Recognition by Variance: Learning Rules for Spatiotemporal Patterns

Omri Barak
omri.barak@weizmann.ac.il
Misha Tsodyks
misha@weizmann.ac.il
Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel

Recognizing specific spatiotemporal patterns of activity, which take place at timescales much larger than the synaptic transmission and membrane time constants, is a demand from the nervous system exemplified, for instance, by auditory processing. We consider the total synaptic input that a single readout neuron receives on presentation of spatiotemporal spiking input patterns. Relying on the monotonic relation between the mean and the variance of a neuron's input current and its spiking output, we derive learning rules that increase the variance of the input current evoked by learned patterns relative to that obtained from random background patterns. We demonstrate that the model can successfully recognize a large number of patterns and exhibits a slow deterioration in performance with increasing number of learned patterns. In addition, robustness to time warping of the input patterns is revealed to be an emergent property of the model. Using a leaky integrate-and-fire realization of the readout neuron, we demonstrate that the above results also apply when considering spiking output.

1 Introduction ─────────────────────────────────

Recognizing the spoken word *recognize* presents no particular problem for most people, and yet the underlying computation our nervous system has to perform is far from trivial. The stimulus presented to the auditory system can be viewed as a pattern of energy changes on different frequency bands over a period of several hundred milliseconds. This timescale should be contrasted with the membrane and synaptic transmission time constants of the individual neurons, which are at least an order of magnitude smaller. The auditory modality is not the only one concerned with the recognition of spatiotemporal patterns of activity; reading braille and understanding sign language are among the exemplars of this rule. Songbirds provide an experimentally accessible model system to study this ability, where Margoliash and Konishi (1985) showed that specific learned (bird's own) songs elicit a higher neuronal response than nonlearned songs with similar

statistics (same dialect). The representation of external stimuli in the brain is a spatiotemporal pattern of spikes (Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997). Experiments on many different systems demonstrated a high-precision spiking response to dynamic stimuli, thereby presenting a spatiotemporal pattern of spikes to the next level in the hierarchy (Aertsen, Smolders, & Johannesma, 1979; Berry, Warland, & Meister, 1997; Uzzell & Chichilnisky, 2004; Bair & Koch, 1996). These experiments motivated us to limit the current work to the study of spiking patterns.

The computational task at hand can be formulated in the following way: given the statistical ensemble of all possible spatiotemporal spiking patterns, a randomly chosen specific finite subset is designated as *learned* patterns, while the remaining possible patterns are termed *background* patterns. The task is to build a model that recognizes a learned pattern as a familiar one by producing a larger output when presented with it, compared to when presented with a typical background pattern. The model therefore reduces the high-dimensional input to a one-dimensional output. We emphasize that in the task that we consider in this letter, the selected group of learned patterns is to be distinguished from infinitely many other random patterns, as opposed to the task of classification, where two (or more) sets of selected patterns are to be distinguished between each other.

Several network models were proposed to address this task. Hopfield and Brody (2001) introduced a network of integrate-and-fire neurons capable of recognizing spoken words, encoded as a spatiotemporal pattern of spikes, regardless of time warp effects. Jin (2004) used a spiking neural network to implement a synfire chain that responds only to specific spatiotemporal sequences of spikes, regardless of variations in the intervals between spikes.

Although the actual biological solution probably includes a network of neurons, there is an advantage in modeling the same task with a single neuron. Single-neuron models are usually more amenable to analytic treatment, thereby facilitating the understanding of the model's mechanisms. A well-known single neuron model that performs classification is the perceptron (Minsky & Papert, 1969). It might seem that the recognition problem for spatiotemporal spike patterns can be reduced to it by a simple binning over time, in which the instantaneous spatial patterns of each time bin of each learned pattern are all considered as separate input patterns for a normal perceptron. This, however, is misleading, as spatiotemporal patterns cannot be shuffled over time without destroying their temporal information, while binned patterns contain no temporal information to start with. Bressloff and Taylor (1992) introduced a two-layer model of time-summating neurons that can be viewed as a discrete-time single-neuron model with an exponentially decaying response function for the inputs. This model solves the first problem mentioned above by adding the temporal memory of the response function. While having an architecture similar to ours, the task it performs is different: the mapping of one spatiotemporal pattern to another. Such a mapping task neither necessitates integration of

the input over time, nor does it reduce the dimensionality of the input. Hence, the analysis performed there does not apply to our case.

In this article, we consider the input current to a single readout neuron, receiving a spatiotemporal pattern of spikes through synaptic connections. Theoretical results show that the firing rate of an integrate-and-fire model neuron is a monotonic function of both the mean and the variance of its input current (Tuckwell, 1988). Rauch, La Camera, Luscher, Senn, and Fusi (2003) demonstrated experimentally the same behavior in vitro by using in vivo–like current injections to neocortical pyramidal neurons. It is important to note that the variance referred to in this work is the variance over *time*, and not the variance over *space* referred to in other works (see, e.g., Silberberg, Bethge, Markram, Pawelzik, & Tsodyks, 2004). The proposed framework, along with the above results, allows us to derive learning rules for the synaptic weights, independent of the particular response properties of the readout neuron. We then use these rules in computer simulations to illustrate and quantify the model's ability to recognize spatiotemporal patterns.

