

Figure 7 Learning multineural spike statistics. (**a**,**b**) Pairwise synchrony (see **Fig. 3b**). In **a**, spike times of eight afferents in a ⊕ (black) and a ⊖ (gray) pattern, realizing distinct pairings of synchronous input spikes. In **b**, scatter plot of pairs (ω, ω') of learned efficacies. Black circles show the efficacies grouped into the pairs that are active synchronously in ⊕ patterns: for example, $\omega = \omega_1$ and $\omega' = \omega_4$ (see **a**). Gray circles show the same efficacies now grouped into the pairs that are active synchronously in ⊖ patterns: for example, $\omega = \omega_7$ and $\omega' = \omega_4$ (see **a**). ($N = 500$, $\tau = 15$ ms). (**c**,**d**) Third-order interactions (**Fig. 3c**). In **c**, generalization error during learning of discrimination between patterns with different third-order interactions as a function of number of pattern presentations ($N = 168$, $\tau = 15$ ms). In **d**, time evolution of the histogram of synaptic efficacies, starting from a Gaussian distribution. Color-coded histograms of synaptic efficacies (logarithmic color scale, right bar) versus number of pattern presentations. Learning suppressed all but the three weights of a single group of interacting afferents that developed large magnitudes. Two of these were positive efficacies (upper branch) and the other was negative (lower branch). Sharp drop in generalization error in **c** accompanies the emergence of this solution.

preparations[25–28]. In addition, local voltage computations similar to those discussed above can be implemented by the thresholding properties of voltage-gated calcium channels and NMDA receptors, whose involvement in the induction of long-term plasticity is well established[25,29,30]. However, the voltage in the tempotron rule is the postsynaptic potential, which in the case of an erroneous output spike (on a ⊖ pattern) does not incorporate the voltage produced by the active conductances. The tempotron incorporates the effect of opening the active conductances by erasing all the synaptic inputs that arrive after an output spike.

Most experimentally derived plasticity rules, such as spike timing–dependent plasticity (STDP), characterize synaptic changes solely in terms of the temporal contiguity of the presynaptic spike and the postsynaptic potential or spiking. Indeed, several computational studies have shown that such 'unsupervised' rules can extract salient statistical features in the input spike trains[31–36]. In contrast, the tempotron solves the more challenging problem of learning to respond to presynaptic spike sequences by making the appropriate decision. Such learning necessarily requires supervisory signals, which instruct the neuron about the correct decision. An important question concerning the biological feasibility of tempotron learning is through which pathways supervisory signals arrive at the site of plasticity and how they are translated into the appropriate synaptic changes.

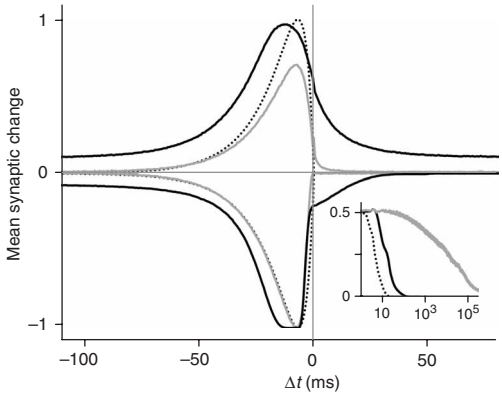

Figure 8 Spike timing dependence of different learning rules. Mean synaptic changes versus the time difference Δt between presynaptic spikes and postsynaptic voltage maximum, for voltage convolution–based (solid black), reinforcement (solid gray) and tempotron learning (dotted black, equation (2)). Minima are scaled to −1. Positive (negative) curves correspond to mean synaptic changes following errors on ⊕ (⊖) patterns. Data were collected during the late phase of learning, after the fraction of misclassified patterns dropped below 0.25. The learning task consisted of random latency patterns (**Fig. 3a**) with $N = 500$ afferents and a load of $\alpha = 1$ ($\tau = 15$ ms). The inset (same color code and parameter values) shows the fraction of misclassified patterns versus the number of cycles of pattern presentations during learning. All data were averaged over 1,000 realizations (solid and dotted black) and 10 realizations (gray).

