Random walks for spike-timing-dependent plasticity

Alan Williams*
Neurological Sciences Institute, Oregon Health & Science University, 505 NW 185th Avenue, Beaverton, Oregon 97006, USA

Todd K. Leen†
Department of Computer Science and Engineering, OGI School of Science & Engineering, Oregon Health & Science University, 505 NW 185th Avenue, Beaverton, Oregon 97006, USA

Patrick D. Roberts‡
Neurological Sciences Institute, Oregon Health & Science University, 505 NW 185th Avenue, Beaverton, Oregon 97006, USA

(Received 23 December 2003; revised manuscript received 3 May 2004; published 31 August 2004)

Random walk methods are used to calculate the moments of negative image equilibrium distributions in synaptic weight dynamics governed by spike-timing-dependent plasticity. The neural architecture of the model is based on the electrosensory lateral line lobe of mormyrid electric fish, which forms a negative image of the reafferent signal from the fish's own electric discharge to optimize detection of sensory electric fields. Of particular behavioral importance to the fish is the variance of the equilibrium postsynaptic potential in the presence of noise, which is determined by the variance of the equilibrium weight distribution. Recurrence relations are derived for the moments of the equilibrium weight distribution, for arbitrary postsynaptic potential functions and arbitrary learning rules. For the case of homogeneous network parameters, explicit closed form solutions are developed for the covariances of the synaptic weight and postsynaptic potential distributions.

DOI: 10.1103/PhysRevE.70.021916 PACS number(s): 87.19.La, 87.18.Sn, 75.10.Nr

I. INTRODUCTION

Spike-timing-dependent plasticity (STDP) [1] is a form of synaptic weight dynamics found experimentally in certain neural systems [2–4]. The key feature of STDP is the dependence of synaptic weight changes on the precise relative timing of presynaptic and postsynaptic spikes; this timing dependence distinguishes STDP from earlier hypothesized forms of activity-dependent plasticity [5–7] in which weight changes depend only on correlations between presynaptic and postsynaptic spike rates. Models of STDP assume that the weight change due to each presynaptic and postsynaptic spike pair is given by some function of the time between them, called the spike-timing-dependent *learning rule* [8–13]. Changes due to all pairs of presynaptic and postsynaptic spike pairs are then summed to give the weight change due to presynaptic and postsynaptic spike trains.

In a previous article [14], we investigated the *mean* weight dynamics in a system in which STDP has been found experimentally: the electrosensory lateral line lobe (ELL), a cerebellum-like structure in mormyrid electric fish [3]. The mormyrid fish uses an adaptation mechanism based on STDP to habituate central neural responses to the predictable sensory input due solely to its own electric organ discharge (EOD). In order for the adaptation to predictable temporal patterns to be maintainable, the synaptic weight configuration giving rise to a negative image of predictable patterns must be a stable equilibrium for the mean weight dynamics induced by the spike-timing-dependent learning rule. Conditions for the existence and stability of such negative image equilibria were first explored in [15] and extended to arbitrary spike-timing-dependent learning rules and arbitrary postsynaptic potential functions in [14].

However, the equilibrium weight *distribution* in the presence of noise—and in particular, that distribution's variance—is also behaviorally important, since fluctuations in the weights due to noise lead to fluctuations in the negative image, which impacts the detectability of external objects. The methods of our previous article [14] were sufficient to calculate the equilibrium mean, but not any higher moments of the equilibrium weight distribution. This is a serious limitation in the biological setting, for two reasons: first, because in principle the variance could be so large that the fluctuations are more physiologically relevant than the mean and, second, because even if the variance is small, it is important to be able to calculate it quantitatively in order to make specific predictions about the impact of fluctuations on detectability. In the present article, we derive implicit expressions for all moments of the equilibrium weight distribution and explicit expressions for the variance (and the third and fourth moments in the single-weight case) for STDP learning rules with stable learning dynamics.

Our approach is to express the weight dynamics as a discrete time, inhomogeneous random walk. From the master equation of this walk we derive a differential equation for the Fourier transform of the equilibrium weight distribution. Taylor expansion of this equation yields recurrence relations for the moments.

The structure of the paper is as follows. In Sec. II we summarize the background facts about random walks, master equations, and characteristic functions that will be used in the present paper. In Sec. III we describe the architecture and

*Electronic address: williaal@ohsu.edu
†Electronic address: tleen@cse.ogi.edu
‡Electronic address: robertpa@ohsu.edu

dynamical assumptions of the model, and in Sec. IV we derive the random walk for the weight dynamics for arbitrary system parameters. In Sec. V we illustrate the method for deriving recurrence relations for the moments of the equilibrium weight distribution by applying the method in the simplest possible setting: the case of a single synaptic weight. We then in Sec. VI apply the method to the full architecture, with arbitrary system parameters. In Sec. VII we specialize to the case of homogeneous system parameters, deriving more explicit analytical results for the covariance of the equilibrium weight and postsynaptic potential distributions. Finally in Sec. VIII we compute the weight and postsynaptic potential covariances for several examples of biological interest and compare our predictions with Monte Carlo simulations. In Sec. IX we summarize our findings, discuss their biological relevance, and suggest future experiments to test the quantitative predictions of the model.

II. RANDOM WALKS, MASTER EQUATIONS, AND CHARACTERISTIC FUNCTIONS

The term *random walk* refers to any stochastic process in which the state variables change only at discrete times. The changes in state variables are called *steps*; from any given position there is a set of possible steps, each having a certain probability (or probability density). The set of possible steps may be discrete or continuous, and both the step values and step probabilities may depend on position.

Random walks have been used extensively to model other physical systems (see the bibliography in [16]), and a large body of mathematical technique has been developed for their analysis [17]. But they have not previously been applied to STDP, where the standard approach has been to use the Fokker-Planck equation [9,10,13]. Given that the Fokker-Planck equation is at best an approximation[1] when applied to discrete stochastic processes [18], whereas random walk methods are exact, we believe it would be prudent to explore the utility of random walk methods for the analysis of STDP.

Random walks are natural models for systems having temporally discrete dynamics. Since the synaptic weight changes in STDP are due to temporally discrete events (spikes or spike pairs), random walks are natural models for STDP.

Suppose a state variable w undergoes a random walk. Let the possible steps from position w be $j_w(x)$ for x in some index set X. Let the step $j_w(x)$ occur with probability density $\rho_w(x)$ in x. Let $P_n(w)$ be the probability distribution for w after n steps. We wish to derive the equation of motion for $P_n(w)$, usually referred to as the *master equation*.

If the state variable is w' after n steps and w after $n+1$ steps, then $w=w'+j_{w'}(x)$ for some x. The probability for the state variable to be between w and $w+dw$ after $n+1$ steps is therefore

$$P_{n+1}(w)dw = \int dx\rho_w(x)[P_n(w')dw'].$$

Hence the master equation is

$$P_{n+1}(w) = \int dx\rho_{w'}(x)P_n(w')\frac{dw'}{dw}.$$

The quantity dw'/dw compensates for any change in the density of states from time n to time $n+1$, due to position dependence of the set of step values. From $w=w'+j(x,w')$ we have

$$\frac{dw'}{dw} = \frac{1}{1+\frac{\partial}{\partial w'}j_{w'}(x)}, \tag{1}$$

and hence the master equation is

$$P_{n+1}(w) = \int dx\rho_{w'}(x)P_n(w')\frac{1}{1+\frac{\partial}{\partial w'}j_{w'}(x)}. \tag{2}$$

Suppose the set of step values is independent of position; then $\partial j_{w'}(x)/\partial w'=0$, and the density of states factor in the master equation is identically 1. Denoting by $j(x)$ the common set of step values, we also have w' explicitly in terms of w and x: $w'=w-j(x)$. For such walks the master equation takes the simpler form

$$P_{n+1}(w) = \int dx\ \rho_{w-j(x)}(x)P_n(w-j(x)). \tag{3}$$

All walks considered in the present paper will turn out to be of this type.

A probability distribution $P(w)$ is an equilibrium (stationary) distribution for the random walk if $P_n=P$ implies $P_{n+1}=P$; in other words, the dynamics of the walk leaves P unchanged. Hence $P(w)$ is an equilibrium distribution for the walk in Eq. (3), if and only if it satisfies

$$P(w) = \int dx\ \rho_{w-j(x)}(x)P(w-j(x)). \tag{4}$$

To calculate the moments of a probability distribution $P(w)$, we will find it useful to invoke a property of its Fourier transform (often referred to as the *characteristic function*):

$$\hat{P}(k) = \int dw\ e^{ikw}P(w). \tag{5}$$

Taking the derivative with respect to k in Eq. (5) and evaluating at $k=0$ yields

$$\left.\frac{d^n\hat{P}(k)}{dk^n}\right|_{k=0} = \left.\left(\int dw(iw)^n e^{ikw}P(w)\right)\right|_{k=0} = i^n\int dww^nP(w) = i^n\langle w^n\rangle. \tag{6}$$

Hence the moments of $P(w)$ are, up to powers of i, just the

[1] Moreover, the conditions under which the approximation is a good one, especially for the nonlinear Fokker-Planck equation, are far from clear [18]. Further discussion of this issue, in the context of STDP, will be the subject of a future paper.

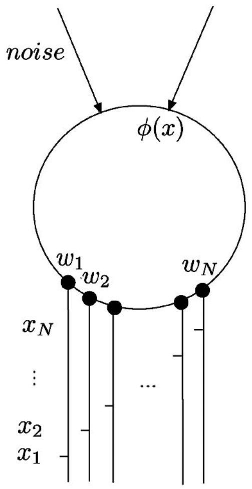

FIG. 1. Schematic of the architecture. The postsynaptic cell receives inputs from N presynaptic neurons, a repeated sensory input $\phi(x)$, and a noisy input. Presynaptic cell i spikes at time x_i in each period of ϕ and has synaptic weight w_i onto the postsynaptic cell.

derivatives of the characteristic function $\hat{P}(k)$ evaluated at $k=0$.

For further background on random walks, see [17].

III. FRAMEWORK

The model consists of a single postsynaptic cell representing a medium ganglion (MG) cell, a cell type in the ELL that shows strong adaptation to changing sensory input [3]. The MG cell is driven by a repeated sensory input (primary sensory reafference), an array of presynaptic cells whose spikes are time-locked to the repeated sensory input (the efference copy of the motor command), and noise (representing other unspecified inputs) [19–21] (Fig. 1). This basic architecture is derived from the mormyrid ELL, but is sufficiently general to capture the dynamics of other neural systems hypothesized to have an array of time-delayed, time-locked inputs through plastic synapses [22,23].

The framework for the neural dynamics is the spike response (SR) model [24,25], without refractoriness, as described in our previous report [14]. Much of the details of our MG cell model have appeared previously [14,21], and here we shall outline our general methods and comment on important differences between previous treatments and the present framework.

The repeated sensory input is the postsynaptic potential (PSP) in the postsynaptic cell due to primary sensory afferents, over a single EOD sweep. Each time-locked presynaptic cell i spikes (exactly once) at a fixed time within each sweep of the repeated sensory input, causing a corrsponding PSP in the postsynaptic cell.

The total membrane potential in the postsynaptic cell is the sum of the repeated sensory input, the noisy input, and the PSPs due to time-locked presynaptic spikes, weighted by synaptic efficacies (weights) w_i. This membrane potential causes the postsynaptic cell to generate postsynaptic, dendritic spikes [3] at a certain (noisy) rate. We assume that each presynaptic spike causes a constant change in the weight w_i (nonassociative learning) and each postsynaptic and presynaptic spike pair causes a change in w_i according to a spike-timing-dependent learning rule—i.e., a function of the time difference between the postsynaptic and presynaptic spikes (associative learning).

Let the resulting period (pulse width) be T, and introduce two time variables: $x \in [0,T)$ for the time within each period of the sensory input and $t=nT$, $n \in \mathbb{Z}$, for the time of initiation of each such period [20,21,26]. General dynamical quantities will be functions of the pair (x,t). The time-locked presynaptic cell i spikes at a fixed time in each period. Denote this time by x_i. Let $w_i(x,t)$ be the synaptic weight of presynaptic cell i, and let $\mathcal{E}_i(s)$ be the PSP evoked by a spike in cell i at time s after the spike. We will assume that $\mathcal{E}_i$ is causal: $\mathcal{E}_i(s)=0$ for $s<0$. Let α_i be the nonassociative weight change due to a presynaptic spike by cell i and $\mathcal{L}_i(s)$ the associative weight change due to a postsynaptic spike at time s after a presynaptic spike by cell i. Let $\phi(x)$ be the periodic sensory input and $U(x,t)$ the total postsynaptic potential due to the non-noisy inputs.