2 Methods

2.1 Description of the Model. Our model (see Figure 1A) consists of a single readout neuron receiving N afferent spike trains within a fixed time window T. Each spike contributes a synaptic current with an initial amplitude W_i depending on the synapse, and an exponential decay with a synaptic time constant τ_S. The total synaptic current is the input current to the neuron. The synaptic weights are normalized according to $\sum_{i=1}^{N} W_i^2 = 1$, and we assume $T \gg \tau_S$.

We chose input patterns where each of the N neurons fires exactly n spikes, enabling the precise temporal information to provide the information for recognition. For each learned pattern $\mu \in \{1, \ldots, p\}$, each neuron $i \in \{1, \ldots, N\}$ fires spike number $k \in \{1, \ldots, n\}$ at time $t_{i,k}^{\mu} \in [0, T]$. An input pattern, $\mathbf{t}^{\mu} = \{t_{i,k}^{\mu}\}_{i=1,k=1}^{N,n}$, is therefore defined as a fixed realization of spike times drawn from a common probability distribution $P(\mathbf{t})$, where each $t_{i,k}$ is an independent random variable distributed uniformly in $[0, T]$ (see Figure 1B).

The input current $I^{\mu}(t)$, upon presentation of pattern μ, is determined according to

$$I^{\mu}(t) = \sum_{i}^{N} W_i I_i^{\mu}(t)$$

$$\tau_S \dot{I}_i^{\mu}(t) = -I_i^{\mu}(t) + \xi_i^{\mu}(t)$$

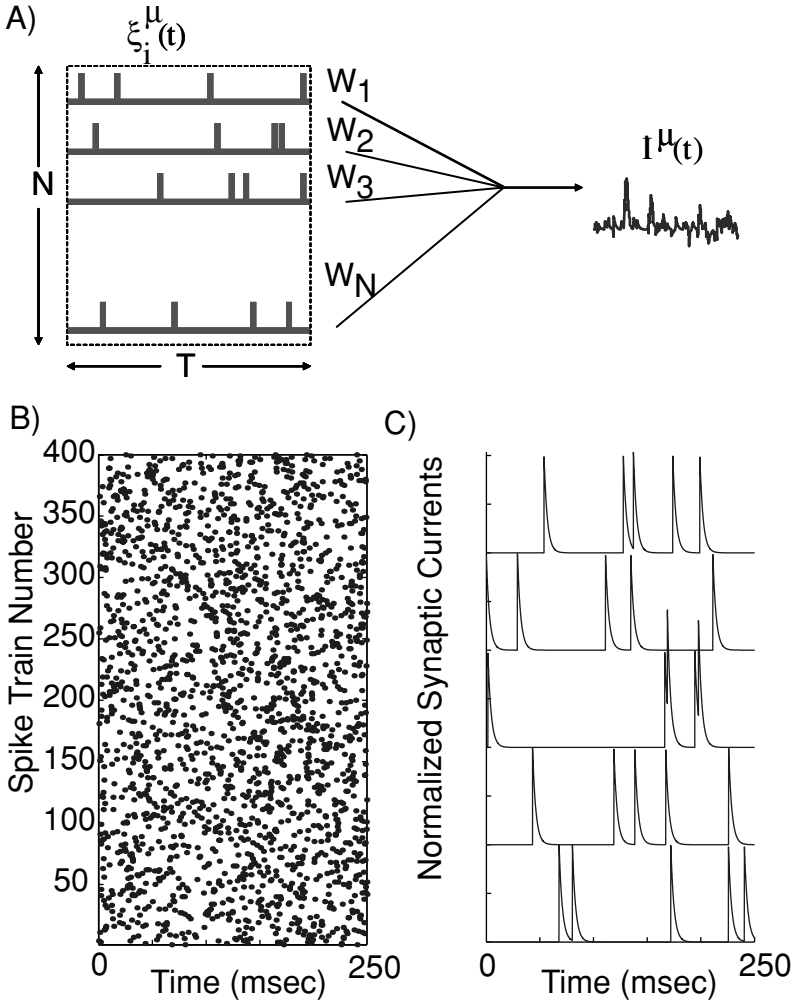$$\xi_i^{\mu}(t) = \sum_{k=1}^{n} \delta(t - t_{i,k}^{\mu}). \tag{2.1}$$

Figure 1: The model and sample input. (A) The components of the model. A spatiotemporal input pattern $\xi_i^\mu(t)$ is fed via synapses to a readout neuron. Each presynaptic spike from neuron i contributes a decaying exponential with time constant τ_S and initial height W_i to the input current. (B) A typical input pattern; each dot represents a spike. (C) Normalized synaptic currents ($W = 1$) for five input neurons (arbitrary units).

Here $I_i^\mu(t)$ are normalized synaptic currents, resulting from the spike train of a single input neuron, with the synaptic weights factored out (see Figure 1C).

The following leaky integrate-and-fire model (Lapicque, 1907; Tuckwell, 1988) will be used for an illustration of a typical readout neuron:

$$\tau_m \dot{V} = -V + I,$$ (2.2)

where I is the input current and V the membrane potential. Once V reaches a threshold, an output spike is generated, and V is immediately set to a constant reset value. We omitted a refractory period in this model since the output rates are not high.