In the tempotron, the magnitudes of synaptic changes are fully determined by the time course of the postsynaptic voltage. The polarity of synaptic changes is the same for all synapses and only depends on the desired response. Hence, the role of the supervisory signal is to determine this global polarity. One way to implement the control of polarity is through classic error feedback. In such a scenario, an error trial is followed either by a signal that results in long-term potentiation (LTP) if the neuron failed to spike on a ⊕ pattern or by a signal resulting in long-term depression (LTD) if the neuron erroneously fired on a ⊖ pattern. The minimal requirements for such a scheme are twofold. First, synapses must be able to maintain voltage-dependent 'eligibility' traces over substantial periods of time, such that the subsequent delivery of the error feedback can be translated into the correct amount of change. Second, there must be neuromodulatory signaling pathways that can be recruited to activate either LTP or LTD after an error trial, depending on the type of error. One problematic aspect of this feedback scenario is that it relies on the calculation and transmission of the error by the supervisory system, which may considerably delay the execution of the appropriate synaptic modification. This sequence of calculations can be simplified by realizing that the tempotron requires only the availability of a signal that instructs the desired response and not an explicit computation of error. We can build on this feature to construct a scheme in which a synaptic depression process is induced independently of the desired response whenever the tempotron fires an output spike. In contrast, potentiation does not depend on postsynaptic spiking but is induced only in the presence of a neuromodulator released by the supervisory system. This activation of the LTP process is the point of interaction through which

the tempotron is instructed that the desired response is to spike. Neuromodulatory control of LTP occurs for several neuromodulatory pathways, such as dopamine and acetylcholine[37–39]. Finally, it is necessary that coactivation of both processes results in cancellation of potentiation and depression, such that no changes in synaptic efficacies are expressed if the neuron spikes correctly in response to a ⊕ pattern. On a ⊖ pattern, no plasticity is induced when the neuron remains quiescent, as neither LTP nor LTD is elicited. However, the failure to spike on ⊕ patterns results in potentiation whereas spiking on ⊖ patterns results in depression, producing the correct polarity of synaptic changes on erroneous ⊕ and ⊖ trials.

Recent reports provide support for the long-standing hypothesis[40,41] that the polarity of expressed synaptic change results from the balance between separately activated intracellular induction pathways for LTP and LTD (refs. 42,43), both of which converge onto the trafficking of postsynaptic AMPA receptors[44–46]. Due to its intricate dependence on a large number of intracellular state and signaling variables, this balance between a depression process modulated by calcineurin[41] and a potentiation process mediated by Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII; ref. 47) offers a large number of possible interaction sites for both neuromodulatory and state-dependent regulations[37,41,48]. At present, these intracellular signaling cascades are not yet fully exposed, precluding the determination of the feasibility of any detailed scenario of balancing between LTP and LTD processes. Independent of the details of how tempotron learning is implemented, our work predicts that, unlike simple STDP rules, some synaptic plasticity in behaving animals is governed by two factors: the temporal contiguity of a presynaptic spike and a postsynaptic depolarization, and a global neuromodulatory signal that determines the polarity of the expressed long-term synaptic changes.

Our results highlight the ability of simple neuronal circuits to learn and implement spike timing–based computation. This supports the robust decoding of a broad range of spatiotemporal neural codes, in contrast to earlier assertions[9,13]. A natural question is what are the computational limitations of the tempotron. In the tempotron, output spikes are elicited in response to spatiotemporal input features whose duration is on the order of the postsynaptic integration window (τ in our notation). As shown here, the tempotron can learn to detect the presence of such temporally restricted features anywhere in a stimulus time interval (T in our notation), even when this interval is significantly larger than τ. However, the tempotron is expected to perform poorly on classification tasks that require it to discriminate between spike patterns on the basis of temporal features that extend beyond a single integration time[49]. For instance, this limitation restricts the tempotron's ability to learn the conjunction of temporally local features separated by long times. Solving such demanding tasks requires the addition of working memory mechanisms or slow synaptic dynamics[50] and possibly also multilayer architectures.