In contrast to our previous approach [14], we will assume that, in each period of ϕ, either zero or one postsynaptic spike occurs. The probability density (in x, for a given t) for a postsynaptic spike to occur at (x,t) is assumed to be $\boldsymbol{(}1/T)f\boldsymbol{(}U(x,t)\boldsymbol{)}$ for some positive and strictly increasing function $f:\mathbb{R}\to[0,1]$. The probability of zero postsynaptic spikes in the period beginning at t is then $1-(1/T)\int_0^T dx f\boldsymbol{(}U(x,t)\boldsymbol{)}$. Heuristically, the function f is the effective gain function of the postsynaptic cell in the presence of the noisy inputs, with the maximum slope of f indicating the noise level: high or low noise corresponds to an f with small or large maximum slope, respectively.

We will implement changes in weights as discrete steps with no internal time course. We update weights synchronously, once per sweep of the periodic sensory input, at time $x=0$ for each $t=nT$, $n \in \mathbb{Z}$. The value of w_i in the period beginning at $(0,t)$ is then independent of x and will be denoted $w_i(t)$. In the present treatment, we impose no boundaries on the weight values because the weight equilibria and equilibrium variances are such that weights are almost always in the region that would be enclosed by biological bounds.

To simplify the derivation of the weight dynamics, we will assume that $\mathcal{E}_i(s), \mathcal{L}_i(s)$ are zero or negligible for $|s| > \tau_E, \tau_L$, respectively, with $\tau_E, \tau_L \ll T$. We will also impose the slow learning rate assumption $T \ll \tau_w$, where τ_w is the time scale over which weights undergo significant relative change. The existence of approximate negative image states requires [14] that the spacing of presynaptic spike times be much smaller than the widths of $\mathcal{E}_i$ and $\mathcal{L}_i$: $\delta \ll \tau_E, \tau_L$. These time-scale assumptions can be summarized as

$$\delta \ll (\tau_E, \tau_L) \ll T \ll \tau_w.$$

Typical values from the mormyrid ELL are $\delta < 1$ ms [27], $\tau_E \sim 20$ ms [3], $\tau_L \sim 40$ ms [3], $T \sim 80$ ms [[27] (b)], and $\tau_w \sim 10^2 T$ [3].

IV. WEIGHT DYNAMICS

We now derive the random walk for the weight dynamics by computing the possible weight changes $\Delta w_i(t) = w_i(t+T) - w_i(t)$ and their corresponding probabilities. The details here are very similar to those in [14], deviating only in the treatment of postsynaptic spike generation. Instead of a variable number of spikes per EOD cycle, occurring at a mean rate per unit time, we now have a single postsynaptic spike per cycle whose occurrence is given by a probability density.

The nonassociative change in $w_i(t)$ due to the single presynaptic spike at (x_i, t) is α_i. For the associative change due to presynaptic and postsynaptic spike pairs, the calculation is identical to that in [14]; for a pair consisting of a presynaptic spike at (x_i, t) and postsynaptic spike at (x, t), the change in $w_i(t)$ is approximately $\mathring{\mathcal{L}}_i(x - x_i)$, where $\mathring{\mathcal{L}}_i(s) = \sum_{n=-\infty}^{\infty} \mathcal{L}_i(s - nT)$ is the periodization of $\mathcal{L}_i$ with period T.

A postsynaptic spike between t and $t+T$ occurs with a probability density $(1/T) f\big(U(x,t)\big)$ in x, with the probability of zero postsynaptic spikes being $1-(1/T)\int_0^T dx f\big(U(x,t)\big)$. Hence the change in w_i due to postsynaptic spikes between t and $t+T$ is $\mathring{\mathcal{L}}_i(x)$ with density $(1/T) f(U(x,t))$ in x and 0 with probability $1-(1/T)\int_0^T dx f\big(U(x,t)\big)$. The total change in $w_i(t)$ due to both nonassociative and associative learning is therefore

$$\Delta w_i(t) = \begin{cases} \alpha_i + \mathring{\mathcal{L}}_i(x), & \text{density } f\big(U(x,t)\big), \\ \alpha_i, & \text{probability } 1 - (1/T)\displaystyle\int_0^T dx\ f\big(U(x,t)\big). \end{cases} \tag{7}$$

The calculation of the non-noisy component of the postsynaptic potential, $U(x,t)$, is the same as in [14]; we find that

$$U(x,t) = \phi(x) + \sum_{j=1}^{N} w_j(t) \mathring{\mathcal{E}}_j(x - x_j), \tag{8}$$

where $\mathring{\mathcal{E}}_i(s) = \sum_{n=-\infty}^{\infty} \mathcal{E}_i(s - nT)$ is the periodization of $\mathcal{E}_i$ with period T. Defining $\tilde{f}$ by $\tilde{f}\big(x, w(t)\big) = f\big(\phi(x) + \sum_{j=1}^{N} w_j(t) \mathring{\mathcal{E}}_j(x)\big)$, we have from Eqs. (7) and (8) the following expression for the total change in $w_i(t)$:

$$\Delta w_i(t) = \begin{cases} \alpha_i + \mathring{\mathcal{L}}_i(x), \ \text{density}(1/T)\tilde{f}\big(x, w_1(t), \ldots, w_N(t)\big), \\ \alpha_i, \ \text{probability } 1 - (1/T)\displaystyle\int_0^T dx \tilde{f}\big(x, w_1(t), \ldots, w_N(t)\big). \end{cases} \tag{9}$$

Equation (9) defines the random walk for the weight dynamics. It is discrete time (steps occur only at $t = nT$, $n \in \mathbb{Z}$), continuous space (steps can take a continuum of values), and inhomogenous (step probabilities depend on position).

The common periodicity of the functions $\mathring{\mathcal{E}}_i$, $\mathring{\mathcal{L}}_i$, and ϕ is an important feature, allowing the systematic use of Fourier techniques.

V. ONE WEIGHT

To illustrate the technique in the simplest possible setting, we first examine the case of a single weight. If there is only one weight $w_1(t)$, then without loss of generality we may take $x_1 = 0$ by translating ϕ if necessary. Writing $w(t)$, α, $\mathring{\mathcal{L}}$, and $\mathring{\mathcal{E}}$ for $w_1(t)$, α_1, $\mathring{\mathcal{L}}_1$, and $\mathring{\mathcal{E}}_1$, the random walk, Eq. (9), for the weight dynamics becomes

$$\Delta w(t) = \begin{cases} \alpha + \mathring{\mathcal{L}}(x), & \text{density}(1/T)\tilde{f}\big(x, w(t)\big), \\ \alpha, & \text{probability } 1 - (1/T)\displaystyle\int_0^T dx \tilde{f}\big(x, w(t)\big), \end{cases} \tag{10}$$

where

$$\tilde{f}\big(x, w(t)\big) = f\big(\phi(x) + w(t)\mathring{\mathcal{E}}(x)\big).$$

From the random walk for the weight dynamics we derive the moments of the equilibrium weight distribution in three steps. First we write the master equation for the time evolution of the probability distribution of the weight and the corresponding functional equation for the equilibrium (stationary) distribution. Taking the Fourier transform yields a differential equation for the Fourier transform of the equilibrium distribution. Taylor expansion of this equation yields recurrence relations for the moments.

Notice that the set of step values in the walk (10) is independent of w; hence the equilibrium distribution $P(w)$ must satisfy Eq. (4). From the step values and step probabilities in Eq. (10) we have

$$\begin{aligned} P(w) = {} & \left[1 - \frac{1}{T}\int_0^T dx \tilde{f}(x, w - \alpha)\right] P(w - \alpha) \\ & + \frac{1}{T}\int_0^T dx \tilde{f}\big(x, w - [\alpha + \mathring{\mathcal{L}}(x)]\big) P\big(w - [\alpha + \mathring{\mathcal{L}}(x)]\big). \end{aligned} \tag{11}$$

Taking the Fourier transform $\int dw\, e^{ikw}$ on both sides, changing variables, and rearranging yields

$$\begin{aligned} \hat{P}(k)[1 - e^{ik\alpha}] = {} & \frac{1}{T}\int_0^T dx [e^{ik[\alpha + \mathring{\mathcal{L}}(x)]} - e^{ik\alpha}] \\ & \times \int dw' e^{ikw'} \tilde{f}(x, w') P(w'). \end{aligned} \tag{12}$$

A physiologically plausible spike output function f would take the form of a smooth, monotonically increasing sigmoid, but for maximal simplicity we assume f is piecewise linear:

$$f(u)=\begin{cases}0, & u<-V-\theta,\\ \dfrac{1}{2T}\left(1+\dfrac{u-\theta}{V}\right), & -V-\theta\leqslant u\leqslant V-\theta,\\ \dfrac{1}{T}, & u>V-\theta,\end{cases} \tag{13}$$

so that $\tilde{f}$ is given by

$$\tilde{f}(x,w)=\begin{cases}0, & U(x)<-V-\theta,\\ \dfrac{1}{2T}\left(1+\dfrac{U(x)-\theta}{V}\right), & -V-\theta\leqslant U(x)\leqslant V-\theta,\\ \dfrac{1}{T}, & U(x)>V-\theta,\end{cases} \tag{14}$$

with $U(x)=\phi(x)-\theta+w\mathring{\mathcal{E}}(x)$.

We further assume that the equilibrium weight distribution $P(w)$ is zero or negligible for w such that $U(x)<-V-\theta$ or $U(x)>V-\theta$. This is a *confinement condition* on the equilibrium postsynaptic potential $U(x)$ and will be justified later. Note that the confinement condition helps justify the piecewise linear assumption on f, since the more "confined" the postsynaptic potential $U(x)$, the better our piecewise linear f approximates a smooth sigmoid in the region where $U(x)$ is concentrated. If the confinement condition holds, then in Eq. (12) we may replace $\tilde{f}(x,w')$ under the integral by the following linear function of w:

$$\frac{1}{2T}\left(1+\frac{\phi(x)-\theta+w\mathring{\mathcal{E}}(x)}{V}\right).$$

Using $\int dw\, e^{ikw}wP(w)=\hat{P}'(k)$, we then obtain

$$\hat{P}(k)\left[1-e^{ik\alpha}-\frac{1}{T}\int_0^T dx\frac{1}{2}\left(1+\frac{\phi(x)-\theta}{V}\right)\eta(x)\right]$$
$$=\frac{1}{i}\hat{P}'(k)\frac{1}{T}\int_0^T dx\frac{1}{2}\frac{\mathring{\mathcal{E}}(x)}{V}\eta(x), \tag{15}$$

where $\eta(x)=e^{ik[\alpha+\mathring{\mathcal{L}}(x)]}-e^{ik\alpha}$. By Eq. (6), the moments of $P(w)$ are (up to powers of i) just the derivatives of $\hat{P}(k)$ at $k=0$; since those derivatives are implicitly constrained by Eq. (15), the moments of $P(w)$ are constrained by Eq. (15). Specifically, the Taylor expansion of Eq. (15) around $k=0$ will yield a hierarchy of recurrence relations for the derivatives of $\hat{P}(k)$ and hence for the moments of $P(w)$. The Taylor expansions of the exponentials are

$$e^{ik\alpha}=\sum_{n=0}^{\infty}\frac{i^n}{n!}\alpha^n k^n,$$

$$e^{ik(\alpha+\mathring{\mathcal{L}}(x))}=\sum_{n=0}^{\infty}\frac{i^n}{n!}[\alpha+\mathring{\mathcal{L}}(x)]^n k^n.$$

For the expansion of the characterisitic function $\hat{P}(k)$ we expand the exponential in the definition of $\hat{P}(k)$ and invert the order of summation and integration:

$$\hat{P}(k)=\int dw\, e^{ikw}P(w)=\sum_{m=0}^{\infty}\frac{i^m}{m!}k^m\int dw\, w^m P(w)$$
$$=\sum_{m=0}^{\infty}\frac{i^m}{m!}\langle w^m\rangle k^m.$$

From this it follows that

$$\frac{1}{i}\hat{P}'(k)=\frac{1}{i}\sum_{m=0}^{\infty}\frac{i^m}{m!}\langle w^m\rangle k^{m-1}m=\sum_{m=0}^{\infty}\frac{i^m}{m!}\langle w^{m+1}\rangle k^m.$$