3 Results

3.1 Moments of the Input Current.
As discussed in section 1, we focus our attention on the mean and variance of the input current to the neuron upon presentation of an input pattern. In doing so, we are relying on the monotonic relations between these values and the spiking output of the readout neuron.

The mean input current upon presentation of pattern μ (assuming $T \gg \tau_S$) is computed (see the appendix) to be $\overline{I^\mu} = \frac{\tau_S}{T} \sum_{i=1}^N W_i n \, (\overline{f} = \frac{1}{T} \int_0^T f \, dt)$. Since there is no pattern-specific information contained in the mean current, it cannot be used for recognition. We therefore consider the variance of the input current to the readout neuron on presentation of pattern μ (abbreviated as variance of pattern μ), and calculate it to be

$$\text{Var}(I^\mu) \stackrel{def}{=} \overline{(I^\mu)^2} - \left(\overline{I^\mu}\right)^2 = \mathbf{W}^T C^\mu \mathbf{W},$$ (3.1)

where

$$C_{ij}^\mu = \overline{\widetilde{I_i^\mu} \widetilde{I_j^\mu}} = \frac{\tau_S}{2T} \sum_{k_i, k_j = 1}^n e^{-\frac{\left| t_{i,k_i}^\mu - t_{j,k_j}^\mu \right|}{\tau_S}} - \left(\frac{n\tau_S}{T}\right)^2$$

$$\widetilde{I_i^\mu} \stackrel{def}{=} I_i^\mu - \overline{I_i^\mu}.$$ (3.2)

We see that the variance depends on the matrix C^μ, which is the temporal covariance matrix of the normalized synaptic currents (see Figure 1C) resulting from presentation of pattern μ. The ijth element of C^μ can be described as counting the fluctuations in the number of coincidences between pattern μ's spike trains i and j, with a temporal resolution of τ_S.

Our aim is to maximize the variance for all learned patterns, leading to the cost function $-\sum_\mu \text{Var}(I^\mu)$. This is not a priori the best choice for a cost function, as different functions can perhaps take into account the spread of variances for the different patterns, but it is the most natural one.

Our problem is therefore to maximize $\mathbf{W}^T C \mathbf{W}$ (see equation 3.1), where $C = \sum_\mu C^\mu$, with the constraint $\sum_{i=1}^N W_i^2 = 1$.

3.2 Derivation of Learning Rules. The solution to the maximization problem is known to be the leading eigenvector of C; however, we are also interested in deriving learning rules for $\mathbf{W}$. By performing stochastic projected gradient descent (Kelley, 1962) on the cost function with the norm constraint on $\mathbf{W}$, we derive the following rule:

$$\dot{W}_i(t) = \eta(t) \left[\tilde{I}^\mu(t) \left(\tilde{I}_i^\mu(t) - \tilde{I}^\mu(t) W_i(t) \right) \right], \tag{3.3}$$

which is similar to the online Oja's rule (Oja, 1982), with the input vector composed of the instantaneous normalized synaptic currents and the output being the total input current to the neuron. This is to be expected since the optimal $\mathbf{W}$ is the leading eigenvector of the covariance matrix of the normalized synaptic currents.

Since Oja's rule requires many iterations to converge, it is of interest to also look for a "one-shot" learning rule. We construct a correlation-based rule by using the coincidences between the different inputs,

$$W_i = \kappa \sum_{j=1}^N C_{ij}, \tag{3.4}$$

with κ a normalization factor.

Figure 2 demonstrates the convergence of Oja's rule to the optimal set of weights $\mathbf{W}$. It can also be seen that the correlation rule is a step in the right direction, though suboptimal. We also verified that the correlation between the leading eigenvector and the weight vector resulting from Oja's rule converges to 1, ensuring that the convergence is in the weights, and not only in performance.

3.3 Simulation Results. The main goal of the simulations was to assess the performance of the model for an increasing number of learned patterns. Unless otherwise stated, the parameter values in all simulations were $N = 400$ inputs, $T = 250$ msec, $n = 5$ spikes, and $\tau_S = 3$ msec. All numerical simulations were performed in MATLAB. As explained in section 2, we compare the input current variance for learned and background patterns.

As shown earlier, the Oja learning rule converges to the leading eigenvector of the matrix C (see Figure 2), and we therefore used the MATLAB *eig* function instead of applying the learning rule in all simulations. An obvious exception is the illustration of convergence for the Oja rule.