METHODS

Neuron model. Numerical simulations of the tempotron were based on exact integration of the voltage dynamics of a leaky integrate-and-fire neuron driven by exponentially decaying synaptic currents (see equation (1)). For concreteness, we held the ratio between the membrane and the synaptic time constants (τ and τ_s, respectively) fixed at $\tau/\tau_s = 4$. With this ratio, the normalization coefficient V_0 (see equation (1)) equaled 2.12. The important temporal parameter is the ratio τ/T. We chose $T = 500$ ms and $\tau = 15$ ms (except in **Fig. 4**). PSP amplitudes of individual afferents were given by the synaptic efficacies ω_i. In this model, the sign of the synaptic efficacy could change by learning. This can be realized by altering the balance between excitatory and inhibitory pathways. We checked that the tempotron learning rule also worked (although with a reduced capacity) when the signs of the synaptic efficacies were not allowed to change. The neuron was simulated for up to 10,000 plastic synapses; results shown are for $N = 500$ synapses, except for capacity measurements which are also shown for $N = 1{,}000$ and $N = 1{,}500$ synapses (**Fig. 4a**) and for the learning of higher-order spike statistics where three spikes arrived per afferent and $N = 168$ (**Fig. 7c,d**).

An output spike was elicited when $V(t)$ crossed the firing threshold V_{thr}. After a spike at t_{spike}, the voltage was reset smoothly to the resting value by shunting all synaptic inputs that arrived after the occurrence of the spike—namely, by restricting the sum in equation (1) to $t_i < t_{spike}$ (see Tempotron Learning, below). We used $V_{thr} = 1$ and $V_{rest} = 0$.

Input patterns. As stated above, for all patterns, the duration of the input time interval [0, T] was set to $T = 500$ ms. In random patterns (**Figs. 3a**, **4–6** and **8**), each afferent spiked once at a time drawn independently from a uniform distribution. Perceptron-like patterns (**Fig. 4a**, gray curve) consisted of a random half of afferents spiking in synchrony while the other half remained silent. Patterns were assigned to input classes ⊕ or ⊖ with a probability of one-half. Jittered realizations of patterns (**Fig. 6**) were generated from the original patterns by adding independent Gaussian noise with mean = 0 and standard deviation σ_{in} to all spike arrival times while keeping the class identity of the original pattern. In learning patterns of pairwise synchrony (**Fig. 7a,b**), afferents were randomly grouped into $N/2$ pairs, the identities of which were different for each of the two classes ⊕ and ⊖ (**Fig. 3b**). In each pattern, each pair of afferents contributed one synchronous spike pair at a time drawn randomly from a uniform distribution. In learning patterns with higher-order spike correlations (**Fig. 7c,d**), afferents were assigned into groups of three, with identical and fixed identities for both pattern classes (**Fig. 3c**). Each afferent spiked three times during the input interval. In ⊕ patterns, two of the spikes of each group member formed synchronous spike pairs with its two neighbors. In ⊖ patterns, one spike of each group member engaged in a triplet of synchronized spikes. In both classes, the remaining spikes did not interact. To prevent overlap between events within each interacting group of three afferents, spike times were drawn from a uniform distribution but separated by a distance of $\tau + \tau_s$.

Tempotron learning. Synapses were modified after each error trial. We defined the single pattern cost functions $E_{\pm} = \pm (V_{thr} - V(t_{max}))\ \Theta(\pm (V_{thr} - V(t_{max}))$ for ⊕ and ⊖ patterns, respectively. Here t_{max} denotes the time of the maximum postsynaptic potential, and the Heaviside step function, defined by $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$, ensures zero cost on successful trials. In gradient-based learning, changes in the synaptic efficacies are given by $\Delta\omega_i \propto -\mathrm{d}E_{\pm}/\mathrm{d}\omega_i$. Using equation (1) (Results), we found that $\mathrm{d}E_{\pm}/\mathrm{d}\omega_i = 0$ for patterns that did not result in erroneous decisions, whereas for erroneous patterns,

$$-\frac{\mathrm{d}E_{\pm}}{\mathrm{d}\omega_i} = \pm \sum_{t_i < t_{max}} K(\Delta t_i) \pm \frac{\partial V(t_{max})}{\partial t_{max}} \frac{\mathrm{d}t_{max}}{\mathrm{d}\omega_i} \qquad (4)$$

The ± sign refers to potentiation and depression following errors on ⊕ and ⊖ patterns, respectively; the sum includes all presynaptic spikes arriving at synapse ω_i before t_{max}, and $\Delta t_i = t_{max} - t_i$ denotes the time difference between the activation of a synapse and the voltage maximum. Because $\partial V(t_{max})/\partial t_{max} = 0$ by definition of t_{max}, only the first term in equation (4) contributes to the synaptic update, yielding the learning rule shown in equation (2). This also holds for ⊖ patterns because of the smooth reset of V following an output spike. We used $\lambda = 10^{-4}/V_0$ for the maximum synaptic change, except for results of the tempotron's performance for different values of τ and N (**Fig. 4**) where we used $\lambda = 3 \times 10^{-3}\ T/(\tau N V_0)$.