By substituting these expansions into Eq. (15) and equating coefficients of k^μ on both sides, we obtain the following relations:

$$\sum_{m=0}^{\mu}\binom{\mu}{m}[\gamma^{\phi}_{\mu-m}\langle w^m\rangle-\gamma^{E}_{\mu-m}\langle w^{m+1}\rangle]=0,$$
$$\mu=0,1,2,\ldots, \tag{16}$$

where for brevity we have defined

$$\gamma^{\phi}_n=\delta_{n,0}-\alpha^n-\frac{1}{T}\int_0^T dx\frac{1}{2}\left(1+\frac{\phi(x)-\theta}{V}\right)\{[\alpha+\mathring{\mathcal{L}}(x)]^n-\alpha^n\},$$

$$\gamma^{E}_n=\frac{1}{T}\int_0^T dx\frac{1}{2}\frac{\mathring{\mathcal{E}}(x)}{V}\{[\alpha+\mathring{\mathcal{L}}(x)]^n-\alpha^n\}.$$

The relations (16) are lower triangular[2] and hence are easily rearranged to yield explicit recurrence relations for the moments in terms of moments of lower degree only:

$$\langle w^\mu\rangle=-\frac{\gamma^{\phi}_{\mu}}{\mu\gamma^{E}_{1}}-\frac{1}{\mu\gamma^{E}_{1}}\sum_{m=1}^{\mu-1}\langle w^m\rangle\psi_{\mu,m},$$
$$\mu=1,2,\ldots, \tag{17}$$

where

$$\psi_{\mu,m}=\binom{\mu}{m}\gamma^{\phi}_{\mu-m}-\binom{\mu}{m-1}\gamma^{E}_{\mu-m+1}.$$

We may now compute the central moments $M_k=\langle(w-\langle w\rangle)^k\rangle$ by expressing $\langle w^n\rangle$ in terms of the $\{M_k\}$:

$$\langle w^n\rangle=\langle(w-\langle w\rangle+\langle w\rangle)^n\rangle=\sum_{k=0}^{n}\binom{n}{k}M_k\langle w\rangle^{n-k}. \tag{18}$$

Substituting into Eq. (17) and rearranging yields

[2] One could also derive moment equations via the more direct route of Taylor expanding, in w, the equilibrium condition (11) for $P(w)$, but the resulting moment equations are not triangular. In fact they are fully coupled (each equation involving all moments, in general) and hence not readily solvable.

$$M_\mu = -\left(\frac{\gamma_1^\phi}{\gamma_1^E}\right)^\mu - \frac{\gamma_\mu^\phi}{\mu\gamma_1^E} - \frac{1}{\mu\gamma_1^E}\sum_{m=1}^{\mu-1}\left(\frac{\gamma_1^\phi}{\gamma_1^E}\right)^m \psi_{\mu,m} + \sum_{k=2}^{\mu-1} M_k \left\{ -\binom{\mu}{k}\left(\frac{\gamma_1^\phi}{\gamma_1^E}\right)^{\mu-k} - \frac{1}{\mu\gamma_1^E}\sum_{m=k}^{\mu-1}\binom{m}{k}\left(\frac{\gamma_1^\phi}{\gamma_1^E}\right)^{m-k}\psi_{\mu,m}\right\}. \tag{19}$$

For $\mu=2,3,4$ we obtain

$$M_2 = -\frac{1}{2}\frac{\gamma_2^\phi}{\gamma_1^E} + \frac{1}{2}\frac{\gamma_1^\phi\gamma_2^E}{(\gamma_1^E)^2},$$

$$M_3 = -\frac{1}{3}\frac{\gamma_3^\phi}{\gamma_1^E} + \frac{1}{3}\frac{\gamma_1^\phi\gamma_3^E}{(\gamma_1^E)^2} + \frac{1}{2}\frac{\gamma_2^\phi\gamma_2^E}{(\gamma_1^E)^2} - \frac{1}{2}\frac{\gamma_1^\phi(\gamma_2^E)^2}{(\gamma_1^E)^3},$$

$$M_4 = \frac{3}{4}\frac{(\gamma_2^\phi)^2}{(\gamma_1^E)^2} - \frac{3}{2}\frac{\gamma_1^\phi\gamma_2^\phi\gamma_2^E}{(\gamma_1^E)^3} + \frac{3}{4}\frac{(\gamma_1^\phi)^2(\gamma_1^E)^2}{(\gamma_1^E)^4} - \frac{1}{4}\frac{\gamma_4^\phi}{\gamma_1^E} + \frac{1}{4}\frac{\gamma_1^\phi\gamma_4^E}{(\gamma_1^E)^2} + \frac{1}{2}\frac{\gamma_2^\phi\gamma_3^E}{(\gamma_1^E)^2} + \frac{1}{2}\frac{\gamma_3^\phi\gamma_2^E}{(\gamma_1^E)^2} - \frac{\gamma_1^\phi\gamma_2^E\gamma_3^E}{(\gamma_1^E)^3} - \frac{3}{4}\frac{\gamma_2^\phi(\gamma_2^E)^2}{(\gamma_1^E)^3} + \frac{3}{4}\frac{\gamma_1^\phi(\gamma_2^E)^3}{(\gamma_1^E)^4}. \tag{20}$$

We can see from M_3 alone that in general the equilibrium weight distribution is not Gaussian. For generic PSP $\mathcal{E}$ and learning rule $\mathcal{L}$ there are no polynomial relations among the coefficients γ_n^E and γ_n^ϕ, hence M_3 is generically nonzero.

To determine the dependence of the moments on step size, we multiply both α and $\mathcal{L}$, and hence the steps of the random walk, by a scalar λ. The coefficients γ_n^E and γ_n^ϕ are then both $O(\lambda^n)$, and substitution into Eq. (20) yields

$$M_2 = O(\lambda),$$

$$M_3 = O(\lambda^2),$$

$$M_4 = 3M_2^2 + O(\lambda^3).$$

Hence as $\lambda \to 0$ the skew and kurtosis approach Gaussian values:

$$(\text{skew}) = \frac{M_3}{M_2^{3/2}} = O(\lambda^{1/2}) \to 0,$$

$$(\text{kurtosis}) = \frac{M_4}{M_2^2} = 3 + O(\lambda) \to 3.$$

VI. MULTIPLE WEIGHTS

We now apply the technique to the case of multiple weights w_i, $i=1,2,\ldots,N$. The algebra is more complicated, but the structure of the derivation is identical to the single-weight case. For notational compactness we introduce the vector notation

$$w(t) = \begin{pmatrix} w_1(t) \\ \vdots \\ w_N(t) \end{pmatrix}, \quad \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix},$$

$$\mathring{\mathcal{E}}(x) = \begin{pmatrix} \mathring{\mathcal{E}}_1(x-x_1) \\ \vdots \\ \mathring{\mathcal{E}}_N(x-x_N) \end{pmatrix}, \quad \mathring{\mathcal{L}}(x) = \begin{pmatrix} \mathring{\mathcal{L}}_1(x-x_1) \\ \vdots \\ \mathring{\mathcal{L}}_N(x-x_N) \end{pmatrix}.$$

The random walk for the weight vector $w(t)$ takes place in $\mathbb{R}^{\mathbb{N}}$, with the walk for each component $w_i(t)$ given by Eq. (10). In vector notation the walk for $w(t)$ is then

$$\Delta w(t) = \begin{cases} \alpha + \mathring{\mathcal{L}}(x), \ \text{density}(1/T)\tilde{f}(x,w(t)), \\ \alpha, \ \text{probability}1 - (1/T)\int_0^T dx\tilde{f}(x,w(t)), \end{cases} \tag{21}$$

where

$$\tilde{f}(x,w(t)) = f[\phi(x) + w(t)\cdot\mathring{\mathcal{E}}(x)]$$

and the center dot $(\cdot)$ indicates the vector dot product.

Again, the step sizes are independent of position, so the equilibrium condition, Eq. (4), applies. We have

$$P(w) = \left[1 - \frac{1}{T}\int_0^T dx\tilde{f}(x,w-\alpha)\right]P(w-\alpha) + \frac{1}{T}\int_0^T dx\tilde{f}(x,w - [\alpha + \mathring{\mathcal{L}}(x)])P(w-[\alpha+\mathring{\mathcal{L}}(x)]). \tag{22}$$

As before, we take the (now n-dimensional) Fourier transform on both sides. Applying $\int dw\, e^{ik\cdot w}$, changing variables, and rearranging yields

$$\hat{P}(k)[1-e^{ik\cdot\alpha}] = \frac{1}{T}\int_0^T dx\,\eta(x)\int dw' e^{ik\cdot w'}\tilde{f}(x,w')P(w'), \tag{23}$$

where

$$\eta(x) = e^{ik\cdot[\alpha+\mathring{\mathcal{L}}(x)]} - e^{ik\cdot\alpha}. \tag{24}$$

We now assume that the postsynaptic gain function f is piecewise linear and given by Eq. (13); hence, $\tilde{f}$ is given by Eq. (14), with $U(x)=\phi(x)-\theta+w\cdot\mathring{\mathcal{E}}(x)$. And as before, we assume that $P(w)$ is negligible for w such that $U(x)<-V-\theta$ or $U(x)>V-\theta$, a confinement condition on $P(w)$, which will be justified later. Then we may replace $\tilde{f}(x,w')$ under the integral by the linear function of w:

$$\frac{1}{2T}\left(1 + \frac{\phi(x)-\theta+w\cdot\mathring{\mathcal{E}}(x)}{V}\right).$$

Using

$$\int dw\ e^{ik\cdot w}w_jP(w) = \frac{1}{i}\frac{\partial\hat{P}(k)}{\partial k_j},$$

we obtain the following first-order partial differential equation (PDE) for $\hat{P}(k)$:

$$\hat{P}(k)\left[1-e^{ik\cdot\alpha}-\frac{1}{T}\int_0^T dx\frac{1}{2}\left(1+\frac{\phi(x)-\theta}{V}\right)\eta(x)\right]$$
$$=\sum_{j=1}^{N}\frac{1}{i}\frac{\partial\hat{P}(k)}{\partial k_j}\frac{1}{T}\int_0^T dx\frac{1}{2}\frac{\mathring{\mathcal{E}}_j(x-x_j)}{V}\eta(x). \tag{25}$$

Taylor expansion of both sides of this equation around $k=0$ will yield recurrence relations for the moments of w. The Taylor expansion of a function g on $\mathbb{R}^{\mathbb{N}}$ is given by

$$g(k_1,\ldots,k_N)=\sum_{n=0}^{\infty}\frac{1}{n!}\sum_{0}\binom{n}{s_1\cdots s_N}$$
$$\times\left.\frac{\partial^n g(z_1,\ldots,z_N)}{\partial_{z_1}^{s_1}\cdots\partial_{z_N}^{s_N}}\right|_{z=0}\prod_l k_l^{s_l}. \tag{26}$$

The expansions of the complex exponentials in Eq. (25) are thus

$$e^{ik\cdot\alpha}=\sum_{n=0}^{\infty}\sum_{s}\frac{i^n}{n!}\binom{n}{s}\prod_l\alpha_l^{s_l}\prod_l k_l^{s_l},$$

$$e^{ik\cdot(\alpha+\mathring{\mathcal{L}}(x))}=\sum_{n=0}^{\infty}\sum_{s}\frac{i^n}{n!}\binom{n}{s}\prod_l[\alpha+\mathring{\mathcal{L}}(x)]_l^{s_l}\prod_l k_l^{s_l}, \tag{27}$$

where in the sums on the right, $s=(s_1s_2\cdots s_N)^T$ with each s_i a nonnegative integer and $\Sigma_{i=1}^N\ s_i=n$. For brevity we write $\binom{n}{s}$ for the multinomial coefficient in Eq. (26).