Figure 3A shows that both rules lead to higher input variances for the learned patterns relative to the background patterns, and, as expected, the
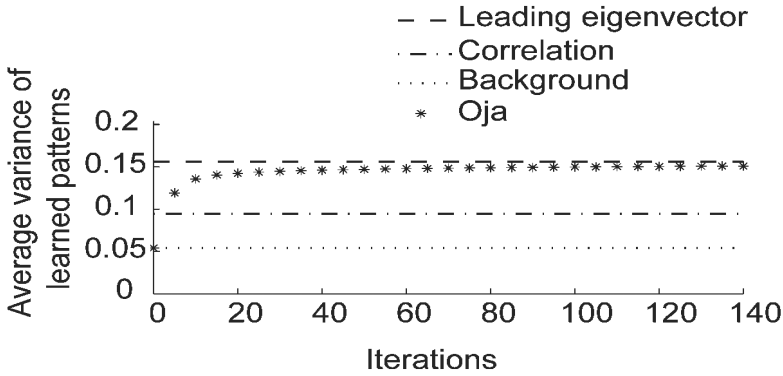
Figure 2: Convergence of Oja's rule. The average variance of five learned patterns. Oja's rule starts at the variance of the background patterns but quickly converges to the upper limit defined by the leading eigenvector of C. The correlation rule can be seen to be better than the background but suboptimal. The Oja learning step declined with iterations according to $\eta_i = 0.02/(1 + \lceil i/5 \rceil)$ (see equation 3.3).

Oja rule performs better. This is also illustrated for a specific pair of patterns in Figure 3B, where the difference in variances is evident. Although we took a neural model independent approach, we illustrate with a specific realization of the readout neuron—a leaky integrate-and-fire model (see section 2 and equation 2.2)—that the difference in the input variance can be used to elicit a pattern-specific spiking response (see Figures 3C and 3D).

In order to quantify the performance of the model, we defined the probability for false alarm (see Figure 4A) as the probability of classifying a background pattern as a learned one, if a classification threshold is put at the variance of the worst (or 10th percentile) learned pattern variance. Probability for false alarm was calculated for each condition by creating p learned patterns, choosing the appropriate (Oja or correlation) weights, and calculating the variance of learned patterns. Then an additional 500 patterns were generated to estimate the distribution of background variances for each set of weights. The fraction of background variances above the minimal (or 10th percentile) learned pattern variance was calculated. The probability for false alarm was estimated as the mean of repeating this process 50 times. Figure 4B shows that for a small number of patterns, both learning rules perform quite well. However, as the number of patterns increases, the correlation rule quickly deteriorates, while the Oja rule exhibits a much later and slower decline in performance.
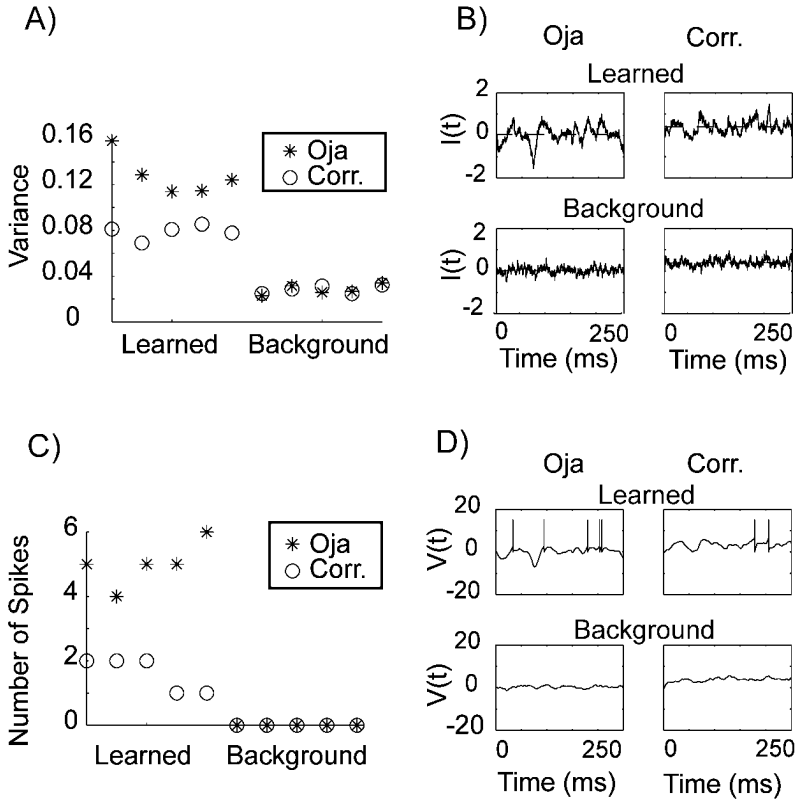
Figure 3: Simulation results for recognition. (A, B) Input current to the readout neuron. Note that in B, the mean current for the correlation rule is larger than that of the Oja rule, but there is no difference in mean between learned and background patterns. (C, D) Membrane potential and spike output for the leaky integrate-and-fire readout neuron with $\tau_m = 10$ msec and threshold of 2.6 and 6.5 for Oja and correlation rules, respectively.

Figure 4: Quantification of the model's performance. (A) The estimation of probability for false alarm. Sixty patterns were learned and their variances plotted (see the dots at the bottom of the figure). The variances of 500 background patterns were used to estimate their distribution. The results of choosing different recognition thresholds are shown. (B, C) Probability for false alarm as a function of the number of learned patterns for 0% (B) and 10% (C) miss. Notice that the x-axis is in log-scale, illustrating the slow degradation in performance. Values of 0 and 1 are actually less than 0.001 and more than 0.999, respectively, due to the finite sample used to estimate the background variance distribution. Error bars are smaller than the markers.