Voltage convolution–based learning. In voltage convolution–based learning, changes in synaptic efficacies following error trials were governed by the convolution of the PSP kernel with the postsynaptic potential (equation (3)). Other functional forms that capture the temporal correlation between presynaptic spiking and voltage are possible. The voltage maximum operation was approximated by thresholding υ_i, yielding

$$\Delta\omega_i^{\pm} = \begin{cases} \pm\lambda & \upsilon_i > \kappa \\ 0 & \upsilon_i \leq \kappa \end{cases}$$

for synaptic changes on $\oplus$ and $\ominus$ patterns. We used a maximal synaptic change of $\lambda = 8 \times 10^{-5}$ and a plasticity induction threshold of $\kappa = 10^{-3}$. When the system was near its capacity limit, performance could be improved by increasing all weights in which $\upsilon_i \leq \kappa$ by a small amount ($\Delta\omega_i^+ = \varepsilon\lambda$, $\varepsilon = 0.01$) after each erroneous $\oplus$ pattern trial.

Reinforcement learning. We used a method of reinforcement learning that is based on stochastic synapses (ref. 24). This was implemented by generating, in each trial, independent Gaussian noise ξ_i with mean = 0 and standard deviation = 0.01 and replacing all ω_i in equation (1) with $\omega_i + \xi_i$. Following each error trial, all efficacies were corrected by $\Delta\omega_i = -\lambda\xi_i$.

Momentum term. To accelerate learning, we used a momentum heuristic in all learning rules[15]. In this scheme, synaptic updates $\Delta\omega_i^{\text{current}}$ consist not only of the correction, which is given by the learning rule $\Delta\omega_i$, but also incorporate a fraction of the previous synaptic change $\Delta\omega_i^{\text{previous}}$, such that $\Delta\omega_i^{\text{current}} = \Delta\omega_i + \mu\,\Delta\omega_i^{\text{previous}}$. The parameter μ, which we set to 0.99 throughout the present study, implements a decaying trace of former synaptic changes. As a result, the maximal step size λ was adaptively scaled with a factor ranging from $1/(1-\mu) \approx 100$ if $\Delta\omega_i$ was constant: that is, the learning kept modifying in the same direction, to 1 if the direction of learning oscillated.

Number of contributing synapses. To quantify the number of synapses that contribute to a spike decision at time t_{dec} in response to an input spike pattern, we defined the relative contribution of each synapse, γ_i, as the magnitude of the synaptic efficacy scaled by the PSP kernel, $\gamma_i = |\omega_i|\,K(t_{\text{dec}} - t_i)$. The effective number of synapses contributing to a spike decision was then measured as

$$N_{\text{dec}} = \frac{\left(\sum_{i=1}^{N} \gamma_i\right)^2}{\sum_{i=1}^{N} \gamma_i^2}$$

Note that if M afferents contributed $\gamma_i = \gamma$ and the contribution of the rest was zero, $N_{\text{dec}} = M$ as expected.

Data statistics, parameter choices and initial conditions. Averages over independent realizations were obtained by repeating simulations with identical system parameters but different random number generator seeds (**Figs. 4–6** and **8**). As a result, both input spike patterns and initial synaptic efficacies were different for each realization. When we studied the tempotron near its critical capacity (**Fig. 4**), a small fraction of runs did not converge to zero error within the maximum runtime of our simulations, as expected for any finite N near critical capacity. All data shown complied with the criterion that the probability of misclassifying a pattern after learning was below 10^{-5}.

Additional information on parameter choices, specifications of initial conditions of numerical simulations and Gaussian mixture estimation is presented in **Supplementary Methods** online.