As before, for the expansion of the characteristic function $\hat{P}(k)$ we expand the exponential in the definition of $\hat{P}(k)$ and invert the order of summation and integration:

$$\hat{P}(k)=\int dw\ e^{ik\cdot w}P(w)=\sum_{m=0}^{\infty}\sum_{r}\frac{i^m}{m!}\binom{m}{r}$$
$$\times\left[\int dw\ P(w)\prod_l w_l^{r_l}\right]\prod_l k_l^{r_l}=\sum_{m=0}^{\infty}\sum_{r}\frac{i^m}{m!}\binom{m}{r}$$
$$\times\langle w_1^{r_1}w_2^{r_2}\cdots w_N^{r_N}\rangle\prod_l k_l^{r_l}, \tag{28}$$

where $r=(r_1r_2\cdots r_N)^T$ with each r_i a nonnegative integer and $\Sigma_{i=1}^N\ r_i=m$. From this expansion of $\hat{P}(k)$ it follows that

$$\frac{1}{i}\frac{\partial\hat{P}(k)}{\partial k_j}=\sum_{m=0}^{\infty}\sum_{r}\frac{i^m}{m!}\binom{m}{r}\langle w_1^{r_1}\cdots w_N^{r_N}\rangle r_j\prod_l k_l^{r_l-\delta_{lj}}. \tag{29}$$

Using the combinatorial identity

$$\frac{i^{m-1}}{m!}\binom{m}{r}r_j=\frac{i^{m-1}}{(m-1)!}\binom{m-1}{r_1\cdots r_j-1\cdots r_N},$$

we may reindex Eq. (29) to yield

$$\frac{1}{i}\frac{\partial\hat{P}(k)}{\partial k_j}=\sum_{m=0}^{\infty}\sum_{r}\frac{i^m}{m!}\binom{m}{r}\langle w_1^{r_1}\cdots w_j^{r_j+1}\cdots w_N^{r_N}\rangle\prod_l k_l^{r_l}. \tag{30}$$

When the expansions, Eqs. (27), (28), and (30), are substituted into Eq. (25), equating the coefficients of $\Pi_l k_l^{q_l}$ on both sides yields

$$\frac{1}{\mu!}\binom{\mu}{q}\langle w_1^{q_1}w_2^{q_2}\cdots w_N^{q_N}\rangle=\sum_{r+s=q}\frac{1}{n\,!\,m!}\binom{m}{r}\binom{n}{s}$$
$$\times[\gamma_s^{\phi}\langle w_1^{r_1}w_2^{r_2}\cdots w_N^{r_N}\rangle$$
$$+\sum_{j=1}^{N}\gamma_s^{E_j}\langle w_1^{r_1}\cdots w_j^{r_j+1}\cdots w_N^{r_N}\rangle], \tag{31}$$

where

$$\gamma_s^{\phi}=\frac{1}{T}\int_0^T dx\frac{1}{2}\left(1+\frac{\phi(x)-\theta}{V}\right)\left(\prod_l[\alpha+\mathring{\mathcal{L}}(x)]_l^{s_l}-\prod_l\alpha_l^{s_l}\right)$$
$$+\prod_l\alpha_l^{s_l},$$

$$\gamma_s^{E_j}=\frac{1}{T}\int_0^T dx\frac{1}{2}\frac{\mathring{\mathcal{E}}_j(x-x_j)}{V}\left(\prod_l[\alpha+\mathring{\mathcal{L}}(x)]_l^{s_l}-\prod_l\alpha_l^{s_l}\right),$$

and $q=(q_1q_2\cdots q_N)^T$, each q_i a non-negative integer, with $\Sigma_{i=1}^N\ q_i=\mu$. A slight simplification follows from $\gamma_0^{\phi}=1$ and $\gamma_0^{E}=0$: the quantity on the left side of Eq. (31) is canceled by the term on the right side with $s=0$ and $r=q$. The resulting recurrence relations are

$$0=\sum_{\substack{r+s=q\\ m<\mu}}\frac{1}{n\,!\,m!}\binom{m}{r}\binom{n}{s}[\gamma_s^{\phi}\langle w_1^{r_1}w_2^{r_2}\cdots w_N^{r_N}\rangle$$
$$+\sum_{j=1}^{N}\gamma^{E_j}\langle w_1^{r_1}\cdots w_j^{r_j+1}\cdots w_N^{r_N}\rangle],$$

$$0\leqslant q_i\leqslant\mu,\quad\sum_{i=1}^{N}q_i=\mu,\mu=1,2,\ldots. \tag{32}$$

For each choice of q we obtain a single linear equation involving moments of total order at most $\mu=\Sigma_i\ q_i$. Regarding the moments of total order μ as unknowns to be solved for in terms of moments of total order less than μ, we have a linear system with the same number of equations as unknowns. The coefficient matrix of this system involves the quantities $\gamma_{\cdot}^{\phi}$ and $\gamma_{\cdot}^{E}$. For generic $\mathcal{E}$ and $\mathcal{L}$ there are no polynomial relations among these quantities; hence the determinant of the coefficient matrix is generically nonzero, and the system can be inverted to give the moments of total order μ in terms of $\gamma_{\cdot}^{\phi}$, $\gamma_{\cdot}^{E}$, and the moments of total order less than μ. The complete moment hierarchy can thus be obtained: first moments of total order 1, then moments of total order 2, and so on.

A. Equilibrium mean

For $\mu=1$ we must have $q_j=\delta_{ij}$ for some i. Since in Eq. (32) only terms with $m<\mu$ appear, and $m=\Sigma_j r_j$, the only possibility for r is $r=0$, and then $s_j=q_j=\delta_{ij}$. The recurrence relation Eq. (32) then becomes

$$0=\alpha_i+\frac{1}{T}\int_0^T dx\frac{1}{2}\left(1+\frac{\phi(x)-\theta}{V}\right)\mathring{\mathcal{L}}_i(x-x_i)+\sum_{j=1}^{N}\langle w_j\rangle\frac{1}{T}\int_0^T dx\frac{1}{2V}\mathring{\mathcal{E}}_j(x-x_j)\mathring{\mathcal{L}}_i(x-x_i). \tag{33}$$

Allowing i to vary over all possible values $1,2,\ldots,N$, we have N linear equations in the N unknowns w_i, which can be written in vector form as

$$C\langle w\rangle=d, \tag{34}$$

with the matrix C and vector d given by

$$C_{ij}=-\frac{1}{2V}\frac{1}{T}\int_0^T dx\ \mathring{\mathcal{E}}_j(x-x_j)\mathring{\mathcal{L}}_i(x-x_i), \tag{35}$$

$$d_i=\alpha_i\frac{1}{T}\int_0^T dx\frac{1}{2}\left(1+\frac{\phi(x)-\theta}{V}\right)\mathring{\mathcal{L}}_i(x-x_i). \tag{36}$$

The overall minus sign in the definition of C is for later convenience. For generic $\mathcal{E}$ and $\mathcal{L}$ the matrix C is invertible, and we have $\langle w\rangle=C^{-1}d$. The physical meaning of this relation can be illuminated by rewriting Eq. (33) as follows:

$$0=\alpha_i+\frac{1}{T}\int_0^T dx\left(1+\frac{\phi(x)-\theta+\sum_{j=1}^{N}\langle w_j\rangle\mathring{\mathcal{E}}_j(x-x_j)}{V}\right)\mathring{\mathcal{L}}_i(x-x_i)=\alpha_i+\frac{1}{T}\int_0^T dx\langle f\rangle(x)\mathring{\mathcal{L}}_i(x-x_i),$$

where $\langle f\rangle(x)$ we define to be the value of $f(x)$ when $w=\langle w\rangle$. Now add and subtract $\alpha_i(1/T)\int_0^T dx\langle f\rangle(x)$ to obtain

$$0=\left[1-\frac{1}{T}\int_0^T dx\langle f\rangle(x)\right]\alpha_i+\frac{1}{T}\int_0^T dx\langle f\rangle(x)[\alpha_i+\mathring{\mathcal{L}}_i(x-x_i)]=\langle\Delta w_i\rangle. \tag{37}$$

We find that the equilibrium mean weight vector $\langle w\rangle$ is that for which the mean weight change is zero for all weights. This condition is obvious on independent grounds and could have been used to calculate $\langle w\rangle$ directly, without recourse to the moment hierarchy relations. But for moments of total order 2 or higher, transparent conditions such as this are not available; in that case we have no choice but to solve Eq. (32).

Given the equilibrium mean weights $\langle w\rangle$, we can calculate the equilibrium mean postsynaptic potential $\langle U\rangle(x)$ via

$$\langle U\rangle(x)=\phi(x)+\mathring{\mathcal{E}}(x)\cdot\langle w\rangle=\phi(x)+\mathring{\mathcal{E}}(x)\cdot C^{-1}d,$$

provided C is invertible.

B. Equilibrium variance

We now take $\mu=2$ and $q_k=\delta_{ik}+\delta_{jk}$ in Eq. (32). After some simplification, using C, d, $\langle f\rangle$, and $\langle w\rangle$ from above, we obtain

$$0=-\sum_{k=1}^{N}C_{jk}\langle w_kw_i\rangle-\sum_{k=1}^{N}C_{ik}\langle w_kw_j\rangle-\langle w_i\rangle d_j-\langle w_j\rangle d_i+\frac{1}{T}\int_0^T dx\langle f\rangle(x)\times\{[\alpha_i+\mathring{\mathcal{L}}_i(x-x_i)][\alpha_j+\mathring{\mathcal{L}}_j(x-x_j)]-\alpha_i\alpha_j\}.$$

This can be rearranged to give

$$\sum_{k=1}^{N}C_{jk}\langle w_kw_i\rangle+\sum_{k=1}^{N}C_{ik}\langle w_kw_j\rangle-\langle w_i\rangle\sum_{k=1}^{N}C_{jk}\langle w_k\rangle-\langle w_j\rangle\sum_{k=1}^{N}C_{ik}\langle w_k\rangle=\frac{1}{T}\int_0^T dx\langle f\rangle(x)\{[\alpha_i+\mathring{\mathcal{L}}_i(x-x_i)][\alpha_j+\mathring{\mathcal{L}}_j(x-x_j)]\}+\left[1-\frac{1}{T}\int_0^T dx\langle f\rangle(x)\right]\alpha_i\alpha_j=-\langle\Delta w_i\,\Delta w_j\rangle.$$

In vector form this becomes

$$C(\langle ww^T\rangle-\langle w\rangle\langle w\rangle^T)+(\langle w^2\rangle-\langle w\rangle\langle w\rangle^T)C^T=\langle\Delta w\,\Delta w^T\rangle. \tag{38}$$

The covariance of a vector random variable v is $\operatorname{cov} v=\langle vv^T\rangle-\langle v\rangle\langle v\rangle^T$. Equation (38) then takes the compact form

$$C(\operatorname{cov} w)+(\operatorname{cov} w)C^T=\operatorname{cov}\Delta w, \tag{39}$$

where we have used the equilibrium mean condition $\langle\Delta w\rangle=0$ on the right side. Equation (39) is a Lyapunov equation [28] for cov w, giving the equilibrium weight covariance in terms of C (which depends on $\mathcal{E}$ and $\mathcal{L}$) and covΔw (which depends on $\langle f\rangle$, α, and $\mathcal{L}$). Both C and covΔw can be calculated from the parameters of the system, and then the equilibrium covariance cov w, if it exists, must satisfy Eq. (39).

A theorem of Ostrowski and Schneider [28,29] gives conditions for the existence and uniqueness of solutions to Lyapunov equations. If S is symmetric positive definite and A and $-A$ have no common eigenvalues, then the Lyapunov equation $AH+HA^T=S$ has a unique solution H. Furthermore, H is symmetric and has the same *inertia* (number of eigenvalues with positive, zero, or negative real part) as A.

Since covΔw is necessarily symmetric positive definite, the theorem says that a symmetric solution cov w to Eq. (39) exists uniquely provided C and $-C$ have no common eigenvalues, and cov w is positive definite if and only if all eigenvalues of C have positive real part.

The condition that C and $-C$ have no common eigenvalues is true for generic C and hence for generic $\mathcal{E}$ and $\mathcal{L}$. The condition that cov w be positive definite is needed in order to interpret cov w as the covariance matrix of a probability distribution; we say cov w is *physical* if it is positive definite. Denoting by λ_n^C the nth eigenvalue of C, we then have the following physicality condition:

$$\text{cov}\ w\ \text{physical} \Leftrightarrow \text{Re}\lambda_n^C > 0 \quad \text{for all } n. \tag{40}$$

A theorem of Heinz [28,30] says that if all eigenvalues of A have positive real part and all eigenvalues of B have negative real part, then the (unique) solution X to the equation $AX-XB=Y$ is given by

$$X = \int_0^\infty ds\ e^{-sA} Y e^{sB}, \tag{41}$$

where the matrix exponentials are defined via Taylor expansions. The assumptions on the eigenvalues of A and B ensure that the integral in Eq. (41) converges, and one can show by direct substitution that the resulting X satisfies $AX-XB=Y$. If the physicality condition (40) holds, then C and $-C^T$ satisfy the conditions for A and B, respectively, and we obtain

$$\text{cov}\ w = \int_0^\infty dse^{-sC}(\text{cov}\ \Delta\ w)e^{-sC^T}. \tag{42}$$

This gives the equilibrium covariance matrix explicitly in terms of system parameters.