A) Probability Density vs. Variance, showing distribution with "0% Miss" and "10% Miss" thresholds, "False Alarms: 22%" and "False Alarms: 3%".

B) 0% Miss — Probability for False Alarm vs. Number of Learned Patterns (Oja, Corr.)

C) 10% Miss — Probability for False Alarm vs. Number of Learned Patterns (Oja, Corr)
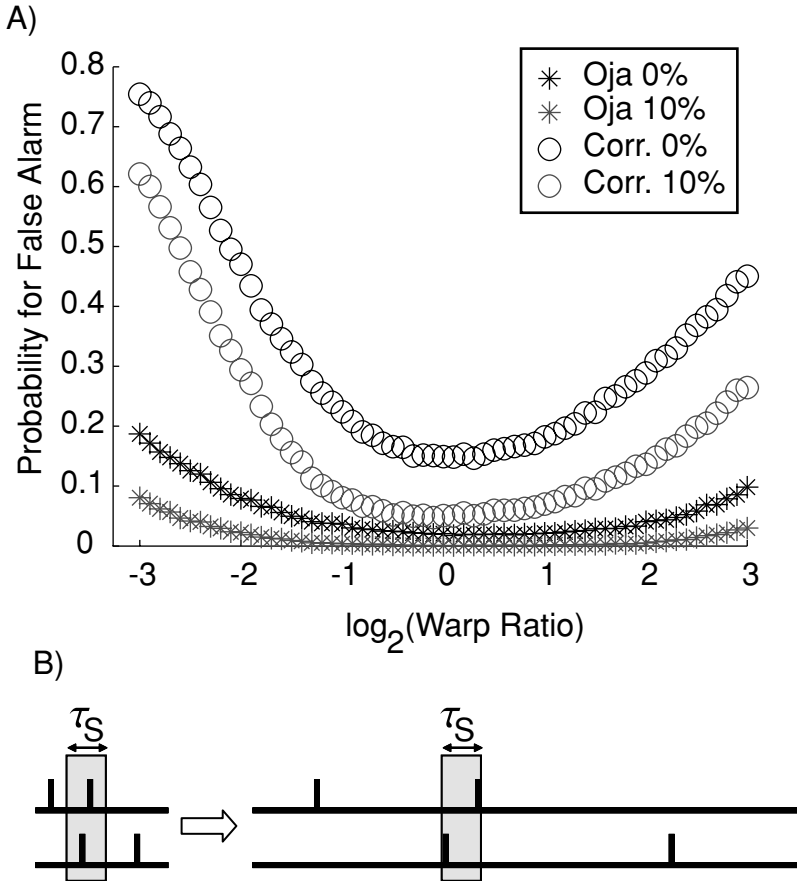
A)

B)
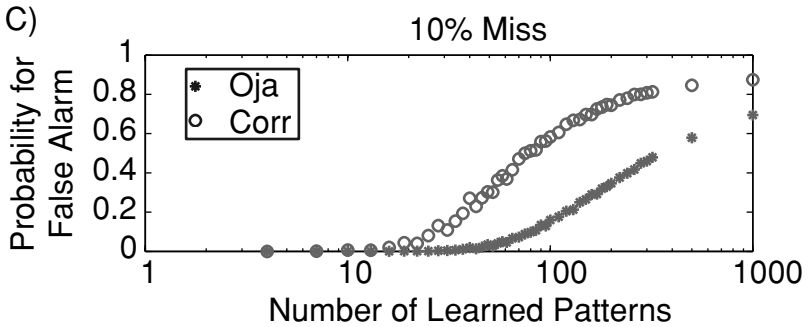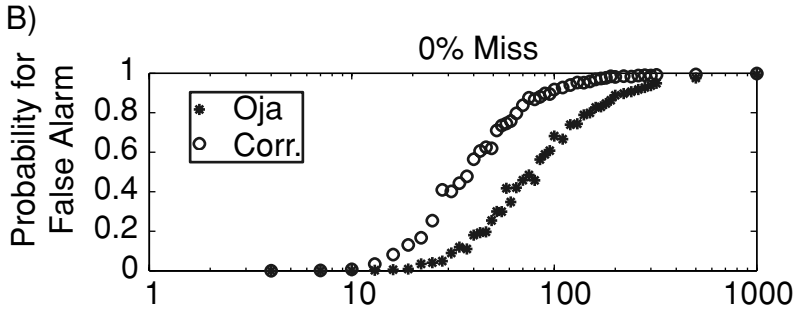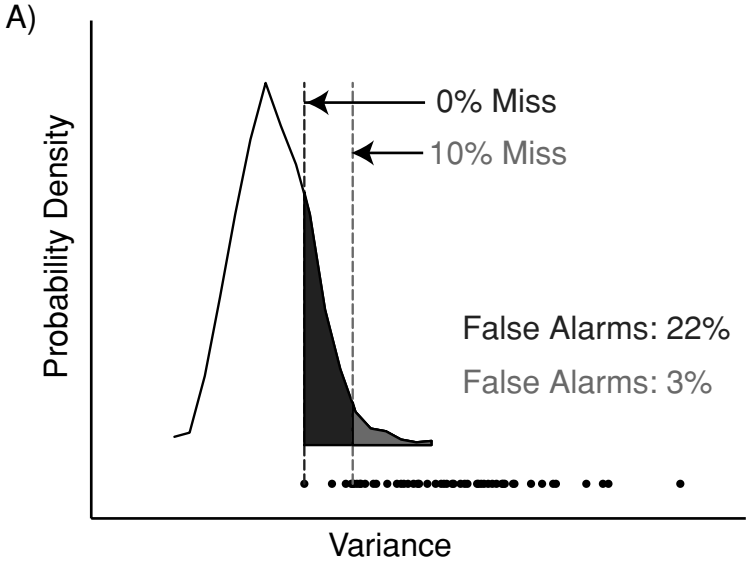
Figure 5: Robustness to time warp. (A) Twenty patterns were learned at $T = 250$ msec, and then the model was tested with warped versions of various ratios. The performance hardly changes over a large range. Note that the x-axis is in $\log_2$ scale. (B) Schematic explanation of the reason for this robustness. The model relies on the number and quality of input coincidences, not on their location, making it robust to time warp.