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

ACKNOWLEDGMENTS

We thank L. de Hoz, A. Globerson, M. Gutnick, D. Hansel, O. White and Y. Yarom for comments. We acknowledge computational resources provided by the Harvard University Bauer Center for Genomic Research. This work was supported in part by the German Research Foundation, the Minerva Foundation, the European Commission's Improving Human Potential Program, the Israel Science Foundation (Center of Excellence no. 8006/00) and the Defense Research & Development Directorate (MAFAT).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

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

1. Meister, M., Lagnado, L. & Baylor, D.A. Concerted signaling by retinal ganglion cells. *Science* **270**, 1207–1210 (1995).
2. deCharms, R.C. & Merzenich, M.M. Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* **381**, 610–613 (1996).
3. Neuenschwander, S. & Singer, W. Long-range synchronization of oscillatory light responses in the cat retina and lateral geniculate nucleus. *Nature* **379**, 728–732 (1996).
4. Wehr, M. & Laurent, G. Odour encoding by temporal sequences of firing in oscillating neural assemblies. *Nature* **384**, 162–166 (1996).
5. Johansson, R.S. & Birznieks, I. First spikes in ensembles of human tactile afferents code complex spatial fingertip events. *Nat. Neurosci.* **7**, 170–177 (2004).
6. Hopfield, J.J. Pattern recognition computation using action potential timing for stimulus representation. *Nature* **376**, 33–36 (1995).
7. Thorpe, S., Delorme, A. & Van Rullen, R. Spike-based strategies for rapid processing. *Neural Netw.* **14**, 715–725 (2001).
8. Castelo-Branco, M., Goebel, R., Neuenschwander, S. & Singer, W. Neural synchrony correlates with surface segregation rules. *Nature* **405**, 685–689 (2000).
9. Nirenberg, S., Carcieri, S., Jacobs, A. & Latham, P. Retinal ganglion cells act largely as independent encoders. *Nature* **411**, 698–701 (2001).
10. Perkel, D.H. & Bullock, T.H. Neural coding. *Neurosci. Res. Program Bull.* **6**, 221–348 (1968).
11. Singer, W. Time as coding space? *Curr. Opin. Neurobiol.* **9**, 189–194 (1999).
12. Meister, M. & Berry, M. II. The neural code of the retina. *Neuron* **22**, 435–450 (1999).
13. Shadlen, M. & Movshon, J. Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* **24**, 67–77 (1999).
14. Minsky, M.L. & Papert, S.A. *Perceptrons* (MIT Press, Cambridge, Massachusetts, 1969).
15. Hertz, J., Krogh, A. & Palmer, R.G. *Introduction to the Theory of Neural Computation* (Westview Press, Boulder, Colorado, 1991).
16. Gardner, E. The space of interactions in neural network models. *J. Phys. A* **21**, 257–270 (1988).
17. Brunel, N., Hakim, V., Isope, P., Nadal, J.P. & Barbour, B. Optimal information storage and the distribution of synaptic weights: perceptron versus purkinje cell. *Neuron* **43**, 745–757 (2004).
18. Vaadia, E. *et al.* Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* **373**, 515–518 (1995).
19. MacLeod, K., Backer, A. & Laurent, G. Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature* **395**, 693–698 (1998).
20. Brody, C. & Hopfield, J.J. Simple networks for spike-timing-based computation, with application to olfactory processing. *Neuron* **37**, 843–852 (2003).
21. Schnitzer, M.J. & Meister, M. Multineuronal firing patterns in the signal from eye to brain. *Neuron* **37**, 499–511 (2003).
22. Kuhn, A., Aertsen, A. & Rotter, S. Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Comput.* **15**, 67–101 (2003).
23. Bressloff, P.B. & Taylor, J.G. Perceptron-like learning in time-summating neural networks. *J. Phys. A* **25**, 4373–4388 (1992).
24. Seung, H.S. Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron* **40**, 1063–1073 (2003).
25. Artola, A., Brocher, S. & Singer, W. Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature* **347**, 69–72 (1990).
26. Ngezahayo, A., Schachner, M. & Artola, A. Synaptic activity modulates the induction of bidirectional synaptic changes in adult mouse hippocampus. *J. Neurosci.* **20**, 2451–2458 (2000).
27. Sjostrom, P., Turrigiano, G. & Nelson, S. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* **32**, 1149–1164 (2001).
28. Lisman, J. & Spruston, N. Postsynaptic depolarization requirements for LTP and LTD: a critique of spike timing-dependent plasticity. *Nat. Neurosci.* **8**, 839–841 (2005).
29. Cummings, J.A., Mulkey, R.M., Nicoll, R.A. & Malenka, R.C. Ca^{2+} signaling requirements for long-term depression in the hippocampus. *Neuron* **16**, 825–833 (1996).
30. Malenka, R.C. & Nicoll, R.A. Long-term potentiation–a decade of progress? *Science* **285**, 1870–1874 (1999).
31. Kempter, R., Gerstner, W. & van Hemmen, J.L. Hebbian learning and spiking neurons. *Phys. Rev. E* **59**, 4498–4515 (1999).
32. Song, S., Miller, K.D. & Abbott, L.F. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* **3**, 919–926 (2000).
33. Gütig, R., Aharonov, R., Rotter, S. & Sompolinsky, H. Learning input correlations through non-linear temporally asymmetric Hebbian plasticity. *J. Neurosci.* **23**, 3697–3714 (2003).
34. Guyonneau, R., VanRullen, R. & Thorpe, S.J. Neurons tune to the earliest spikes through STDP. *Neural Comput.* **17**, 859–879 (2005).
35. Legenstein, R., Naeger, C. & Maass, W. What can a neuron learn with spike-timing-dependent plasticity? *Neural Comput.* **17**, 2337–2382 (2005).
36. Worgotter, F. & Porr, B. Temporal sequence learning, prediction, and control: a review of different models and their relation to biological mechanisms. *Neural Comput.* **17**, 245–319 (2005).
37. Foehring, R.C. & Lorenzon, N.M. Neuromodulation, development and synaptic plasticity. *Can. J. Exp. Psychol.* **53**, 45–61 (1999).
38. Seamans, J. & Yang, C. The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog. Neurobiol.* **74**, 1–58 (2004).
39. Wespatat, V., Tennigkeit, F. & Singer, W. Phase sensitivity of synaptic modifications in oscillating cells of rat visual cortex. *J. Neurosci.* **24**, 9067–9075 (2004).
40. Lisman, J. A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. *Proc. Natl. Acad. Sci. USA* **86**, 9574–9578 (1989).