Since the postsynaptic potential $U(x)$ is a deterministic function of the synaptic weight vector w, the weight covariance cov w determines the covariance of the postsynaptic potential. From $U(x)=\phi(x)+\mathring{\mathcal{E}}(x)w$, we have

$$\text{cov}(U(x),U(y)) = \mathring{\mathcal{E}}(x)^T \text{cov}\ w\mathring{\mathcal{E}}(y) \tag{43}$$

for any pair of times x,y in the interval $[0,T]$. Of particular interest is the diagonal variance of $U(x)$:

$$\text{cov}(U(x),U(x)) = \mathring{\mathcal{E}}(x)^T \text{cov}\ w\mathring{\mathcal{E}}(x). \tag{44}$$

Our derivation of the equilibrium moment hierarchy equations relied on the equilibrium distribution of $U(x)$ being negligible on the "tails" of the postsynaptic spike probability function f. We will show in the next section, for the case of homogeneous parameters, that the confinement condition on $U(x)$ can always be satisfied by adjusting the rates of associative and nonassociative learning.

Note that for a spatially extended PSP $\mathcal{E}$, Eq. (44) implies that the diagonal variance of $U(x)$ depends on the full matrix cov w; in other words, it depends not only on the diagonal variances of the synaptic weights w, but also on the off-diagonal correlations between different synaptic weights.

VII. MULTIPLE WEIGHTS, HOMOGENEOUS PARAMETERS

For maximal generality in the foregoing analysis, we have allowed the postsynaptic potential functions and spike-timing-dependent learning rules to be different for different presynaptic neurons and have allowed the presynaptic spike times to be arbitrary. Further analytical progress can be made in the case where the system parameters are homogeneous; i.e., the postsynaptic potential functions and spike-timing-dependent learning rules are the same for all presynaptic neurons, and the presynaptic spike times are regularly spaced.

For such parameters it will turn out that the matrix C, the coefficient matrix in the Lyapunov equation (39) for cov w, has a special form: it is *circulant* [31]. The matrix covΔw on the right side of the Lyaponov equation for cov w is not circulant in general, but it is circulant if the postsynaptic spike probability density $\langle f\rangle(x)$ is independent of x. Now it was shown in [14] that in the case of homogeneous parameters, if the spacing δ between presynaptic spike times is sufficiently small and provided certain other constraints hold, the (mean) equilibrium weight vector has the property that the mean total postsynaptic potential $\langle U\rangle(x)$ is approximately constant.[3] In that case the mean equilibrium postsynaptic spike density $\langle f\rangle(x)$ is also approximately constant, and the matrix covΔw is approximately a circulant matrix D. The Lyapunov equation for cov w is then approximately

$$C(\text{cov}\ w) + (\text{cov}\ w)C^T = D, \tag{45}$$

with solution given by

$$\text{cov}\ w = \int_0^\infty ds\ e^{-sC} D e^{-sC^T}. \tag{46}$$

The eigenvalues and eigenvectors of circulant matrices are easily calculated; furthermore, all circulant matrices can be *simultaneously* diagonalized. Simultaneous diagonalization of C, C^T, and D in Eq. (46) will yield an explicit solution for cov w in terms of the eigenvectors and eigenvalues of C and D, which will themselves be written as explicit functions of the system parameters.

Let $\mathcal{E}(s)$, $\mathcal{L}(s)$, and α denote the common postsynaptic potential function, associative learning rule, and nonassociative learning rule, respectively. Let the spike time for presynaptic cell i be $x_i=(i-1)\delta$, $i=1,2,\ldots,N$, $\delta=T/N$. We then have

$$C_{ij} = -\frac{1}{2V}\frac{1}{T}\int_0^T dx\mathring{\mathcal{L}}(x-x_i)\mathring{\mathcal{E}}(x-x_j), \tag{47}$$

and for $\langle f\rangle(x)$ approximately the constant $\langle f\rangle$ we have cov$\Delta w \simeq D$, where

$$D_{ij} = \alpha^2(1-\langle f\rangle) + \langle f\rangle\frac{1}{T}\int_0^T dx[\alpha+\mathring{\mathcal{L}}(x-x_i)][\alpha+\mathring{\mathcal{L}}(x-x_j)].$$

By periodicity of $\mathring{\mathcal{L}}$, this can be simplified to

$$D_{ij} = \alpha^2 + 2\langle f\rangle\alpha\beta + \langle f\rangle\frac{1}{T}\int_0^T dx\mathring{\mathcal{L}}(x-x_i)\mathring{\mathcal{L}}(x-x_j), \tag{48}$$

where $\beta=(1/T)\int_0^T dx\mathring{\mathcal{L}}(x)$.

A matrix A is circulant [31] if each row of A equals the row above it shifted one entry to the right (and wrapped around at the edges); in other words,

[3] The present model differs from the model in [14] in having a postsynaptic spike probability density instead of a mean postsynaptic spike rate, but the argument is unaffected.

$$A_{(i+1)\mathrm{mod}\ N,(j+1)\mathrm{mod}\ N} = A_{ij} \quad \text{for all } i,j.$$

We now show that both C and D are circulant. First, let $g(x)$ and $h(x)$ be any periodic functions of x with period T, and let the $\{x_i\}$ be regularly spaced on $[0,T]$ as defined above. Let A be the matrix defined by

$$A_{ij} = \int_0^T dx\ f(x-x_i)g(x-x_j). \tag{49}$$

Taking (i,j) to $\bigl((i+1)\mathrm{mod}\ N,(j+1)\mathrm{mod}\ N\bigr)$ in Eq. (49) shifts the argument of both functions by $-\delta$, and by periodicity this does not change the value of the integral. Hence any matrix of the form (49) is circulant.

The constant matrices (all of whose entries are the same) are also circulant, and circulant matrices are closed under addition, scalar multiplication, and transposition. Hence by Eqs. (47) and (48), C and D are both circulant and so is C^T.

It is easily shown [31] that the vectors $u^{(n)}$, $n=1,2,\ldots,N$, with components

$$u_l^{(n)} = e^{2\pi i(l-1)n/N}, \quad k=1,2,\ldots,N, \tag{50}$$

are a complete set of eigenvectors for any circulant matrix A, with corresponding eigenvalue λ_n given by

$$\lambda_n = \sum_{l=1}^{N} A_{jl} e^{2\pi i(j-l)n/N}. \tag{51}$$

The expression on the right in Eq. (51) is independent of j because A_{jl} and the complex exponential both depend only on $(j-l)$ mod N. It is easily checked from Eq. (51) that adding a constant matrix (all entries the same) to a nonzero circulant matrix has no effect on its eigenvalues.

Let R be the unitary matrix whose nth column is the vector $u^{(n)}$, and let Λ be the diagonal matrix with entries λ_n. Then

$$A = R\Lambda R^*,$$

where R^* is the complex conjugate transpose of R.

In the present context it will be convenient to define wave numbers k_n so that the argument of the complex exponential in Eq. (50) is $ik_n x_l$; this we can arrange by taking $k_n = 2\pi n/T$, $n=1,2,\ldots,N$. From Eq. (51), the eigenvalues of C and D are then

$$\lambda_n^C = -\frac{1}{2V}\sum_{l=1}^{N} e^{ik_n(x_j-x_l)}\frac{1}{T}\int_0^T dx \mathring{\mathcal{L}}(x-x_j)\mathring{\mathcal{E}}(x-x_l),$$

$$\lambda_n^D = \langle f\rangle \sum_{l=1}^{N} e^{ik_n(x_j-x_l)}\frac{1}{T}\int_0^T dx \mathring{\mathcal{L}}(x-x_j)\mathring{\mathcal{L}}(x-x_l).$$

By periodicity of $\mathring{\mathcal{E}}$ and $\mathring{\mathcal{L}}$ and regular spacing of the $\{x_i\}$, these can be rewritten as

$$\lambda_n^C = -\frac{1}{2V}\sum_{l=1}^{N} e^{ik_n x_l}\frac{1}{T}\int_0^T dx \mathring{\mathcal{L}}(x-x_l)\mathring{\mathcal{E}}(x), \tag{52}$$

$$\lambda_n^D = \langle f\rangle \sum_{l=1}^{N} e^{ik_n x_l}\frac{1}{T}\int_0^T dx \mathring{\mathcal{L}}(x-x_l)\mathring{\mathcal{L}}(x). \tag{53}$$

Let Λ^C and Λ^D be the diagonal matrices with entries λ_n^C and λ_n^D, and let R be the unitary matrix defined above with entries

$$R_{jl} = u_j^{(l)} = e^{ik_l x_j}.$$

Then $C = R\Lambda^C R^*$ and $D = R\Lambda^D R^*$. Transposition takes eigenvalues to their complex conjugates, so $C^T = R\overline{\Lambda^C}R^*$. From $RR^* = I$ and Taylor expansion it follows that $e^{sC} = \mathrm{R}e^{s\Lambda^C}R^*$ and $e^{sC^T} = \mathrm{R}e^{s\overline{\Lambda^C}}R^*$. Substitution into Eq. (45) then yields a diagonalization of cov w:

$$\mathrm{cov}\ w = R\left[\int_0^\infty ds\ e^{-s\Lambda^C}\Lambda^D e^{-s\overline{\Lambda^C}}\right]R^* = R\Lambda^w R^*,$$

where Λ^w is the diagonal matrix with entries

$$\lambda_n^w = \int_0^\infty ds\ e^{-s\lambda_n^C}\lambda_n^D e^{-s\overline{\lambda_n^C}} = \frac{\lambda_n^D}{2\ \mathrm{Re}\ \lambda_n^C}, \tag{54}$$

provided Re $\lambda_n^C > 0$. Since D is symmetric positive definite (it is, by construction, a physical covariance matrix), we have λ_n^D real and positive for all n. Recall that in order for the solution of the Lyapunov equation (45) to be positive definite, all eigenvalues of C must have positive real part—i.e., Re $\lambda_n^C > 0$ for all n. If this physicality condition is satisfied, then the eigenvalues of cov w given by Eq. (54) are real and positive. These eigenvalues, with λ_n^C and λ_n^D given by Eqs. (52) and (53), are the variances associated with the independent components of the equilibrium weight distribution. The corresponding eigenvectors are the $u^{(n)}$, with components $u_j^{(n)} = e^{ik_n x_j}$.

Since $V > 0$, the condition for physicality of the covariance is

$$\mathrm{Re}\sum_{l=1}^{N} e^{ik_n x_l}\frac{1}{T}\int_0^T dx \mathring{\mathcal{L}}(x-x_l)\mathring{\mathcal{E}}(x) < 0 \text{ for all } n.$$

This coincides with the condition derived in [14] for stability of the *mean* weight state. Roughly speaking, it follows that if there exists an equilibrium weight distribution $P(w)$ (with finite covariance matrix), then the mean of the distribution must be stable. We do not address the stability of the equilibrium distribution (or, equivalently, the stability of all moments of the equilibrium distribution) in the present paper, but a natural conjecture would be that if the equilibrium distribution $P(w)$ exists, then it is necessarily stable.

From cov $w = R\Lambda^w R^*$ we can now write down explicit expressions for the equilibrium covariance of any pair of weights:

$$\mathrm{cov}(w_j,w_l) = \sum_{n,m=1}^{N} R_{jn}\Lambda_{nm}^w R_{ml}^* = \sum_{n=1}^{N} R_{jn}\overline{R_n}\lambda_n^w = \sum_{n=1}^{N} e^{ik_n(x_j-x_l)}\lambda_n^w, \tag{55}$$

with λ_n^w given by Eq. (54) and λ_n^C, λ_n^D given by Eqs. (52) and (53).

Note that $\mathrm{cov}(w_j,w_l)$ depends on j and l only via the difference (x_j-x_l) mod T, due to periodicity and translational invariance of the architecture for homogeneous parameters. Also, the covariance of the weights depends only on the associative part $\mathcal{L}$ of the learning rule, since the nonassociative part α does not appear in Eq. (55). This is not surprising, since the role of α is essentially analagous to that of a constant externally applied force in a physical system. Such a force changes the position of the equilibrium, but does not alter the dynamics around the equilibrium.

A. Confinement

Our derivation of the moment hierarchy relations, Eqs. (32), relied on the assumption that the equilibrium weight distribution was negligible on the "tails" of the piecewise linear postsynaptic gain function f. This places a constraint on the mean $\langle U\rangle(x)$ and diagonal variance $\mathrm{cov}\big(U(x),U(x)\big)$ of the postsynaptic potential: they must be such that the mean is a large number of standard deviations away from the tails. For each x, let $r(x)$ be the standard deviation of $U(x)$ divided by the distance from $\langle U\rangle(x)$ to the nearest tail—i.e., to $V-\theta$ or $-V-\theta$. The parameter $r(x)$ will be referred to as the *confinement parameter* for the system. The confinement condition holds provided $\langle U\rangle(x)$ is in the interval $(-V-\theta,V-\theta)$ and $r(x)\ll 1$, for all x.