Since the matrix C depends on the number and temporal precision of the coincidences in the input, and not on their exact location in time (see equation 3.2), we expected that the model's performance would be robust to time warp effects (see Figure 5B). To quantify this robustness, a set of 20 patterns was learned to obtain the synaptic weights **W**. Warped versions of both learned and background patterns were formed by setting $\mathbf{t}^\mu \rightarrow \alpha\mathbf{t}^\mu$, $T \rightarrow \alpha T$. The original weights were then used to calculate

the variance of the warped versions and thus determine the probability for false alarm. Note that this warp changes both the firing rate and the average pattern variance, resulting in a different threshold for each α. A biological realization of this dynamic threshold can use the firing rate as a plausible tuning signal. Figure 5A shows that indeed there is a large plateau, indicating a small decrease in performance for a large range of expansion or contraction of the patterns.

3.4 Capacity Estimate. The performance results presented in the previous section are based on numerical simulations. In order to get a rough analytical estimate of the capacity, we first approximate the input spikes as Poisson trains to facilitate calculations. We then compare the variance of a random pattern to that obtained by the correlation rule. The capacitance will be defined as the limit where the two variances coincide.

Since the C^μ matrices depend on coincidences with a temporal resolution of τ_S, we approximate the normalized synaptic currents by using independent random binary variables x_{it}^μ:

$$x_{it}^\mu = \frac{1}{\sqrt{2}} \begin{cases} -\frac{n\tau_S}{T}, & \text{with probability } 1 - \frac{n\tau_S}{T} \\ 1 - \frac{n\tau_S}{T}, & \text{with probability } \frac{n\tau_S}{T} \end{cases} \tag{3.5}$$

$$< x > = 0 \tag{3.6}$$

$$< x^2 > = \frac{n\tau_S}{2T} \tag{3.7}$$

$$< x^m > \cong \frac{n\tau_S}{2^{m/2}T}, \quad m > 1 \tag{3.8}$$

$$C_{ij}^\mu = \frac{\tau_S}{T} \sum_{t=0}^{T/\tau_S} x_{it}^\mu x_{jt}^\mu, \tag{3.9}$$

where $< f(t) > = \int f \, dP(\mathbf{t})$.

For a background pattern $\{x_{it}\}$, the W's and the x_{it}'s are independent variables, since the W's are determined by the learned patterns. Thus, the variance of a background pattern is given by

$$< \text{Var}_{bg} > = \sum_{i,j=1}^N W_i W_j \frac{\tau_S}{T} \sum_{t=0}^{T/\tau_S} < x_{it} x_{jt} >$$

$$= \sum_{i=1}^N W_i^2 < x^2 >$$

$$= < x^2 > = \frac{n\tau_S}{2T}. \tag{3.10}$$

For the correlation rule, the synaptic weights are given by

$$W_i = \kappa \sum_{j=1}^{N} C_{ij}$$

$$= \sum_{\mu=1}^{p} \sum_{j=1}^{N} \frac{\tau_S}{T} \sum_{t=0}^{T/\tau_S} x_{it}^{\mu} x_{jt}^{\mu}. \tag{3.11}$$

We find κ from the normalization constraint:

$$1 = < \sum_{i=1}^{N} W_i^2 >$$

$$= \kappa^2 \left(\frac{\tau_S}{T}\right)^2 \sum_{\mu,\nu=1}^{p} \sum_{i,j,k=1}^{N} \sum_{t,s=0}^{T/\tau_S} < x_{it}^{\mu} x_{jt}^{\mu} x_{is}^{\nu} x_{ks}^{\nu} >$$

$$= \left[\kappa^2 < x^4 > \left(\frac{Np}{T/\tau_S}\right) + < x^2 >^2 \left[Np^2 + \frac{N^2 p}{T/\tau_S} - 2\left(\frac{Np}{T/\tau_S}\right) \right] \right]$$

$$\cong \kappa^2 \frac{Np}{4T/\tau_S} \left[\frac{(N-2)n}{T/\tau_S} + np + 1 \right]. \tag{3.12}$$