41. Mansuy, I. Calcineurin in memory and bidirectional plasticity. *Biochem. Biophys. Res. Commun.* **311**, 1195–1208 (2003).
42. Liu, L. *et al.* Role of NMDA receptor subtypes in governing the direction of hippocampal synaptic plasticity. *Science* **304**, 1021–1024 (2004).
43. Wang, H.X., Gerkin, R.C., Nauen, D.W. & Bi, G.Q. Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nat. Neurosci.* **8**, 187–193 (2005).
44. Malinow, R. & Malenka, R. AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.* **25**, 103–126 (2002).
45. Zhu, J., Qin, Y., Zhao, M., Van Aelst, L. & Malinow, R. Ras and rap control AMPA receptor trafficking during synaptic plasticity. *Cell* **110**, 443–455 (2002).
46. Tomita, S., Stein, V., Stocker, T., Nicoll, R. & Bredt, D. Bidirectional synaptic plasticity regulated by phosphorylation of stargazin-like tarps. *Neuron* **45**, 269–277 (2005).
47. Lisman, J., Schulman, H. & Cline, H. The molecular basis of CaMKII function in synaptic and behavioural memory. *Nat. Rev. Neurosci.* **3**, 175–190 (2002).
48. Centonze, D., Gubellini, P., Pisani, A., Bernardi, G. & Calabresi, P. Dopamine, acetylcholine and nitric oxide systems interact to induce corticostriatal synaptic plasticity. *Rev. Neurosci.* **14**, 207–216 (2003).
49. Mauk, M. & Buonomano, D. The neural basis of temporal processing. *Annu. Rev. Neurosci.* **27**, 307–340 (2004).
50. Buonomano, D. Decoding temporal information: a model based on short-term synaptic plasticity. *J. Neurosci.* **20**, 1129–1141 (2000).