We now argue that by adjusting only the rates of nonassociative and associative learning, the confinement condition can always be satisfied. Multiplying the associative learning rule by a positive scalar factor β and both nonassociative and associative components by a positive scalar factor λ, we have weight changes given by

$$\Delta w(t)=\begin{cases}\lambda\alpha+\beta\mathring{\mathcal{L}}(x), \text{density } (1/T)\tilde{f}(x,w(t)),\\ \lambda\alpha, \quad \text{probability } 1-(1/T)\displaystyle\int_0^T dx\tilde{f}(x,w(t)).\end{cases} \tag{56}$$

The ratio of associative to nonassociative learning rate is parametrized by β, while the overall learning rate is parametrized by λ. Now it was shown in [14] that in the case of homogeneous parameters, under certain mild conditions, the equilibrium mean weight vector has the property that $\langle U\rangle(x)$ is approximately constant (i.e., the equilibrium is an approximate negative image state). Hence $\langle f\rangle$ in Eq. (37) is approximately constant. If it were exactly constant, then Eq. (37) (for homogeneous parameters) would yield, after cancelling λ on top and bottom,

$$\langle f\rangle=\frac{-\alpha}{\alpha+\dfrac{\beta}{T}\int dx\mathring{\mathcal{L}}(x)}.$$

Provided α and $\int dx\mathring{\mathcal{L}}(x)$ have opposite sign (shown in [14] to be necessary for existence of a negative image equilibrium) the right side of this equation can be made to have any desired value by appropriate choice of $\beta>0$. Hence $\langle f\rangle$ can be made to have any desired value by appropriate choice of β; in particular, a range of β exists for which $\langle f\rangle$ falls in the open interval $\big(f(-V-\theta),f(V-\theta)\big)$. Since f is invertible for arguments in $(-V-\theta,V-\theta)$ and $\langle f\rangle=f(\langle U\rangle)$, it follows that by appropriate choice of β, $\langle U\rangle$ can be made to have any value in $(-V-\theta,V-\theta)$. Since $\langle f\rangle(x)$ approximately constant implies $\langle U\rangle$ approximately constant, it follows that the mean postsynaptic potential $\langle U\rangle(x)$ can always be made to lie between the tails, for all x.

It remains to show that the diagonal variance $\mathrm{cov}\big(U(x),U(x)\big)$ can be made sufficiently small so that the distribution of $U(x)$ is negligible on the tails. We do this by holding β fixed and varying λ. Since the matrix C is proportional to λ and the matrix cov Δw is proportional to λ^2, it follows from Eq. (38) that cov w—and hence cov U from Eq. (48)—is proportional to λ. In particular, $\mathrm{cov}\big(U(x),U(x)\big)$ can be made arbitrarily small by taking λ sufficiently small.

Thus, by appropriate choice of β and λ, the confinement condition can always be satisfied. The value of β determines the location of the mean postsynaptic potential, and the value of λ determines the width of the distribution around the mean. The latter fact—that the width of the equilibrium distribution of the postsynaptic potential is proportional to the overall learning rate—has direct behavioral relevance to the mormyrid fish, since it implies a tradeoff between speed of adaptation and accuracy of the adapted state.[4]

B. Dense spacing limit

In the architecture of the mormyrid ELL, the spacing δ between presynaptic spike times is much less than the widths τ_E, τ_L of the PSP $\mathcal{E}$ and learning rule $\mathcal{L}$. In the dense spacing limit the set of discrete weights per unit time $\{w_i/\delta\}$ corresponding to presynaptic spikes at times $\{x_i\}$ becomes a continuum weight density $\mathcal{W}(y)$, with weight $\mathcal{W}(y)dy$ corresponding to presynaptic spike times between y and $y+dy$. Sums over x_i are replaced by integrals over y. The matrices C and D in Eq. (45) become infinite dimensional, with eigenvalues λ_n^C, λ_n^D given by

$$\lambda_n^C=-\frac{1}{2VT}\int_0^T dy\; e^{ik_ny}\int_0^T dx\mathring{\mathcal{L}}(x-y)\mathring{\mathcal{E}}(x), \tag{57}$$

$$\lambda_n^D=-\frac{\langle f\rangle}{T}\int_0^T dy\; e^{ik_ny}\int_0^T dx\mathring{\mathcal{L}}(x-y)\mathring{\mathcal{L}}(x), \tag{58}$$

for $n=0,1,\ldots$. We introduce some useful notation. Let $\mathcal{F}_T[h]$ be the sequence of Fourier coefficients for a function h on $[0,T]$, given by $\mathcal{F}_T[h]_n=\int_0^T dy\; e^{ik_ny}h(y)$ with $k_n=2\pi n/T$, $n=0,1,\ldots$. Let $*_T$ denote convolution on the interval $[0,T]$, $(g*_Th)(x)=\int_0^T dy\; g(x-y)h(y)$. Let $\tilde{h}$ denote the horizontal re-

[4] The fact that the variance is proportional to the learning rate is also true for inhomogeneous parameters, by the same argument. But the confinement of the mean postsynaptic potential $\langle U\rangle(x)$ is unclear in that case, because the equilibrium is not necessarily an approximate negative image. Further work is required to characterize the equilibrium for inhomogeneous parameters.

flection of $h, \tilde{h}(y) = h(-y)$. Then Eqs. (57) and (58) can be written as

$$\lambda_n^C = -\frac{1}{2VT}\mathcal{F}_T[\mathring{\mathcal{L}} *_T \tilde{\mathring{\mathcal{E}}}]_n,$$

$$\lambda_n^D = \frac{\langle f\rangle}{T}\mathcal{F}_T[\mathring{\mathcal{L}} *_T \tilde{\mathring{\mathcal{L}}}]_n.$$

Now we invoke the Fourier convolution theorem $\mathcal{F}_T[g*h] = \mathcal{F}_T[g]\mathcal{F}_T[h]$ and the fact that $\mathcal{F}_T[\tilde{g}] = \overline{\mathcal{F}_T[g]}$, where $\overline{z}$ denotes the complex conjugate of z. This gives

$$\lambda_n^C = -\frac{1}{2VT}\mathcal{F}_T[\mathring{\mathcal{L}}]_n\overline{\mathcal{F}_T[\mathring{\mathcal{E}}]_n}, \tag{59}$$

$$\lambda_n^D = -\frac{\langle f\rangle}{T}\mathcal{F}_T[\mathring{\mathcal{L}}]_n\overline{\mathcal{F}_T[\mathring{\mathcal{L}}]_n}. \tag{60}$$

The eigenvalues of the weight covariance are therefore

$$\lambda_n^{\mathcal{W}} = \frac{\lambda_n^D}{2 \text{ Re } \lambda_n^C} = -\langle f\rangle V\frac{\mathcal{F}_T[\mathring{\mathcal{L}}]_n\overline{\mathcal{F}_T[\mathring{\mathcal{L}}]_n}}{\text{Re}[\mathcal{F}_T[\mathring{\mathcal{L}}]_n\mathcal{F}_T[\mathring{\mathcal{E}}]_n]}. \tag{61}$$

It follows that the covariance of $\mathcal{W}(y)$ and $\mathcal{W}(z)$ is

$$\text{cov }(\mathcal{W}(y),\mathcal{W}(z)) = \sum_{n=0}^{\infty} e^{ik_n(y-z)}\lambda_n^W = -2\pi\langle f\rangle V\mathcal{F}_T^{-1} \times\left[\frac{\mathcal{F}_T[\mathring{\mathcal{L}}]\overline{\mathcal{F}_T[\mathring{\mathcal{L}}]}}{\text{Re}[\mathcal{F}_T[\mathring{\mathcal{L}}]\overline{\overline{\mathcal{F}_T[\mathring{\mathcal{E}}]}}]}\right](y-z), \tag{62}$$

where $\mathcal{F}_T^{-1}[h](x) = (1/2\pi)\Sigma_{n=0}^{\infty}e^{ik_nx}h_n$ is the inverse Fourier transform on $[0,T]$. The covariance of the postsynaptic potential is then

$$\begin{aligned}\text{cov } U(y,z) &= \int_0^T dx\int_0^T dx'\,\mathring{\mathcal{E}}(y-x)\text{cov } \mathcal{W}(x,x')\mathring{\mathcal{E}}(z-x') \\ &= -2\pi\langle f\rangle V\int_0^T dx\int_0^T dx'\,\mathring{\mathcal{E}}(y-x)\mathring{\mathcal{E}}(z-x') \\ &\quad\times \mathcal{F}_T^{-1}\left[\frac{\mathcal{F}_T[\mathring{\mathcal{L}}]\overline{\mathcal{F}_T[\mathring{\mathcal{L}}]}}{\text{Re}[\mathcal{F}_T[\mathring{\mathcal{L}}]\overline{\overline{\mathcal{F}_T[\mathring{\mathcal{E}}]}}]}\right](x-x'). \end{aligned}\tag{63}$$

One special case is worth noting: suppose the PSP and learning rule have identical functional form—i.e., are proportional to one another—$\mathcal{L}(x) = c\mathcal{E}(x)$ for some (real) constant c. Then we have

$$\mathcal{F}_T^{-1}\left[\frac{\mathcal{F}_T[\mathring{\mathcal{L}}]\overline{\mathcal{F}_T[\mathring{\mathcal{L}}]}}{\text{Re}[\mathcal{F}_T[\mathring{\mathcal{L}}]\overline{\overline{\mathcal{F}_T[\mathring{\mathcal{E}}]}}]}\right]x = \mathcal{F}_T^{-1}[c](x) = \frac{c}{2\pi}\delta(x),$$

where $\delta(x)$ is the Dirac delta function. For such a learning rule the covariance of the weight density is

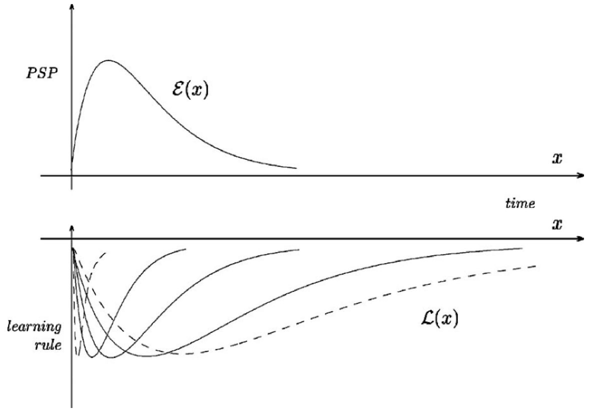

FIG. 2. PSP and learning rules used in the examples. Stability requires $3-2\sqrt{2} < \tau L/\tau E < 3+2\sqrt{2}$. Stable examples are drawn with solid lines; end points of the stable interval are drawn with dashed lines. Arbitrary units.

$$\text{cov}(\mathcal{W}(y),\mathcal{W}(z)) = -\langle f\rangle Vc\,\delta(y-z). \tag{64}$$

In particular, the covariance (and hence the correlation) of $\mathcal{W}(y)$ and $\mathcal{W}(z)$ is zero for $y \neq z$; hence weights corresponding to different presynaptic spike times are statistically independent. This is surprising, since the coupling of weights through the PSP $\mathcal{E}$ and learning rule $\mathcal{L}$ has some nonzero "range," given roughly by the widths of $\mathcal{E}$ and $\mathcal{L}$, and within this range one would expect the weights to necessarily have some nonzero correlation. The result just derived says that in certain exceptional cases this correlation may vanish. The result was derived in the dense spacing limit, but can be expected to hold approximately for the physical case of discrete spacing and also to hold approximately for $\mathcal{L}$ not quite proportional to $\mathcal{E}$; this will be verified in the examples calculated below. Given that the best current experimental measurement of the learning rule in the mormyrid ELL [3] is not inconsistent with $\mathcal{E}$ and $\mathcal{L}$ having the same functional form, this vanishing correlation phenomenon may have biological relevance.

VIII. EXAMPLES

We now compute the equilibrium weight covariances for a class of PSP's and learning rules consistent with those measured in the mormyrid ELL, assuming homogeneous parameters. The PSP we take to be an excitatory alpha function of width τ_E, and the learning rule we take to be alpha function, depressive, and pre-before-post, of width τ_L:

$$\mathcal{E}(x) = \tau_E^2 e^{-x/\tau_E}H(x), \tag{65}$$

$$\mathcal{L}(x) = -\tau_L^2 e^{-x/\tau_L}H(x), \tag{66}$$

where $H(x)$ is the Heaviside function: $H(x)=1$ if $x \geqslant 0$ and 0 otherwise (Fig. 2). In the above expressions both $\mathcal{E}$ and $\mathcal{L}$ have been normalized to unit area, but to ensure confinement of the postsynaptic potential, the learning rule $\mathcal{L}$ (and hence the size of the learning steps) must be made sufficiently small so that the confinement condition is satisfied.