The variance of pattern μ can now be estimated as

$$< \mathrm{Var}_{Corr}(\mu) > = \kappa^2 \left(\frac{\tau_S}{T}\right)^3 \sum_{\eta,\nu=1}^{p} \sum_{i,j,k,l=1}^{N} \sum_{t,s,v=0}^{T/\tau_S} < x_{it}^{\mu} x_{jt}^{\mu} x_{is}^{\nu} x_{ks}^{\nu} x_{jv}^{\eta} x_{lv}^{\eta} >$$

$$= \kappa^2 \frac{nNp\tau_S}{8T} [1 - 8n\tau_S/T + 3np - 6n^2 p\tau_S/T$$

$$+ 7n^2 \tau_S^2/T^2 + n^2 p^2 + 5nN\tau_S/T - 6n^2 N\tau_S^2/T^2$$

$$+ 3n^2 Np\tau_S/T + n^2 N^2 \tau_S^2/T^2]$$

$$\cong \frac{\tau_S n}{2T} \left[\frac{\left(\frac{N}{p^T/\tau_S}\right)^2 n + \left(\frac{N}{p^T/\tau_S}\right)(3n + O(1)) + n + O(1)}{\left(\frac{N}{p^T/\tau_S}\right) n + n + O(1)} \right] \tag{3.13}$$

$$= \begin{cases} \frac{n\tau_S}{2T} \left(\frac{N}{p^T/\tau_S}\right), & N \gg p\frac{T}{\tau_S} \\ \frac{n\tau_S}{2T} (1 + O(1)), & N \ll p\frac{T}{\tau_S}. \end{cases} \tag{3.14}$$

Thus, we see that unless $N \gg p\frac{T}{\tau_S}$, the variance of the learned patterns is similar to that of the background ones; they cannot be successfully

recognized. We conclude that the maximal number of patterns that can be recognized in the network with a correlation learning rule is on the order of $N\tau_S/T$. As we showed in the previous section, the performance of the Oja rule is superior to that of the correlation rule, but we do not have analytical estimates for its capacity.

4 Discussion

By considering the input current to the readout neuron as the model's output, we were able to derive learning rules that are independent of a particular response function of that neuron. These rules enabled recognition of a large number of spatiotemporal spiking patterns by a single neuron. An emergent property of the model, of particular importance when considering speech recognition applications, is robustness of the recognition to time warping of the input patterns. We illustrated that these model-independent rules are applicable to specific spiking models by showing pattern recognition by a leaky integrate-and-fire neuron.

A biological realization of the proposed learning rules requires plasticity to be induced as a function of the input to the neuron, without an explicit role for postsynaptic spikes. While at odds with the standard plasticity rules (Abbott & Nelson, 2000), there are some experimental indications that such mechanisms exist. Sjostrom, Turrigiano, and Nelson (2004) used a long-term depression protocol in which the postsynaptic cell is depolarized but does not generate action potentials. Nonlinear effects in the dendrites, such as NMDA spikes (Schiller, Major, Koester, & Schiller, 2000), can provide a mechanism for coincidence detection. It remains a challenge to see whether our proposed rules are compatible with those and similar mechanisms.

Although our model was defined as a single neuron receiving spike trains, the same framework is applicable for any linear summation of time-decaying events. One alternative realization could be using bursts of spikes as the elementary input events while still converging on a single readout neuron. This realization is in accord with experimental findings in rabbit retinal ganglion cells showing temporal accuracy of bursts in response to visual stimuli (Berry et al., 1997). A more general realization is an array of different local networks of neurons firing a spatiotemporal pattern of population spikes. The readout neuron in this case can be replaced by a single readout network, where the fraction of active cells stands for the magnitude of the input current.

The process of recognition entails several stages of processing, of which our model is but one. As such, our model does not address all the different aspects associated with recognition. For example, if the spike pattern is reversed, our model, which is sensitive to coincidences and not to directionality, will recognize the reversed pattern as if it was the original. Yet we know from everyday experience that reversed words sound completely different. A possible solution to this problem lies in the transformation

between the sensory stimulus and the spike train received by our model, which is not the immediate recipient of the sensory data. Since the transformation is not a simple linear one, time reversal of the stimulus will, in general, not lead to time reversal of the spike train associated with this stimulus.

Finally, we present some directions for future work. The performance of the model was quantified by defining the probability for a false alarm (incorrect identification of a background pattern as a learned one). Computer simulations showed a very slow degradation in the model's performance when the number of learned patterns increased. These results, however, are numerical and sample only a small volume of the parameter space. A more complete understanding of the model's behavior could be achieved by analytical estimation of this performance, which remains a challenge for a future study.

An emergent feature of the model, which is of great ecological significance, is robustness to time warping of the input patterns. There are, however, a variety of other perturbations to which recognition should be robust. These include background noise, probabilistic neurotransmitter release, and jitter in the spike timing. The robustness of the model to these perturbations was not addressed in this study.

Appendix: Derivation of the Moments of the Input Current

Starting from equation 2.1, we calculate the normalized synaptic currents,

$$I_i^\mu(t) = \sum_{k=1}^{n} \Theta(t - t_{i,k}^\mu) e^{-\frac{t - t_{i,k}^\mu}{\tau_S}}, \tag{A.1}$$

where Θ is the Heaviside step function.