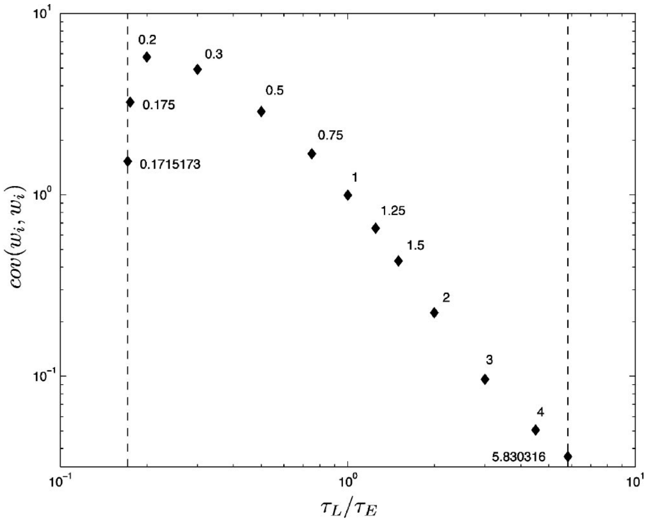

FIG. 3. Diagonal variance of weights, for alpha functions $\mathcal{E}$ and $\mathcal{L}$ and for various values of τ_L/τ_E. The larger of τ_L and τ_E was taken to be $0.2T$ in all cases. Diagonal variance vs τ_L/τ_E, log-log plot. Dotted lines indicate the boundary of the stable interval, $\tau_L/\tau_E = 3\pm 2\sqrt{2}$. Dimensionless units.

It was shown in [14] that in order for the mean weight dynamics to be stable near the (negative image) equilibrium, the time constants τ_E and τ_L must satisfy

$$3-2\sqrt{2} < \frac{\tau_L}{\tau_E} < 3+2\sqrt{2}.$$

For τ_L/τ_E in this stable range, we calculated the equilibrium covariance of the synaptic weights and of the postsynaptic potential and verified our predictions by direct Monte Carlo simulation of the underlying random walk. The number of presynaptic cells was taken to be $N=50$, and to ensure that the confinement condition was well satisfied, the rates of nonassociative and associative learning were adjusted so that the confinement parameter was $r(x)=0.2$ for all x (i.e., the tails were five standard deviations away from the mean postsynaptic potential). By translational symmetry for homogeneous parameters, the diagonal variances (w_i, w_i) are independent of i, and the off-diagonal covariance (w_i, w_j) depends only on $(x_i - x_j)$mod T. The covariance matrix is then completely described by the diagonal variance (a single number) and the correlation of weight w_i with the "midpoint" weight $w_{N/2}$, for $i=1,2,\ldots,N$; the correlation in this case is just the covariance normalized by the diagonal variance. The diagonal variance is shown in Fig. 3, and the correlation is shown in Fig. 4, for various values of τ_L/τ_E between $3-2\sqrt{2}$ and $3+2\sqrt{2}$. Note the approximate vanishing of off-diagonal correlation for τ_L/τ_E near 1, as expected from the analytic calculation in the dense-spacing limit. The manner in which the correlation deviates from an approximate delta function as τ_L/τ_E deviates from 1 also shows an interesting pattern: for τ_L/τ_E slightly greater than 1, the near-diagonal (near-neighbor) correlation is positive, while for τ_L/τ_E slightly less than 1, the near-neighbor correlation is negative. But for τ_L/τ_E substantially greater than or less than 1, the near-neighbor correlation is positive in both cases. The magnitude of off-diagonal correlation tends to increase as τ_L/τ_E moves away from 1 in either direction. Near the limits of the stable range of τ_L/τ_E, the near-neighbor correlation is close to 1 and the "antipodal" correlation (correlation with weights a half period away) is close to -1. Such strong long-range correlation and anticorrelation was also observed numerically in [15] in mean weight dynamics for parameters near the boundary of the stable region, with breakdown of stability being characterized by the appearance of traveling waves.

The correlation of the postsynaptic potential is shown in Fig. 5. For τ_L/τ_E near 1 the correlation is everywhere positive. As τ_L/τ_E deviates from 1, the correlation decreases, and long-range anticorrelations appear. As τ_L/τ_E deviates still further, the anticorrelation decreases in range and increases in magnitude, and a positive long-range correlation appears. For τ_L/τ_E near the limits of the stable range, the midrange and long-range (antipodal) correlations approach -1 and $+1$, respectively, similar to the behavior of the synaptic weight correlation. The "scalloped" appearance of these curves for large τ_L/τ_E is due to τ_E being not much larger than the spacing $\delta = T/50$ between presynaptic spike times, resulting in only marginal overlap of adjacent PSP's. For fixed PSP width τ_E, such scalloping should vanish as the spacing of presynaptic spike times goes to zero. It is believed [27(b)] that in the mormyrid ELL the spacing of presynaptic spike times is sufficiently dense that this scalloping would be insignificant.

Comparison with direct Monte Carlo simulation of the random walk revealed excellent agreement with prediction, provided confinement was well satisfied; results for $\tau_L/\tau_E = 5.814$, near the upper end of the stable range, are shown in Fig. 6. As above, nonassociative and associative learning rates were adjusted so that the confinement parameter $r(x)$ was 0.2 for all x (i.e., the tails were five standard deviations away from the equilibrium mean). Weights were taken to be initially uncorrelated, with mean equal to the predicted mean and variance equal to the predicted (diagonal) variance; the initial correlation was then the discrete Dirac delta function. To quantify convergence we used the mean absolute value of the relative discrepancy between the predicted and actual (ensemble mean) correlation. Translation invariance of the correlation allowed us to reduce the size of fluctuations in the simulation estimate by averaging not just over the ensemble but also over the population of $N=50$ weights in each member of the ensemble.[5] Using this measure, the correlation in the simulation converged to within 1%–2% of the predicted correlation in approximately 10^7 time steps (Fig. 6).

IX. DISCUSSION

Since changes in synaptic weights in STDP are due to temporally discrete events (spikes or spike pairs), the dynam-

[5] Although the predicted correlation is translation invariant, the fluctuations around the prediction are not necessarily uncorrelated. For our purposes this is harmless; it simply means that we do not obtain as large a reduction in fluctuation size by population averaging as we would by using a 50-times larger ensemble.

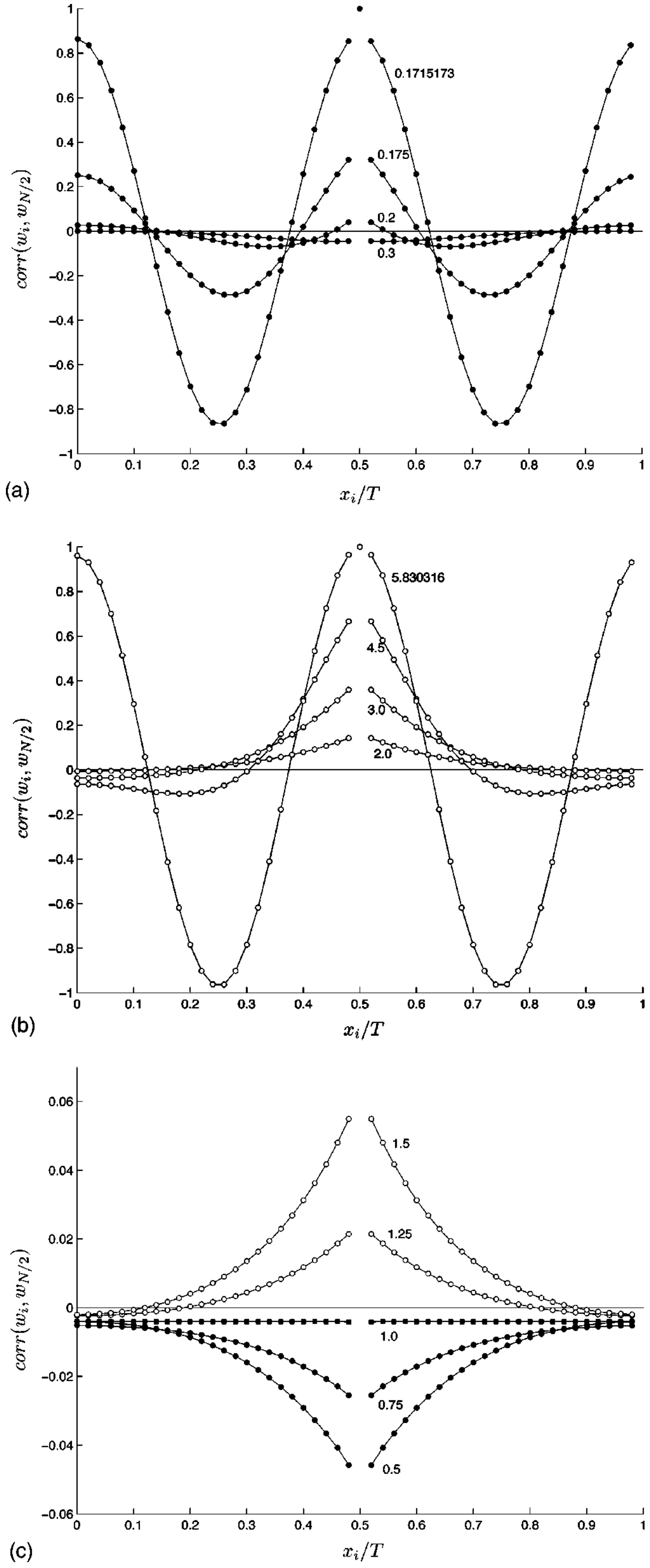

FIG. 4. Correlation of weights, for alpha functions $\mathcal{E}$ and $\mathcal{L}$ and for various values of τ_L/τ_E. The larger of τ_L and τ_E was taken to be $0.2T$ in all cases. Curves are labeled by the value of τ_L/τ_E, and for clarity curves are not joined to the point (0.5,1) which all curves have in common. (a) Correlation of w_i with $w_{N/2}$, versus x_i/T, for τ_L/τ_E significantly less than 1. (b) Same for τ_L/τ_E significantly greater than 1. (c) Same for τ_L/τ_E near 1, with expanded vertical scale. Dimensionless units.

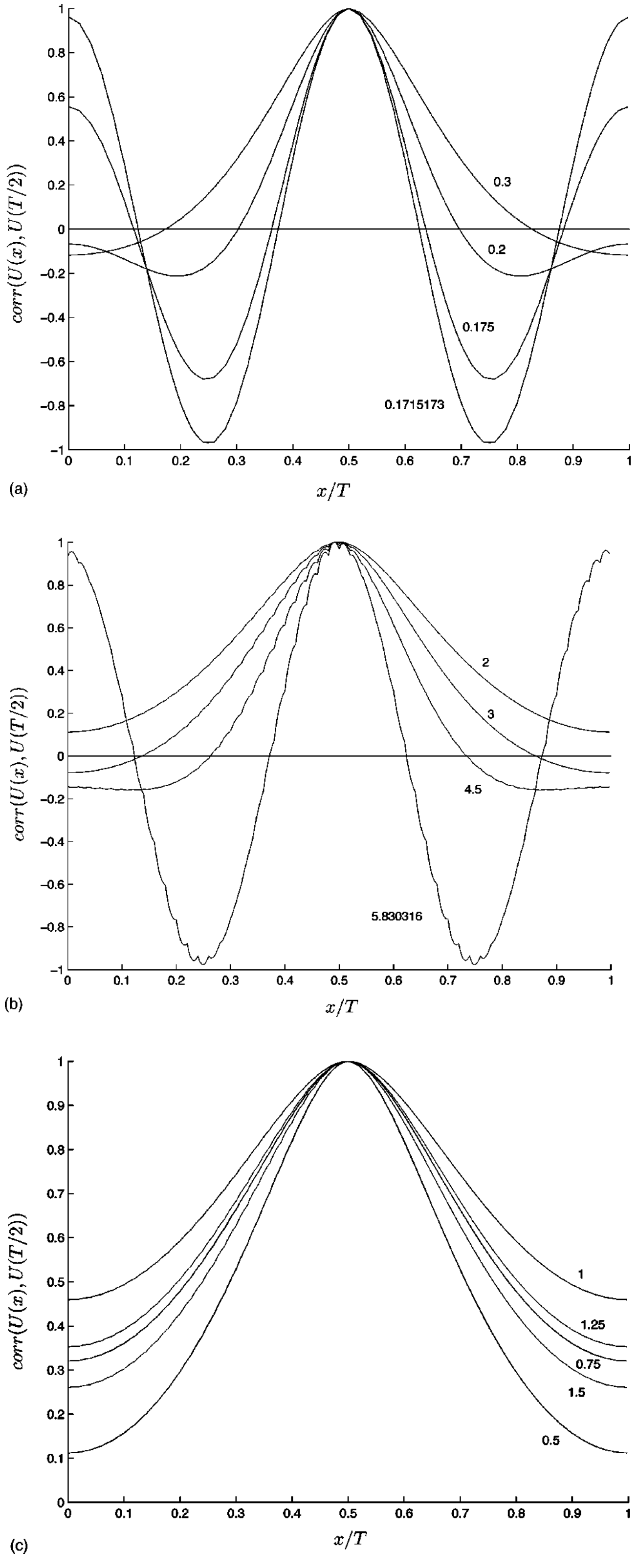

FIG. 5. Correlation of postsynaptic potential, for alpha functions $\mathcal{E}$ and $\mathcal{L}$ and for various values of τ_L/τ_E. The larger of τ_L and τ_E was taken to be $0.2T$ in all cases. (a) Correlation of $U(x)$ with $U(T/2)$, versus x/T, for τ_L/τ_E significantly less than 1. (b) Same for τ_L/τ_E significantly greater than 1. (c) Same for τ_L/τ_E near 1, with expanded vertical scale. Curves are labeled by the value of τ_L/τ_E. Dimensionless units.

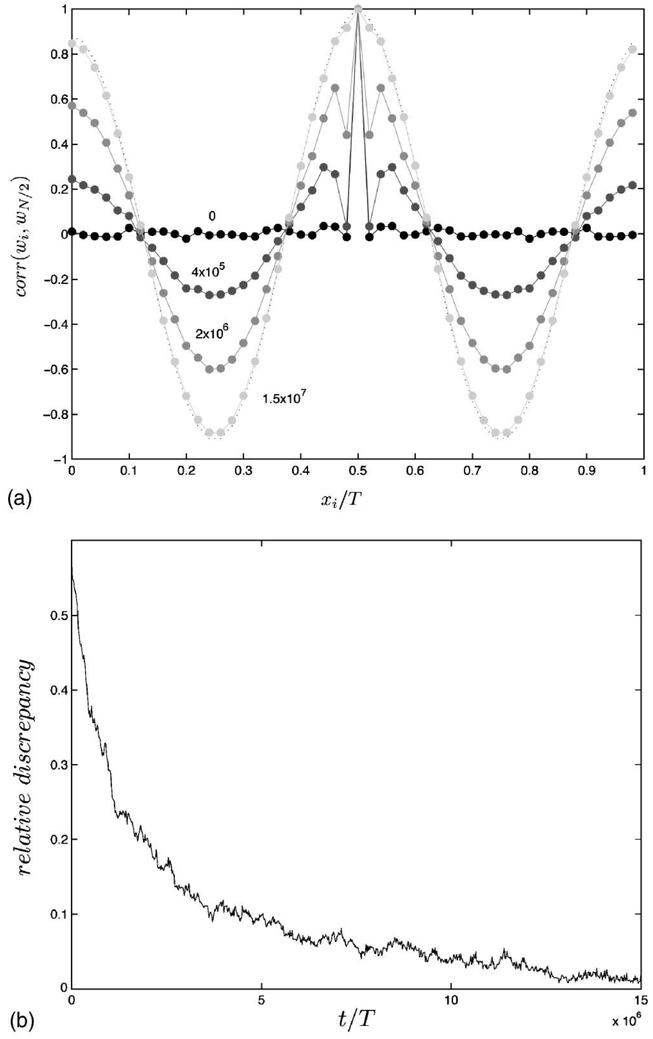

FIG. 6. Convergence of weight correlation to predicted equilibrium values in Monte Carlo simulations, for $L/E=5.81$, $N=50$, confinement parameter =0.2. (a) Time evolution of population-averaged correlation; curves labeled by time, t/T. Dotted curve indicates prediction. (b) Relative discrepancy between predicted and actual correlation, vs time t/T. Dimensionless units.

ics of such plasticity, in the presence of noise, is naturally modeled as a discrete-time random walk. There is a large body of mathematical technique for the analysis of such processes [17].

From the weight dynamics expressed as a random walk one can write down a master equation for the time evolution of the weight probability distribution. From the master equation we obtain a functional equation for the equilibrium weight distribution. Taking the Fourier transform of this equation yields a differential equation for the characteristic function of the equilibrium distribution, and Taylor expansion then yields a hierarchy of recurrence relations for the equilibrium moments. From the moments of the equilibrium weight distribution we also obtain the moments of the postsynaptic membrane potential.

For the case of a single weight, we explicitly calculate moments up to fourth order. The distribution is shown to be generically non-Gaussian, but the skew and kurtosis approach Gaussian values as the learning rate (size of steps) goes to zero.

For the case of multiple weights we explicitly calculate moments up to second order. The mean weight vector satisfies a simple matrix-vector equation, which is equivalent to the condition that the mean step in the equilibrium state be zero for all weights. The weight covariance matrix satisfies a Lyapunov equation. An explicit solution to this equation, in the form of a matrix integral, is obtained. For this solution to be the covariance matrix of some probability distribution it must be positive definite, which imposes a constraint on the PSP $\mathcal{E}$ and the associative learning rule $\mathcal{L}$.

For the case of multiple weights with homogeneous parameters, further analytical progress can be made. The Lyapunov equation for the weight covariance matrix can be fully diagonalized and the covariance of any pair of weights found in closed form. From this we also obtain explicit expressions for the covariance of the postsynaptic potential between any pair of times. The physicality condition—that the weight covariance matrix be positive definite—takes an especially simple form in this case, closely related to the condition derived in [14] for stability of the mean-weight state.

In the limit of dense spacing of presynaptic spike times, the expression for the weight covariance is further simplified. In the special case where $\mathcal{E}$ and $\mathcal{L}$ have the same functional form, we find, surprisingly, that weights corresponding to distinct presynaptic spike times are statistically independent. This result can be expected to hold approximately for discrete presynaptic spike times and for learning rules not quite identical to $\mathcal{E}$ in functional form.

Numerical calculation of the equilibrium weight covariance and postsynaptic potential covariance was carried out for a class of examples relevant to the mormyrid ELL: both $\mathcal{E}$ and $\mathcal{L}$ alpha function in form, with $\mathcal{E}$ excitatory and $\mathcal{L}$ depressive pre-before-post. For the synaptic weights, off-diagonal correlation is near zero for $\tau_L/\tau_E=1$ and tends to increase in magnitude as τ_L/τ_E moves away from 1. Values of τ_L/τ_E near the boundary of the stable range show large long-range anticorrelations. The correlation of the postsynaptic potential is everywhere positive for $\tau_L/\tau_E=1$, but long-range anticorrelations develop as τ_L/τ_E moves away from 1. These numerical predictions were found to be in excellent agreement with direct Monte Carlo simulations of the underlying random walk.

One of the basic results of this paper is that the variance of the equilibrium weight distribution is proportional to learning rate (i.e., to the magnitude of the weight changes induced by individual spikes or spike pairs). A slow learning rate leads to a small variance in equilibrium weight distribution and hence a more accurate negative image; a fast learning rate gives a large variance in equilibrium weight distribution and a less accurate negative image. Detectability of sensory objects is improved by a more accurate negative image; thus to optimize detectability the learning rate should be slow. However, if the fish's own discharge is changing (due to changes in water conductivity or body shape, for example), then the negative image must be updated to remain accurate. Such adaptability of the negative image favors a fast learning rate, to allow the negative image to keep up with changes in the discharge. The twin requirements of detectability and adaptability are thus in direct conflict: any one choice of learning rate represents a compromise between them. A natural hypothesis is that the learning rate in the mormyrid ELL is the slowest learning rate sufficient to provide adaptability of the negative image on time scales over which the fish's discharge varies in the wild. A faster rate would not significantly improve adaptability and would degrade detectability; a slower rate would unacceptably degrade adaptability.

Verification of this hypothesis concerning optimality of the physiological learning rate, and other quantitative predictions of the present paper, requires further experimental work. Direct observation of the equilibrium variance of synaptic weights is probably not feasible, but measurement of the equilibrium variance of the postsynaptic membrane potential in MG cells is certainly feasible. Since sources other than the equilibrium weight variance may also contribute to a fluctuating membrane potential, such a measurement can only provide an upper bound for the learning rate consistent with our calculations. Nevertheless, if this upper bound were too slow to be consistent with direct experimental measurement of weight changes due to single spike pairs [3], then our calculation would be inconsistent with experiment, and the model would need to be modified. For a sharper test of our quantitative predictions, further work must be done to characterize other sources of variance in the MG cell membrane potential, so that the contribution due to synaptic weight variance alone can be isolated. We hope the specificity and quantitative nature of our predictions are sufficient to motivate such work.

Although certain details of our model are drawn from a particular biological system, we have sought in the present paper to lay the groundwork for the rigorous mathematical analysis of equilibrium weight distributions arising from STDP in other systems as well. The methods developed here, in particular the random walk approach and the ability to calculate with arbitrary learning rules and arbitrary postsynaptic potential functions, are quite general and should be extendable to systems other than the mormyrid ELL.

ACKNOWLEDGMENTS

We would like to thank Dr. Gerhard Magnus, Dr. Nathaniel Sawtell, and the members of Dr. Curtis Bell's laboratory for insightful discussions. This material is based upon work supported by the National Science Foundation under Grant No. IBN-0114558 and by the National Institute of Mental Health under Grant No. R01-MH60364.

[1] L. F. Abbott and S. B. Nelson, Nature (London) **3**, 1178 (2000).
[2] H. Markram, J. Lübke, M. Frotscher, and B. Sakmann, Science **275**, 213 (1997).
[3] C. C. Bell, V. Han, Y. Sugawara, and K. Grant, Nature (London) **387**, 278 (1997).
[4] Q. Bi and M. Poo, J. Neurosci. **18**, 10464 (1998).
[5] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
[6] T. J. Sejnowski, J. Theor. Biol. **69**, 385 (1977).
[7] E. L. Bienenstock, L. N. Cooper, and P. W. Munro, J. Neurosci. **2**, 32 (1982).
[8] W. Gerstner, R. Kempter, J. L. van Hemmen, and H. Wagner, Nature (London) **383**, 76 (1996).
[9] M. C. W. van Rossum, G. Q. Bi, and G. G. Turrigiano, J. Neurosci. **20**, 88128821 (2000).
[10] J. Rubin, D. D. Lee, and H. Sompolinsky, Phys. Rev. Lett. **86**, 364 (2001).
[11] M. Yoshioka, Phys. Rev. E **65**, 011903 (2002).
[12] V. P. Zhigulin, M. I. Rabinovich, R. Huerta, and H. D. Abarbanel, Phys. Rev. E **67**, 021901 (2003).
[13] H. Cateau and T. Fukai, Neural Comput. **15**, 597 (2003).
[14] A. Williams, P. D. Roberts, and T. K. Leen, Phys. Rev. E **68**, 021923 (2003).
[15] P. D. Roberts, Phys. Rev. E **62**, 4077 (2000).
[16] L. H. Liyanage, C. M. Gulati, and J. M. Hill, Adv. Mol. Relax. Interact. Processes **22**, 53 (1982).
[17] B. D. Hughes, *Random Walks and Random Environments* (Oxford University Press, Oxford, 1995).
[18] N. G. V. Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
[19] R. Kempter, W. Gerstner, and J. L. van Hemmen, Phys. Rev. E **59**, 4498 (1999).
[20] P. D. Roberts, J. Comput. Neurosci. **7**, 235 (1999).
[21] P. D. Roberts and C. C. Bell, J. Comput. Neurosci. **9**, 67 (2000).
[22] R. H. Hahnloser, A. A. Kozhevnikov, and M. S. Fee, Nature (London) **419**, 65 (2002).
[23] D. Ehrlich, J. H. Casseday, and E. Covey, J. Neurophysiol. **77**, 2360 (1997).
[24] W. Gerstner, R. Ritz, and J. L. van Hemmen, Biol. Cybern. **69**, 503 (1993).
[25] W. Gerstner, Phys. Rev. E **51**, 738 (1995).
[26] P. D. Roberts, J. Neurosci. **84**, 2035 (2000).
[27] (a) C. C. Bell, K. Grant, and J. Serrier, J. Neurosci. **68**, 843 (1992); (b) C. C. Bell (private communication).
[28] R. Bhatia and P. Rosenthal, Bull. London Math. Soc. **29**, 1 (1997).
[29] A. Ostrowski and H. Schneider, J. Math. Anal. Appl. **4**, 72 (1962).
[30] E. Heinz, Math. Ann. **123**, 415 (1951).
[31] P. J. Davis, *Circulant Matrices* (Wiley, New York, 1979).