We can now calculate the moments ($\overline{f} = \frac{1}{T} \int_0^T f \, dt$) of the input current upon presentation of pattern μ (assuming $T \gg \tau_S$):

$$\overline{I_i^\mu} = \frac{n \tau_S}{T} \tag{A.2}$$

$$\overline{I^\mu} = \frac{\tau_S}{T} n \sum_{i=1}^{N} W_i. \tag{A.3}$$

The variance of the input current can be calculated as follows:

$$\overline{I_i^\mu I_j^\mu} = \sum_{k_i, k_j = 1}^{n} \frac{\tau_S}{2T} e^{-\frac{\left| t_{i,k_i}^\mu - t_{j,k_j}^\mu \right|}{\tau_S}} \tag{A.4}$$

$$\overline{(I^\mu)^2} = \sum_{i,j=1}^{N} W_i W_j \overline{I_i^\mu I_j^\mu} \tag{A.5}$$

$$Var(I^\mu) \stackrel{def}{=} \overline{(I^\mu)^2} - \left(\overline{I^\mu}\right)^2 \tag{A.6}$$

$$= \sum_{i,j=1}^{N} W_i W_j \left[\overline{I_i^\mu I_j^\mu} - \overline{I_i^\mu}\ \overline{I_j^\mu} \right]$$

$$= \sum_{i,j=1}^{N} W_i W_j \left[\overline{\widetilde{I_i^\mu} \widetilde{I_j^\mu}} \right]$$

$$\stackrel{def}{=} \sum_{i,j=1}^{N} W_i W_j C_{ij}^\mu$$

$$= \mathbf{W}^T C^\mu \mathbf{W},$$

where

$$C_{ij}^\mu = \overline{\widetilde{I_i^\mu} \widetilde{I_j^\mu}} = \frac{\tau_S}{2T} \sum_{k_i,k_j=1}^{n} e^{-\frac{\left| t_{i,k_i}^\mu - t_{j,k_j}^\mu \right|}{\tau_S}} - \left(\frac{n\tau_S}{T} \right)^2$$

$$\widetilde{I_i^\mu} \stackrel{def}{=} I_i^\mu - \overline{I_i^\mu}. \tag{A.7}$$

Acknowledgments

References

Abbott, L., & Nelson, S. (2000). Synaptic plasticity: Taming the beast. *Nat. Neurosci., 3* (Suppl.), 1178–1183.

Aertsen, A., Smolders, J., & Johannesma, P. (1979). Neural representation of the acoustic biotope: On the existence of stimulus-event relations for sensory neurons. *Biol. Cybern., 32*(3), 175–185.

Bair, W., & Koch, C. (1996). Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Comput., 8*(6), 1185–1202.

Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spiketrains. *PNAS, 94*(10), 5411–5416.

Bressloff, P., & Taylor, J. (1992). Perceptron-like learning in time-summating neural networks. *Journal of Physics A: Mathematical and General, 25*(16), 4373–4388.

Hopfield, J., & Brody, C. (2001). What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proc. Natl. Acad. Sci. USA, 98*(3), 1282–1287.

Jin, D. (2004). Spiking neural network for recognizing spatiotemporal sequences of spikes. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys., 69*(2 Pt. 1), 021905.

Kelley, H. (1962). Method of gradients. In G. Leitmann (Ed.), *Optimization Techniques* (pp. 205–254). New York: Academic Press.

Lapicque, L. (1907). Recherches quantitatifs sur l'excitation électrique des nerfs traiteé comme une polarisation. *J. Physiol. Paris, 9*, 622–635.

Margoliash, D., & Konishi, M. (1985). Auditory representation of autogenous song in the song system of white-crowned sparrows, *PNAS, 82*(17), 5997–6000.

Minsky, M., & Papert, S. (1969). *Perceptrons: An Introduction to computational geometry.* Cambridge, MA: MIT Press.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *J. Math. Biol., 15*(3), 267–273.

Rauch, A., La Camera, G., Luscher, H., Senn, W., & Fusi, S. (2003). Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo–like input currents. *J. Neurophysiol., 90*(3), 1598–1612.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code.* Cambridge, MA: MIT Press.

Schiller, J., Major, G., Koester, H., & Schiller, Y. (2000). NMDA spikes in basal dendrites of cortical pyramidal neurons. *Nature, 404*(6775), 285–289.

Silberberg, G., Bethge, M., Markram, H., Pawelzik, K., & Tsodyks, M. (2004). Dynamics of population rate codes in ensembles of neocortical neurons. *J. Neurophysiol., 91*(2), 704–709.

Sjostrom, P., Turrigiano, G., & Nelson, S. (2004). Endocannabinoid-dependent neocortical layer-5 LTD in the absence of postsynaptic spiking. *J. Neurophysiol., 92*(6), 3338–3343.

Tuckwell, H. (1988). *Introduction to theoretical neurobiology: Vol. 2: Nonlinear and stochastic theories.* Cambridge: Cambridge University Press.

Uzzell, V., & Chichilnisky, E. (2004). Precision of spike trains in primate retinal ganglion cells. *J. Neurophysiol., 92*(2), 780–789.